

## Article

# Une analyse bayésienne des probabilités de réponse dans les petits domaines sous une contrainte

par Balgobin Nandram et Hasanjan Sayit

Décembre 2011



# Une analyse bayésienne des probabilités de réponse dans les petits domaines sous une contrainte

Balgobin Nandram et Hasanjan Sayit<sup>1</sup>

## Résumé

De nombreuses enquêtes par sondage comprennent des questions suscitant une réponse binaire (par exemple, obèse, non obèse) pour un certain nombre de petits domaines. Une inférence est requise au sujet de la probabilité d'une réponse positive (par exemple obèse) dans chaque domaine, la probabilité étant la même pour tous les individus dans chaque domaine et différente entre les domaines. Étant donné le peu de données dans les domaines, les estimateurs directs ne sont pas fiables et il est nécessaire d'utiliser des données provenant d'autres domaines pour améliorer l'inférence pour un domaine particulier. Essentiellement, il est supposé a priori que les domaines sont similaires, si bien que le choix d'un modèle hiérarchique bayésien, le modèle bêta-binomial standard, est naturel. L'innovation tient au fait qu'un praticien peut disposer d'information a priori supplémentaire qui est nécessaire au sujet d'une combinaison linéaire des probabilités. Par exemple, une moyenne pondérée des probabilités est un paramètre, et l'information peut être obtenue au sujet de ce paramètre, ce qui rend le paradigme bayésien approprié. Nous avons modifié le modèle bêta-binomial standard pour petits domaines afin d'y intégrer l'information a priori sur la combinaison linéaire des probabilités, que nous appelons une contrainte. Donc, il existe trois cas. Le praticien a) ne spécifie pas de contrainte, b) spécifie une contrainte et le paramètre entièrement et c) spécifie une contrainte et l'information qui peut être utilisée pour construire une loi a priori pour le paramètre. L'échantillonneur de Gibbs « griddy » est utilisé pour ajuster les modèles. Pour illustrer notre méthode, nous prenons l'exemple de l'obésité chez les enfants dans la National Health and Nutrition Examination Survey dans laquelle les petits domaines sont formés par croisement de l'école (cycle secondaire inférieur ou supérieur), de l'ethnicité (blanche, noire, mexicaine) et du sexe (masculin, féminin). Nous procédons à une étude en simulation pour évaluer certaines caractéristiques statistiques de notre méthode. Nous avons montré que le gain de précision au-delà de (a) est dans l'ordre où (b) est plus grand que (c).

Mots clés : Algorithme d'acceptation-rejet ; loi binomiale ; loi bêta généralisée ; échantillonneur de Gibbs « griddy » ; simulation.

## 1. Introduction

L'utilisation de modèles pour « emprunter de l'information » dans l'estimation sur petits domaines est une pratique standard (Rao 2003). Étant donné le peu de données présentes dans chaque domaine, les estimations directes pour les petits domaines ne sont habituellement pas fiables. Notre procédure permet à un praticien d'intégrer l'information a priori au sujet d'une combinaison linéaire de probabilités binomiales, une pour chaque domaine. Il s'agit d'une contrainte que nous incluons sous forme de moyenne pondérée des probabilités de domaine dans le modèle bêta-binomial standard. La moyenne pondérée peut être supposée connue ou inconnue. Dans le cas où cette valeur est inconnue, nous considérons le scénario où il existe une certaine information qui peut être obtenue auprès d'un expert sous forme de loi a posteriori. Cette situation diffère de la pratique courante en échantillonnage fondé sur le plan de sondage dans lequel l'information auxiliaire est intégrée, comme dans les estimateurs par le ratio et par la régression (Cochran 1977). Quand la valeur peut être spécifiée exactement, la précision augmente, parce que l'information a priori est intégrée dans le domaine.

Le modèle bêta-binomial a été étudié extensivement. Par exemple, Nandram et Sedransk (1993), Nandram (1998) et Nandram et Choi (2002) montrent comment faire une inférence prédictive bayésienne des proportions de petit domaine en population finie pour des données binomiales et multinomiales. Ces modèles supposent que les probabilités binomiales ont un effet en commun, ce qui permet le groupement adaptatif des données provenant des petits domaines (ou grappes). Cependant, il est possible d'améliorer encore davantage ces modèles en incluant de l'information supplémentaire en se servant de covariables par la voie de modèles linéaires généralisés (par exemple, voir Ghosh, Natarajan, Stroud et Carlin 1998). Il convient de mentionner que dans aucun de ces travaux il n'est proposé des moyens d'intégrer l'information a priori au sujet de la combinaison linéaire de paramètres du modèle. Des gains de précision importants sont attendus quand ce genre d'information a priori est intégré dans les modèles pour petits domaines ; voir Silvapulle et Sen (2006) pour une longue discussion de l'inférence statistique sous contrainte. Il convient aussi de souligner que Lazar, Meeden et Nelson (2008) ont montré comment inclure des contraintes dans une approche bayésienne non paramétrique au moyen d'un schéma d'urne de

1. Balgobin Nandram et Hasanjan Sayit, Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA 01609-2280. Courriel : balnan@wpi.edu, hs7@wpi.edu.

Polyá à une loi prédictive des paramètres d'une population finie.

Notre procédure est reliée à l'étalonnage externe qui a lieu quand un estimateur préspecifié est obtenu de sources externes, telles qu'une enquête différente, un recensement ou d'autres dossiers administratifs. Dans l'étalonnage, on veut que la somme des parties soit égale au tout. Par exemple, quand des enquêtes sont réalisées au cours du temps, il existe habituellement des enquêtes mensuelles et des enquêtes annuelles qui sont de nettement meilleure qualité que les enquêtes mensuelles. Quand des estimations sont produites d'après les enquêtes mensuelles de telle façon que la somme de ces estimations concorde avec les totaux des enquêtes annuelles, il existe une protection contre l'échec du modèle et, par conséquent, une amélioration des estimations (c'est-à-dire biais réduit et éventuellement augmentation de la précision). Ces problèmes sont fréquents dans les organismes gouvernementaux, spécialement en ce qui concerne l'emploi et les ventes ; voir Hillmer et Trabelsi (1987) pour un exemple portant sur les ventes au détail des quincailleries provenant du U.S. Census Bureau.

L'information a priori issue de l'étalonnage externe améliore la précision, mais peut également produire des estimateurs gravement biaisés. Cela dépendra de la mesure dans laquelle l'enquête courante diffère des précédentes. Nandram, Toto et Choi (2011) ont appliqué l'étalonnage externe pour estimer les valeurs moyennes de petits domaines en population finie. La contrainte est la moyenne de population pour l'ensemble de la population, qui est une valeur préspecifiée pouvant, de nouveau, être tirée d'une enquête antérieure, d'un recensement ou de dossiers administratifs. Dans nos travaux courants, nous n'intégrons pas l'information au sujet d'une combinaison linéaire des valeurs de population finie, mais nous incorporons plutôt l'information au sujet d'une combinaison linéaire des paramètres de superpopulation (ici les probabilités binomiales).

Nous considérons le problème dans lequel les dénombrements binomiaux sont obtenus auprès de petits domaines similaires, et où l'inférence est requise au sujet des probabilités binomiales. Dans la conclusion, nous discutons de la façon d'étendre notre méthode pour obtenir la loi prédictive des proportions de population finie. Le modèle bêta-binomial standard peut s'avérer inadéquat et l'information a priori supplémentaire peut être incorporée. Nous postulons qu'il y a une augmentation de la précision par rapport au modèle bêta-binomial standard pour petits domaines quand l'information a priori au sujet de la moyenne pondérée des probabilités (par exemple, moyenne des probabilités) est incorporée. Autrement dit, nous intégrons l'information a priori au sujet d'une combinaison linéaire de probabilités binomiales (une moyenne pondérée). Les poids peuvent être proportionnels aux tailles de population et, sous une

répartition proportionnelle, ils peuvent être proportionnels aux tailles d'échantillon. Le but de l'intégration d'information a priori au sujet des probabilités binomiales est d'accroître la précision et, en même temps, il faut contrôler le biais.

Il est nettement plus facile pour un praticien d'enquête de spécifier la valeur de la probabilité globale que celle des probabilités des domaines individuels. Autrement dit, la probabilité globale peut être spécifiée avec relativement moins d'erreur que les probabilités individuelles. Naturellement, on peut spécifier la probabilité globale en utilisant de l'information a priori (une enquête antérieure, un recensement ou des dossiers administratifs) et la spécification de la probabilité globale dépendra donc de la qualité de l'information a priori. D'où le problème relève naturellement du paradigme bayésien, parce que nous intégrons l'information a priori au sujet d'un paramètre au moyen d'une distribution. Donc, il y aura un gain de précision à cause de l'information supplémentaire. Cependant, un praticien peut encore procéder en l'absence d'information a priori. On peut utiliser le ratio du succès total et de la taille totale d'échantillon sur les domaines pour former une spécification raisonnable de la probabilité globale qui n'est habituellement pas la probabilité d'intérêt. Cette estimation aura une beaucoup plus grande précision que celle pour les domaines individuels. Il y aura encore un gain de précision, mais de toute évidence, ce gain est dû à l'utilisation de données courantes (utilisation double) et de la contrainte.

Un exemple d'enquête dans laquelle de l'information fiable peut être obtenue pour procéder à l'étalonnage est la National Health Interview Survey (NHIS) qui est réalisée annuellement par le National Center for Health Statistics pour évaluer un aspect de la santé de la population américaine. Il s'agit d'une enquête représentative de la population qui porte sur de nombreux indicateurs de la santé ; l'un de ces indicateurs est le nombre de visites chez le médecin au cours des deux dernières semaines, et une entité informative est la proportion de personnes qui ont rendu au moins une visite à un médecin l'année précédente (par exemple, Nandram et Choi 2002). Ces proportions sont utiles pour les petits domaines formés par croisement de l'âge, de la race et du sexe pour un État particulier l'année précédente. Comme les estimations à l'échelle d'un État varient fort peu d'une année à l'autre, l'estimation globale pour l'année qui précède immédiatement la dernière année peut être utilisée comme valeur de référence fiable pour la dernière année. S'il n'est pas possible d'obtenir une estimation fiable pour l'étalonnage, on peut construire une loi a priori informative. Par exemple, on peut utiliser la méthode des moments pour égaliser la moyenne d'échantillon et la variance d'échantillon des estimations globales pour les quelques années passées à la moyenne et à la variance d'une loi bêta pour obtenir une

loi a posteriori bêta. Dans l'un et l'autre cas, notre procédure peut être appliquée.

Le plan du présent article est le suivant. À la section 2, nous décrivons la méthodologie. En particulier, nous décrivons le modèle bêta-binomial standard et nous élaborons deux autres modèles pour intégrer l'information supplémentaire en utilisant les lois a priori. Nous décrivons aussi l'inférence a posteriori et la façon d'effectuer les calculs non standard. À la section 3, nous décrivons un exemple portant sur l'obésité, et une étude en simulation pour évaluer empiriquement les propriétés statistiques de nos modèles. À la section 4, nous présentons nos conclusions. Nous discutons aussi de la façon de faire une inférence prédictive bayésienne pour les proportions de population finie. Bien que nous discutons de données binaires, nous montrons aussi comment notre méthode peut être étendue à des données polychotomiques.

## 2. Méthodologie

Nous montrons comment intégrer la contrainte dans le modèle bêta-binomial de deux façons, ce qui fournit un ensemble de modèles de rechange. À la section 2.1, nous décrivons les modèles et à la section 2.2., nous décrivons l'inférence a posteriori. Nous essayons d'expliquer l'effet de la contrainte sur les estimations des probabilités en utilisant une approximation. À la section 2.3, nous décrivons les calculs et un nouvel algorithme.

### 2.1 Modèles

Nous supposons que des données binaires sont disponibles en provenance de  $\ell$  petits domaines, et nous supposons que la probabilité qu'un individu réponde dans le  $i^{\text{e}}$  domaine est  $\pi_i$ ,  $i = 1, \dots, \ell$ . Soit  $n_i$  le nombre d'individus échantillonnés dans le  $i^{\text{e}}$  domaine,  $i = 1, \dots, \ell$ . Soit aussi  $s_i$  le nombre d'individus possédant la caractéristique et  $f_i = n_i - s_i$  le nombre d'individus sans la caractéristique dans le  $i^{\text{e}}$  domaine,  $i = 1, 2, \dots, \ell$ . Alors, le modèle hiérarchique bayésien bêta-binomial standard est

$$s_i | \pi_i \sim \overset{\text{ind}}{\text{Binomiale}}(n_i, \pi_i), \quad (1)$$

$$\pi_i | \mu, \tau \sim \overset{\text{iid}}{\text{Bêta}}\{\mu\tau, (1 - \mu)\tau\}, \quad i = 1, \dots, \ell \quad (2)$$

et

$$p(\mu, \tau) = \frac{1}{(1 + \tau)^2}, \quad 0 < \mu < 1, \tau \geq 0. \quad (3)$$

Nous utilisons un prior de rétrécissement pour  $\tau$ , parce qu'il est approprié et non informatif et qu'il n'y a pas de priors conjugués. Les priors de la forme  $p(\tau) \propto 1/\tau$  sont déconseillés ; voir, par exemple, Gelman (2006). D'autres options sont les demi-densités de Cauchy et les densités

gamma (il serait nécessaire de spécifier les hyper-paramètres). Donc, nous appellerons le modèle spécifié par (1), (2) et (3) le modèle non restreint (NR) ou modèle 1.

Ensuite, nous décrivons le modèle restreint, qui est une extension du modèle non restreint. Nous obtenons une simple combinaison linéaire des probabilités binomiales. En posant que  $\tilde{\pi}_i = s_i/n_i$  et

$$\omega_i = \frac{n_i}{\sum_{i=1}^{\ell} n_i}, \quad i = 1, \dots, \ell,$$

nous avons

$$\frac{\sum_{i=1}^{\ell} s_i}{\sum_{i=1}^{\ell} n_i} = \sum_{i=1}^{\ell} \omega_i \tilde{\pi}_i.$$

Donc, en prenant les  $\pi_i$  inconnues, la combinaison linéaire est  $\sum_{i=1}^{\ell} \omega_i \pi_i$ .

Par conséquent, nous devons faire un ajustement dans (2) afin d'incorporer la contrainte  $\sum_{i=1}^{\ell} \omega_i \pi_i = \theta$  conditionnellement à  $\theta$ . Nous le faisons en introduisant la variable  $\phi = \sum_{i=1}^{\ell} \omega_i \pi_i - \theta$  ; de sorte que la contrainte est équivalente à  $\phi = 0$ . Maintenant, une des variables,  $\pi_i$ ,  $i = 1, \dots, \ell$ , est redondante. Il convient de souligner que l'on peut choisir n'importe laquelle des  $\pi_1, \dots, \pi_{\ell}$ , et, sans perte de généralité, pour faciliter l'exposé, nous choisissons  $\pi_{\ell}$ . Donc, pour intégrer la contrainte, nous transformons  $\pi_{\ell}$  en  $\phi = \sum_{i=1}^{\ell} \omega_i \pi_i - \theta$ , en gardant les  $\pi_1, \dots, \pi_{\ell-1}$  non transformées, et nous posons que  $\boldsymbol{\pi}_{(\ell)} = (\pi_1, \dots, \pi_{\ell-1})'$ .

Comme le jacobien est  $1/\omega_{\ell}$ ,

$$p(\boldsymbol{\pi}_{(\ell)}, \phi | \mu, \tau, \theta) =$$

$$\frac{1}{\omega_{\ell}} \prod_{i=1}^{\ell-1} \frac{\pi_i^{\mu\tau-1} (1 - \pi_i)^{(1-\mu)\tau-1}}{B\{\mu\tau, (1 - \mu)\tau\}} \times \frac{\left[ \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}} \right]^{\mu\tau-1} \left[ 1 - \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}} \right]^{(1-\mu)\tau-1}}{B\{\mu\tau, (1 - \mu)\tau\}}, \quad (4)$$

où

$$0 < \pi_i < 1, \quad i = 1, \dots, \ell,$$

$$0 < \mu < 1, \tau > 0, \phi + \theta - \omega_{\ell} \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \phi + \theta,$$

et

$$\pi_{\ell} = \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}}. \quad (5)$$

Notons que la densité a priori conjointe de  $(\boldsymbol{\pi}_{(\ell)}, \phi)$  dans (4) est bien définie. Nous souhaitons prendre  $\phi = 0$  dans (5) pour intégrer la contrainte, mais quand  $\phi = 0$ , la densité conjointe de  $\boldsymbol{\pi}_{(\ell)}$  n'est pas bien définie.

Nous supposons que  $\mu, \tau, \theta$  sont indépendants a priori avec  $p(\mu, \tau, \theta) = p_1(\mu, \tau) p_2(\theta)$ , où

$$p_1(\mu, \tau) = \frac{1}{(1 + \tau)^2}, \quad 0 < \mu < 1, \quad \tau \geq 0$$

comme dans (3), et  $p_2(\theta)$  est donné par

$$\theta \sim \text{B\^eta} \{\mu_0 \tau_0, (1 - \mu_0) \tau_0\}. \tag{6}$$

Pour le modèle restreint, nous considérons deux scénarios. En laissant  $\tau_0 \rightarrow \infty$ ,  $\theta$  devient un point matériel à  $\mu_0$ , et dans ce cas  $\theta = \mu_0$  doit être spécifié par un praticien ; nous appellerons le modèle ajusté le modèle fixe (FI) ou modèle 2. Nous avons un deuxième scénario dans lequel un praticien spécifie  $\mu_0$  et  $\tau_0$  mais non  $\theta$  ; nous appellerons ce modèle ajusté le modèle informatif (IN) ou modèle 3. Donc, il y a trois modèles, y compris le modèle non restreint. Afin de fournir un cadre unifié, il faut que tous les priors soient appropriés. La valeur exacte de  $\theta$  est vraisemblablement inconnue dans la plupart des applications, ce qui peut donner lieu à des estimations n'ayant pas de cohérence interne.

Il convient de souligner que nous avons considéré un modèle supplémentaire pour faciliter l'étude du gain de précision de IN comparativement à FI. Pour les besoins de comparaison, nous voulons imposer un prior approprié mais non informatif à  $\theta$ , de sorte que  $\theta \sim \text{Uniforme}(0, 1)$  n'est pas un choix déraisonnable. En posant que  $\mu_0 = 1/2$ ,  $\tau_0 = 2$ , nous obtenons  $\theta \sim \text{Uniforme}(0, 1)$  avec ce prior, et nous appellerons le modèle ajusté le modèle uniforme (UN) ou modèle 4 ; naturellement, nous ne devons pas spécifier  $\mu_0$  et  $\tau_0$ . Il convient de souligner que le prior correspondant à  $\tau \rightarrow \infty$  est inapproprié car il correspond à  $\theta \sim \text{B\^eta}(0, 0)$ . Nous ne considérons pas davantage ce modèle ; cependant, même si UN n'a pas de contrainte, nous le considérerons brièvement tout au long de l'exposé.

### 2.2 Inférence a posteriori

Nous envisageons de faire une inférence a posteriori au sujet de  $\pi_i, i = 1, \dots, \ell$ . Soit  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_\ell)'$  et  $\boldsymbol{\pi}_{(i)} = (\pi_1, \dots, \pi_{i-1}, \pi_{i+1}, \dots, \pi_\ell)'$  [par exemple,  $\boldsymbol{\pi}_{(\ell)} = (\pi_1, \dots, \pi_{\ell-1})'$  tel que défini plus haut].

Nous utilisons le théorème de Bayes pour trouver les densités a posteriori conjointes de tous les paramètres. Premièrement, sous le modèle non restreint spécifié par (1), (2) et (3), la densité a posteriori conjointe de  $\boldsymbol{\pi}, \mu, \tau$  est

$$g(\boldsymbol{\pi}, \mu, \tau | s) \propto \prod_{i=1}^{\ell} \frac{\pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1-\mu)\tau - 1}}{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}} \times \prod_{i=1}^{\ell} \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \times \frac{1}{(1 + \tau)^2}, \tag{7}$$

$$0 < \pi_i < 1, 0 < \mu < 1, \tau > 0, i = 1, \dots, \ell.$$

*Lemme 1* Sous le modèle non restreint la densité a posteriori conjointe,  $g(\boldsymbol{\pi}, \mu, \tau | s)$ , est appropriée.

Une preuve du lemme 1 est donnée à l'annexe A.

Sous le modèle restreint, la densité a posteriori conjointe de  $\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta, \phi$  est

$$p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta, \phi | s) \propto \prod_{i=1}^{\ell-1} \frac{\pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1-\mu)\tau - 1}}{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}} \times \frac{\left[ \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{s_\ell + \mu\tau - 1} \left[ \frac{\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{f_\ell + (1-\mu)\tau - 1}}{B\{s_\ell + \mu\tau, f_\ell + (1 - \mu)\tau\}} \times \prod_{i=1}^{\ell} \left[ \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \right] \times \theta^{\mu_0 \tau_0 - 1} (1 - \theta)^{(1 - \mu_0) \tau_0 - 1} \times \frac{1}{(1 + \tau)^2}, \tag{8}$$

$$0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0, \phi + \theta - \omega_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \phi + \theta, 0 < \theta < 1. \text{ Notons que } \pi_\ell = (\phi + \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell.$$

Nous obtenons la densité a posteriori conjointe pertinente en intégrant la contrainte ( $\phi = 0$ ) dans (8). C'est-à-dire,  $p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta | s, \phi = 0) \propto p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta, \phi = 0 | s)$ , où

$$p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta | s, \phi = 0) \propto \prod_{i=1}^{\ell-1} \frac{\pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1-\mu)\tau - 1}}{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}} \times \frac{\left[ \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{s_\ell + \mu\tau - 1} \left[ \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right]^{f_\ell + (1-\mu)\tau - 1}}{B\{s_\ell + \mu\tau, f_\ell + (1 - \mu)\tau\}} \times \prod_{i=1}^{\ell} \left[ \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \right] \times \theta^{\mu_0 \tau_0 - 1} (1 - \theta)^{(1 - \mu_0) \tau_0 - 1} \times \frac{1}{(1 + \tau)^2}, \tag{9}$$

$0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0, \theta - \omega_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \theta, 0 < \theta < 1$ . Notons de nouveau que  $\pi_\ell = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell$ . Il convient de souligner que la densité a posteriori conjointe (9) intègre la contrainte,  $\sum_{i=1}^{\ell} \omega_i \pi_i = \theta$ ,

exactement parce que  $\pi_\ell = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell$ ,  $\theta - \omega_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \theta$ . Autrement dit, la densité a posteriori conjointe n'est pas une fonction de  $\pi_\ell$ , et l'inférence a posteriori au sujet de  $\pi_\ell = (\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell$  découle de l'identité. Donc, il n'y a absolument aucune différence entre  $\theta$  et  $\sum_{i=1}^{\ell} \omega_i \pi_i$ .

*Théorème 1* Sous le modèle resreint, la densité a posteriori conjointe,  $p(\boldsymbol{\pi}_{(\ell)}, \mu, \tau, \theta | s, \phi = 0)$ , est appropriée.

Une preuve du théorème 1 est donnée à l'annexe A.

Nous notons la différence entre les densités pour le modèle non restreint donné par (7) et le modèle restreint donné par (9). Essentiellement, le terme

$$\left( \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right)^{s_\ell + \mu\tau - 1} \times \left( 1 - \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell} \right)^{f_\ell + (1-\mu)\tau - 1} \\ \times \theta^{\mu_0\tau_0 - 1} (1 - \theta)^{(1-\mu_0)\tau_0 - 1}$$

dans (9) remplace  $\pi_\ell^{s_\ell + \mu\tau - 1} (1 - \pi_\ell)^{f_\ell + (1-\mu)\tau - 1}$  dans (7). Notons que, dans (9),

$$\pi_\ell = \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_\ell}.$$

Soit  $a_i = s_i + \mu\tau$ ,  $b_i = f_i + (1 - \mu)\tau$ ,  $i = 1, \dots, \ell$ . En outre, soit

$$c_i = \frac{\theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j - \omega_\ell}{\omega_i}$$

et

$$d_i = \frac{\theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j}{\omega_i}, \quad i = 1, \dots, \ell - 1.$$

Alors,

$$p(\pi_i | \boldsymbol{\pi}_{(i)}, \mu, \tau, \theta, s, \phi = 0) \\ \propto \pi_i^{a_i - 1} (1 - \pi_i)^{b_i - 1} (\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}, \quad (10)$$

$c_i < \pi_i < d_i$ ,  $i = 1, \dots, \ell - 1$ . Notons que cette fonction de densité comprend deux termes  $\pi_i^{a_i - 1} (1 - \pi_i)^{b_i - 1}$  et  $(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}$ ; notons l'échange entre  $a_\ell$  et  $b_\ell$  dans le deuxième terme. Le premier terme est la densité a posteriori conditionnelle sous le modèle non restreint, et le deuxième terme est une densité bêta généralisée [c'est-à-dire une loi bêta ( $b_\ell, a_\ell$ ) dans l'intervalle  $(c_i, d_i)$ ]. Donc, la densité bêta non restreinte est ajustée par la densité bêta généralisée. Dans le reste du document, nous désignons par  $\text{GenBêta}(a, b, c, d)$  la variable aléatoire bêta généralisée avec la fonction de densité,

$$p(x) = (x - c)^{a-1} (d - x)^{b-1} / \{(d - c)^{a+b-1} B(a, b)\}, \\ c \leq x \leq d, a > 1, b > 1.$$

Autrement dit,  $(X - c) / (d - c) \sim \text{Bêta}(a, b)$  si et uniquement si  $X \sim \text{GenBêta}(a, b, c, d)$ .

Il convient de mentionner que nous avons placé les domaines par ordre de dénombrement (du plus petit au plus grand). Cela est commode et avantageux tant du point de vue théorique que des calculs.

Afin d'expliquer le gain de précision, nous essayons d'étudier (10) plus en profondeur en faisant deux approximations. Premièrement, comme la contrainte étudiée est assez faible, nous ne nous attendons pas à ce que  $c_i$  soit fort différent de 0 et que  $d_i$  soit fort différent de 1. Sous cette hypothèse, nous pouvons approximer (10) par

$$p_a(\pi_i | \boldsymbol{\pi}_{(i)}, \mu, \tau, \theta, s, \phi = 0) \\ \propto (\pi_i - c_i)^{a_i - 1} (d_i - \pi_i)^{b_i - 1} (\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}, \\ c_i < \pi_i < d_i.$$

Alors, en intégrant la constante de normalisation dans

$p_a(\pi_i | \boldsymbol{\pi}_{(i)}, \mu, \tau, \theta, s, \phi = 0)$ , nous avons

$$p_a(\pi_i | \boldsymbol{\pi}_{(i)}, \mu, \tau, \theta, s, \phi = 0) \\ = \frac{(\pi_i - c_i)^{a_i - 1} (d_i - \pi_i)^{b_i - 1} (\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}}{\int_{c_i}^{d_i} (\pi_i - c_i)^{a_i - 1} (d_i - \pi_i)^{b_i - 1} (\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1} d\pi_i} \\ = \frac{(\pi_i - c_i)^{a_i - 1} (d_i - \pi_i)^{b_i - 1}}{(d_i - c_i)^{a_i + b_i - 1} B(a_i, b_i)} \\ \times \frac{(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}}{E[(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}]}, \quad c_i < \pi_i < d_i, \quad (11)$$

où l'espérance est prise sur la distribution bêta généralisée  $\pi_i \sim \text{GenBêta}(a_i, b_i, c_i, d_i)$ ,  $i = 1, \dots, \ell - 1$ . Mais sous cette dernière densité,  $(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}$  est un estimateur sans biais de  $E[(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}]$ . En outre, par construction  $a_\ell$  et  $b_\ell$  sont relativement grands et, par conséquent, nous nous attendons à ce que  $(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}$  et sa variance soient petits. Alors, notre seconde approximation est

$$(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1} \approx E[(\pi_i - c_i)^{b_\ell - 1} (d_i - \pi_i)^{a_\ell - 1}]. \quad (12)$$

Donc, en combinant (11) et (12), notre approximation finale de (10) est

$$\pi_i | \boldsymbol{\pi}_{(i)}, \mu, \tau, \theta, s, \phi = 0 \sim \text{GenBêta}(a_i, b_i, c_i, d_i). \quad (13)$$

Il découle de (13) que

$$E_r(\pi_i | \boldsymbol{\pi}_{(i)}, \mu, \tau, \theta, s, \phi = 0) \approx c_i + (d_i - c_i) E_u(\pi_i | \mu, \tau, s)$$

et

$$\text{Var}_r(\pi_i | \boldsymbol{\pi}_{(i)}, \mu, \tau, \theta, \mathbf{s}, \phi = 0) \approx (d_i - c_i)^2 \text{Var}_u(\pi_i | \mu, \tau, \mathbf{s}), \quad (14)$$

où  $u$  fait référence au modèle non restreint et  $r$  au modèle restreint. Notons que, quand  $c_i = 0$  et  $d_i = 1$ , nous obtenons  $E_r(\pi_i | \cdot) = E_u(\pi_i | \cdot)$  et  $\text{Var}_r(\pi_i | \cdot) = \text{Var}_u(\pi_i | \cdot)$ . Toutefois, en général, les estimations de  $\pi_i$  seront légèrement différentes d'un scénario à l'autre. Il est également intéressant que  $\text{Var}_r(\pi_i | \cdot) \leq \text{Var}_u(\pi_i | \cdot)$  du moins approximativement. Donc, la contrainte  $\sum_{i=1}^{\ell} \omega_i \pi_i = \theta$  réduira la variabilité, quand les  $\pi_i$  sont estimées. Cela est vrai parce que les  $\pi_i, i = 1, \dots, \ell$ , appartiennent à un simplexe  $\ell - 1$ -dimensionnel dans l'hypercube  $\ell$ -dimensionnel, tandis que pour le modèle non restreint,  $\pi_i, i = 1, \dots, \ell$ , appartient à l'hypercube  $\ell$ -dimensionnel. Nous prévoyons le gain de précision le plus important quand  $\theta$  est entièrement spécifié, suivi par le cas où  $\mu_0$  est spécifié et  $\tau_0 \gg 2$ , et le gain de précision le plus faible quand  $\theta \sim \text{Uniforme}(0, 1)$ .

### 2.3 Calculs

Nous montrons comment tirer des échantillons des modèles non restreints et restreints. Pour le modèle non restreint, nous sommes capables de tirer des échantillons aléatoires de (7) sans utiliser de méthode Monte Carlo par chaîne de Markov. Cependant, pour le modèle restreint, nous utilisons l'échantillonneur de Gibbs « griddy » (Ritter et Tanner 1992) pour tirer les échantillons de (9).

#### 2.3.1 Modèle non restreint

Nous groupons sur les  $\pi_i$ , et tirons les échantillons à partir de  $p(\mu, \tau | \mathbf{s})$  en utilisant des tirages aléatoires à partir d'une grille bivariable et finalement nous obtenons les échantillons à partir des densités Rao-Blackwellisées  $\pi_i | \mu, \tau, \mathbf{s}$ .

Alors,

$$\pi_i | \mu, \tau, \mathbf{s} \stackrel{\text{ind}}{\sim} \text{Bêta}\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}, i = 1, \dots, \ell, \quad (15)$$

Et, en éliminant  $\boldsymbol{\pi}$ , par intégration, nous obtenons

$$p(\mu, \tau | \mathbf{s}) \propto \prod_{i=1}^{\ell} \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \times \frac{1}{(1 + \tau)^2},$$

$0 < \mu < 1, \tau > 0$ . En posant que  $\delta = \tau / \tau + 1$ , nous avons

$$p(\mu, \delta | \mathbf{s}) \propto \left[ \prod_{i=1}^{\ell} \frac{B\{s_i + \mu\tau, f_i + (1 - \mu)\tau\}}{B\{\mu\tau, (1 - \mu)\tau\}} \right]_{\tau = \frac{\delta}{1-\delta}}, \quad 0 < \mu, \delta < 1.$$

Pour commencer, nous tirons  $\mu, \delta | \mathbf{s}$  en utilisant une grille bivariable sur  $(0, 1)^2$  pour obtenir un échantillon de  $M \approx 10,000$  valeurs de  $(\mu^{(h)}, \delta^{(h)})$ ,  $h = 1, \dots, M$ ,  $\tau^{(h)} = \delta^{(h)} / (1 - \delta^{(h)})$ . Ensuite, nous procédons à une augmentation des données en (15) pour obtenir  $\pi^{(h)}, h = 1, 2, \dots, M$ , en utilisant une méthode de composition. Autrement dit, nous tirons simplement  $\pi_i \sim \text{Bêta}\{s_i + \mu^{(h)}\tau^{(h)}, f_i + (1 - \mu^{(h)})\tau^{(h)}\}, i = 1, \dots, \ell, h = 1, \dots, M$ .

Pour exécuter la méthode de la grille bivariable pour échantillonner à partir de la densité a posteriori de  $(\mu, \delta)$ , nous divisons l'intervalle  $(0, 1)$  en 100 sous-intervalles ; donc il ya 10 000 petits carrés dans le carré unique original. Nous obtenons les hauteurs de la densité a posteriori (sans la constante de normalisation) au centre de chacun des 10 000 carrés. Comme ces petits carrés ont la même aire, les hauteurs de la densité bivariable sont proportionnelles aux probabilités a posteriori que  $(\mu, \delta)$  se trouve dans chacun des carrés. Donc, nous avons construit une fonction de masse de probabilité a posteriori conjointe de  $(\mu, \delta)$  sur des grilles très fines. Il est facile de tirer un échantillon de la fonction de masse de probabilité bivariable discrète en utilisant la méthode de distribution cumulative. Il s'agit en fait d'un tirage aléatoire de l'un des 10 000 carrés avec probabilités proportionnelles aux hauteurs des petits carrés. Puis, dans le carré sélectionné, nous choisissons un point au hasard en tirant deux variables aléatoires uniformes (c'est-à-dire *uniform random jittering*). En effet, il s'agit d'un tirage aléatoire très précis à partir de la densité a posteriori conjointe de  $(\mu, \delta)$ . Nous tirons  $M = 10\,000$  échantillons de cette approximation pour l'inférence a posteriori dans une procédure Monte Carlo standard avec échantillons indépendants et non une chaîne de Markov. Étant donné la technique de *random jittering*, les nombres sont différents avec une probabilité égale à l'unité.

#### 2.3.2 Modèle restreint

Nous montrons comment tirer des échantillons du modèle restreint en utilisant l'échantillonneur de Gibbs. La densité a posteriori conditionnelle conjointe de  $\pi_1, \dots, \pi_{\ell-1}$  est

$$p(\pi_1, \dots, \pi_{\ell-1} | \mu, \tau, \theta, \mathbf{s}, \phi = 0) \propto \prod_{i=1}^{\ell-1} \left\{ \pi_i^{s_i + \mu\tau - 1} (1 - \pi_i)^{f_i + (1 - \mu)\tau - 1} \right\} \times \left( \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i \right)^{s_{\ell} + \mu\tau - 1} \left\{ \sum_{i=1}^{\ell-1} \omega_i \pi_i - \theta + \omega_{\ell} \right\}^{f_{\ell} + (1 - \mu)\tau - 1} \quad (16)$$

où

$$\theta < \omega_{\ell}, \theta - \omega_{\ell} < \sum_{i=1}^{\ell-1} \omega_i \pi_i < \theta, \pi_{\ell} = \frac{\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i}{\omega_{\ell}}.$$

Donc, nous obtiendrions les échantillons de  $\pi_1, \dots, \pi_{\ell-1}$  et nous posons

$$\pi_\ell = \frac{\left(\theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i\right)}{\omega_\ell}$$

pour compléter le vecteur  $\pi_1, \dots, \pi_\ell$ . Autrement dit, la contrainte est obtenue exactement. La densité a posteriori conditionnelle de  $\theta$  est

$$p(\theta \mid \boldsymbol{\pi}_{(\ell)}, \mu, \tau, s, \phi = 0)$$

$$\propto \left\{ \theta - \sum_{i=1}^{\ell-1} \omega_i \pi_i \right\}^{s_\ell + \mu\tau - 1} \left\{ \omega_\ell + \sum_{i=1}^{\ell-1} \omega_i \pi_i - \theta \right\}^{f_\ell + (1-\mu)\tau - 1} \\ \times \theta^{\mu_0\tau_0 - 1} (1 - \theta)^{(1-\mu_0)\tau_0 - 1}, \quad (17)$$

où

$$\sum_{i=1}^{\ell-1} \omega_i \pi_i < \theta < \omega_\ell + \sum_{i=1}^{\ell-1} \omega_i \pi_i.$$

La densité a posteriori conditionnelle conjointe de  $\mu$  et  $\tau$  est

$$p(\mu, \tau \mid \boldsymbol{\pi}_{(\ell)}, \theta, s, \phi = 0)$$

$$\propto \frac{q^{\mu\tau} r^{(1-\mu)\tau}}{[B(\mu\tau, (1-\mu)\tau)]^\ell} \times \frac{1}{(1+\tau)^2}, \quad (18)$$

$$0 < \mu < 1, \tau > 0, q = \prod_{i=1}^{\ell} \pi_i, r = \prod_{i=1}^{\ell} (1 - \pi_i).$$

Pour exécuter l'échantillonneur de Gibbs, nous devons tirer des échantillons de (16), (17) et (18), tour à tour, jusqu'à la convergence. Nous tirons  $\mu, \tau$  de  $p(\mu, \tau \mid \boldsymbol{\pi}_{(\ell)}, \theta, s)$  de manière semblable au tirage de  $p(\mu, \tau \mid \boldsymbol{\pi}_{(\ell)})$  dans le modèle non restreint. Il est plus difficile de tirer un échantillon de (16) et (17). Cependant, nous utilisons essentiellement la même méthode pour tirer les échantillons de la densité a posteriori conditionnelle de  $\pi_i, i = 1, \dots, \ell - 1$ , obtenue de (16) et  $\theta$  à partir de (17) qui sont toutes deux proportionnelles au produit de deux fonctions de densité, l'une étant une densité bêta tronquée et l'autre, une densité bêta généralisée. Ensuite, nous élaborons une certaine théorie pour tirer un échantillon d'une telle densité. Pour cela, nous énonçons et prouvons le lemme 2 et le théorème 2.

La fonction de densité d'intérêt est

$$f(x) = A f_1(x) f_2(x), 0 \leq c < x < d \leq 1, \quad (19)$$

où

$$f_1(x) = \frac{x^{g-1}(1-x)^{h-1}}{\int_c^d x^{g-1}(1-x)^{h-1} dx}, c < x < d, g, h > 0, \quad (20)$$

$$f_2(x) = (x-c)^{a-1}(d-x)^{b-1} / \{(d-c)^{a+b-1} B(a, b)\}, \\ c < x < d, a, b > 1, \quad (21)$$

et, naturellement,

$$A = 1 / \int_c^d f_1(x) f_2(x) dx. \quad (22)$$

Il convient de souligner que nous ne supposons pas que  $g, h > 1$ . Si cela était le cas,  $f_1(x)$  et  $f_2(x)$  seraient toutes deux logconcave, ce qui rendrait  $f(x)$  logconcave, et, dans ce cas, on peut tirer un échantillon de  $f(x)$  en utilisant l'échantillonneur par rejet adaptatif (ARS, Gilks et Wild 1992). Nous fournissons un algorithme spécialisé pour tirer un échantillon de  $f(x)$  qui n'est pas logconcave. Même si  $f_1(x)$  était logconcave (c'est-à-dire  $g, h > 1$ ) cet algorithme spécialisé serait encore supérieur à l'échantillonneur ARS, parce que ce dernier est un algorithme d'usage général; voir Robert et Casella (1999, page 59). Notre algorithme requiert moins de calcul et ne nécessite pas la logconcavité; même en présence de logconcavité, l'échantillonneur ARS peut donner de mauvais résultats dans les queues de la fonction de densité.

*Lemme 2* Considérons les fonctions de densité  $f_1(x)$  et  $f_2(x)$  avec  $a, b > 1$ .

(a) Alors

$$\sup_{c < x < d} f_2(x) = \frac{\delta^{a-1}(1-\delta)^{b-1}}{(d-c)B(a, b)}, \delta = (a-1) / (a+b-2).$$

(b) Pour tout  $g > 0, h > 0$ , il existe deux constantes  $H_1$  et  $H_2$  telles que

$$0 < H_1 \leq A^{-1} \leq H_2 < \infty.$$

Une preuve du lemme 2 est donné à l'annexe A.

*Théorème 2* Soit  $F_{g,h}(\cdot)$  la fonction de répartition de la variable aléatoire Bêta( $g, h$ ) et  $F_{g,h}^{-1}(\cdot)$ , son inverse. Soit

$$U, V \stackrel{\text{ind}}{\sim} \text{Uniforme}(0, 1),$$

et soit

$$X = F_{g,h}^{-1}\{UF_{g,h}(d) + (1-U)F_{g,h}(c)\}.$$

Si pour deux nombres réels  $a, b > 1$ ,

$$V \leq \frac{1}{(d-c)^{a+b-2}} \left(\frac{X-c}{\delta}\right)^{a-1} \left(\frac{d-X}{1-\delta}\right)^{b-1},$$

où  $\delta = (a-1) / (a+b-2)$ , alors  $X$  a la densité  $f(x) = A f_1(x) f_2(x)$ .

Une preuve du théorème 2 est donnée à l'annexe A.

Le théorème 1 nous donne l'algorithme suivant pour le tirage des échantillons de  $f(\pi) \propto \pi^{g-1}(1-\pi)^{h-1}(\pi-c)^{a-1}(d-\pi)^{b-1}, c < \pi < d, g, h > 0, a, b > 1$ .



*Algorithme*(a) Tirer  $U \sim \text{Uniforme}(0, 1)$  et poser

$$\pi = F_{g,h}^{-1}\{UF_{g,h}(d) + (1-U)F_{g,h}(c)\}.$$

(b) Tirer  $V \sim \text{Uniforme}(0, 1)$ . Si

$$V \leq \frac{1}{(d-c)^{a+b-2}} \left( \frac{\pi-c}{\delta} \right)^{a-1} \left( \frac{d-\pi}{1-\delta} \right)^{b-1},$$

accepter  $\pi$ , sinon aller à (a).

Comme les tailles d'échantillon binomial sont arrangées par ordre croissant, dans toute application il sera vrai que  $a, b > 1$  et  $g, h > 0$  (éventuellement plus grand que 1). Donc, l'algorithme fonctionnera. En effet, dans tous nos exemples (un présenté ici) et exercices de simulation, l'algorithme s'exécute très rapidement.

Maintenant, nous montrons comment tirer  $\pi_i, i = 1, \dots, \ell$ , et  $\theta$ . Pour  $\pi_i$ ,

$$p(\pi_i | \boldsymbol{\pi}_{(i,\ell)}, \theta, \mu, \tau, s, \phi = 0)$$

$$\propto \pi_i^{a_i-1} (1-\pi_i)^{b_i-1} (\pi_i - c_i)^{b_i-1} (d_i - \pi_i)^{a_i-1}, c_i < \pi_i < d_i,$$

où  $\boldsymbol{\pi}_{(i,\ell)}$  est le vecteur contenant les éléments de  $\boldsymbol{\pi}$ , sauf pour  $\pi_i$  et  $\pi_\ell$ , et  $a_i = s_i + \mu\tau$ ,  $b_i = f_i + (1-\mu)\tau$ ,  $i = 1, \dots, \ell$ ,

$$c_i = \left( \theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j - \omega_\ell \right) / \omega_i,$$

$$d_i = \left( \theta - \sum_{j=1, j \neq i}^{\ell-1} \omega_j \pi_j \right) / \omega_i, i = 1, \dots, \ell - 1.$$

Appliquer le théorème à  $p(\pi_i | \boldsymbol{\pi}_{(i)}, \theta, \mu, \tau, s)$ ,  $a_\ell > 1$ ,  $b_\ell > 1$ ,  $i = 1, \dots, \ell - 1$ .

Pour  $\theta$ , nous avons

$$p(\theta | \boldsymbol{\pi}, \mu, \tau, s, \phi = 0)$$

$$\propto \theta^{\mu_0 \tau_0 - 1} (1-\theta)^{(1-\mu_0) \tau_0 - 1} (\theta - \tilde{c})^{a_\ell - 1} (\tilde{d} - \theta)^{b_\ell - 1}, \tilde{c} < \theta < \tilde{d},$$

où

$$\tilde{c} = \sum_{i=1}^{\ell-1} \omega_i \pi_i, \tilde{d} = \omega_\ell + \sum_{i=1}^{\ell-1} \omega_i \pi_i.$$

De nouveau, appliquer le théorème,  $a_\ell > 1$ ,  $b_\ell > 1$ .

Quand  $\theta$  est entièrement spécifié (c'est-à-dire  $\theta$  n'est pas aléatoire), nous ne devons pas tirer  $\theta$ . Cependant, quand  $\theta \sim \text{Uniforme}(0, 1)$  a priori ( $\mu_0 = 1/2$ ,  $\tau_0 = 2$ ), nous avons une simplification. Dans ce cas,

$$\theta | \boldsymbol{\pi}_{(i)}, \mu, \tau, s, \phi = 0 \sim \text{GenBêta}(a_\ell, b_\ell, \tilde{c}, \tilde{d})$$

et  $\theta = \tilde{c} + (\tilde{d} - \tilde{c})X$ , où  $X \sim \text{Bêta}(a_\ell, b_\ell)$ , a la densité requise.

Pour les modèles restreints ainsi que non restreints, nous utilisons 10 000 itérations pour faire l'inférence a posteriori au sujet des probabilités binomiales,  $\pi_i$ . Sous le modèle non

restreint, il s'agit simplement de tirages aléatoires et aucune surveillance n'est nécessaire. Pour le modèle restreint, en exécutant l'échantillonneur de Gibbs « griddy », nous avons tiré 11 000 itérations, utilisé 1 000 comme « rodage » (un nombre prudent parce que la convergence se produit beaucoup plus tôt que ne le montrent les tracés de courbe), nous n'avons observé que des corrélations négligeables entre les itérations. Donc, nous avons utilisé 10 000 itérations pour faire une inférence au sujet des probabilités binomiales. Pour le modèle non restreint ainsi que pour les trois modèles restreints, il ne faut que quelques secondes sur notre ordinateur alpha  $2 \times 833$  MHz.

### 3. Études numériques

À la section 3.1, nous décrivons un exemple pour illustrer les principales caractéristiques de la contrainte. À la section 3.2, nous décrivons une étude en simulation pour montrer les propriétés fréquentistes des estimateurs de Bayes, et nous donnons un aperçu plus approfondi des différences entre les quatre scénarios. Notons de nouveau que, quand nous exécutons les calculs, il est commode de classer les domaines par taille de façon que le domaine le plus grand soit le dernier.

#### 3.1 Exemple

Nous avons utilisé des données de la troisième National Health and Nutrition Examination Survey (NHANES) pour illustrer notre méthode. Nous avons étudié l'indice de masse corporelle des adolescents et nous avons des données sur l'échantillon obtenu. Les domaines (petits domaines) sont formés par croisement de l'ethnicité (blanche, noire, mexicaine) et du sexe (masculin, féminin). Nous avons classé les adolescents selon qu'ils étaient au secondaire inférieur ou supérieur au moment de l'enquête. Donc, il existe 12 petits domaines. Les données sont présentées dans les quatre premières colonnes du tableau 1 par domaine. Notons que les domaines SIBM, SINP, SIBF et SSNF sont relativement peu peuplés avec 4, 2, 5, 5 adolescents obèses respectivement ; pour les 12 domaines, l'échantillon consiste en 959 adolescents dont 130 sont obèses (c'est-à-dire la proportion globale des individus est de 0,136 environ). Dans la colonne 4 du tableau 1, nous avons présenté les estimations directes par domaine et ces estimations varient de 0,069 à 0,228. Les estimations pour les domaines les plus petits ne sont pas fiables. En outre, quand les modèles bêta-binomiaux sont utilisés, ces estimations régressent vers la moyenne globale d'échantillon de 0,136, ce qui crée un biais éventuel. Notre méthode devrait augmenter la précision au-delà du modèle non restreint, parce que le modèle restreint utilise plus d'information au sujet de la somme pondérée. Clairement, les prédictors basés sur soit le modèle restreint, soit le

modèle bêta-binomial sont biaisés si le modèle spécifié est incorrect.

Nous avons pris  $\mu_0 = 0,136$ , la proportion globale d'échantillon, et  $\tau_0 = 959$ , la taille totale d'échantillon. Des choix moins optimistes peuvent être utilisés. Par exemple,  $\tau_0 = 100$ , disons ; mais ce choix fait peu de différence. Cependant, il convient de souligner qu'utiliser les données observées pour spécifier la distribution a priori peut diminuer artificiellement la variance a posteriori. Habituellement, un praticien des sondages a une spécification appropriée venant d'une enquête antérieure ou d'un recensement. On ne peut pas spécifier des valeurs pour  $\mu_0$  et  $\tau_0$  qui sont entièrement déraisonnables et qui créeront des biais énormes. Ici,  $\tau_0$  est une taille d'échantillon a priori et  $\mu_0$  est une moyenne a priori de  $\theta$ . Cette méthode permet une valeur raisonnable pour  $\theta$  ; nous ajoutons essentiellement un degré d'incertitude au sujet de la connaissance de la combinaison linéaire. Donc, ces spécifications ne sont pas déraisonnables.

Nous avons appliqué notre méthode telle que décrite pour les quatre scénarios. Dans les autres colonnes du tableau 1, nous étudions les estimations des probabilités de petits domaines. Nous présentons la moyenne a posteriori (MP), l'écart-type a posteriori (ETP),

$$REQM = \sqrt{(\hat{\pi} - MP)^2 + ETP^2},$$

où  $\hat{\pi}$  est l'estimation directe et l'intervalle de confiance de densité a posteriori maximale (DPM) à 95 % (Int). Comme prévu, les ETP sont classés approximativement par ordre croissant : modèle 2, modèle 3, modèle 4 et modèle 1 ; dans certains cas, les différences sont importantes. Les MP pour les modèles 1, 2 et 3 sont en partie similaires, mais pour le modèle 4, les MP sont en grande partie plus petites que pour les trois autres modèles. Les modèles 2 et 3 présentent une nette amélioration par rapport au modèle 1, du moins en ce qui concerne la précision. Ce gain devient moins important pour le modèle 4, c'est-à-dire le modèle avec la plus grande incertitude au sujet de  $\theta$ .

**Tableau 1**  
**Comparaison des quatre modèles en utilisant la moyenne a posteriori (MP), l'écart-type a posteriori (ETP), la racine carrée de l'erreur quadratique moyenne (REQM), et les intervalles de crédibilité DPM à 95 % (HPD) (Int) de  $\pi_i$  par domaine (D) pour les données de la NHANES**

D	s	n	$\hat{\pi}$	MP	ETP	REQM	Int	MP	ETP	REQM	Int
				Modèle 1				Modèle 2			
1	4	47	0,085	0,114	0,033	0,044	(0,051 ; 0,179)	0,111	0,032	0,041	(0,049 ; 0,170)
2	2	29	0,069	0,112	0,037	0,057	(0,042 ; 0,183)	0,111	0,036	0,055	(0,041 ; 0,178)
3	10	44	0,227	0,175	0,044	0,068	(0,100 ; 0,264)	0,177	0,041	0,065	(0,108 ; 0,260)
4	5	62	0,081	0,107	0,030	0,040	(0,047 ; 0,159)	0,107	0,027	0,038	(0,054 ; 0,160)
5	10	74	0,135	0,134	0,030	0,030	(0,077 ; 0,194)	0,134	0,028	0,028	(0,080 ; 0,190)
6	12	69	0,174	0,158	0,036	0,039	(0,089 ; 0,227)	0,155	0,031	0,036	(0,095 ; 0,214)
7	8	79	0,101	0,116	0,028	0,031	(0,065 ; 0,173)	0,115	0,027	0,030	(0,065 ; 0,166)
8	5	62	0,081	0,107	0,030	0,040	(0,052 ; 0,169)	0,105	0,029	0,038	(0,042 ; 0,153)
9	28	123	0,228	0,196	0,036	0,048	(0,129 ; 0,262)	0,196	0,032	0,045	(0,131 ; 0,253)
10	10	111	0,090	0,106	0,026	0,030	(0,059 ; 0,155)	0,105	0,024	0,028	(0,061 ; 0,150)
11	16	122	0,131	0,132	0,026	0,026	(0,083 ; 0,183)	0,130	0,023	0,023	(0,090 ; 0,179)
12	20	137	0,146	0,144	0,026	0,026	(0,094 ; 0,194)	0,141	0,022	0,023	(0,100 ; 0,184)
				Modèle 3				Modèle 4			
1	4	47	0,085	0,111	0,033	0,042	(0,044 ; 0,169)	0,109	0,032	0,040	(0,050 ; 0,172)
2	2	29	0,069	0,111	0,037	0,056	(0,039 ; 0,179)	0,108	0,036	0,053	(0,037 ; 0,173)
3	10	44	0,227	0,175	0,043	0,068	(0,093 ; 0,260)	0,170	0,044	0,072	(0,091 ; 0,255)
4	5	62	0,081	0,106	0,029	0,038	(0,050 ; 0,160)	0,103	0,030	0,038	(0,048 ; 0,164)
5	10	74	0,135	0,134	0,029	0,029	(0,077 ; 0,189)	0,129	0,030	0,030	(0,067 ; 0,184)
6	12	79	0,174	0,156	0,034	0,038	(0,090 ; 0,217)	0,151	0,036	0,043	(0,087 ; 0,222)
7	8	69	0,101	0,118	0,028	0,033	(0,062 ; 0,171)	0,111	0,028	0,029	(0,061 ; 0,167)
8	5	62	0,081	0,107	0,030	0,040	(0,051 ; 0,165)	0,102	0,030	0,036	(0,050 ; 0,159)
9	28	123	0,228	0,195	0,034	0,047	(0,138 ; 0,265)	0,189	0,035	0,052	(0,123 ; 0,255)
10	10	111	0,090	0,107	0,024	0,029	(0,062 ; 0,156)	0,104	0,025	0,029	(0,051 ; 0,149)
11	16	122	0,131	0,132	0,024	0,024	(0,086 ; 0,179)	0,126	0,025	0,025	(0,083 ; 0,179)
12	20	137	0,146	0,143	0,024	0,024	(0,095 ; 0,191)	0,137	0,025	0,027	(0,091 ; 0,189)

Nota : Les quatre modèles sont : modèle 1 – pas de contrainte ; modèle 2 –  $\theta$  fixe ; modèle 3 – prior informatif pour  $\theta$  ; modèle 4 – prior uniforme pour  $\theta$ . Les domaines sont formés par croisement de l'école (secondaire inférieur – SI, secondaire supérieur – SS), la race (blanche – B, noire – N, mexicaine – M) et le sexe (masculin – M, féminin – F). Donc, les domaines sont : 1-SIBM, 2-SINF, 3-SIMM, 4-SIBF, 5-SINM, 6-SIMF, 7-SSBM, 8-SSNF, 9-SSMM, 10-SSBF, 11-SSNM, 12-SSMF (par exemple, le premier domaine constitue des garçons blancs au cycle secondaire inférieur).  $n$  est le nombre d'adolescents et  $s$  est le nombre d'adolescents obèses dans chaque domaine. Les données proviennent des 35 plus grands comtés des É.-U. Une estimation de la probabilité globale est  $130 / 959 \approx 0,136$ , et pour le premier domaine  $\hat{p} = 4 / 47 = 0,085$  ; les erreurs-types numériques sont toutes plus petites que 0,001 ;  $REQM = \sqrt{(\hat{\pi} - MP)^2 + ETP^2}$ .

Nous étudions également très brièvement le paramètre de nuisance  $\theta$ . Nous notons que la moyenne pondérée des estimateurs directs des petits domaines est 0,136 (plus exactement 0,135599). Quand  $\theta$  est maintenu fixe à 0,135599, la moyenne pondérée des moyennes a posteriori est 0,136. Quand  $\theta$  possède le prior informatif, la moyenne pondérée des moyennes a posteriori est 0,136, et pour  $\theta$ , la MP est 0,136, l'ETP est 0,008, et un intervalle DPM à 95 % pour  $\theta$  est (0,122 ; 0,152). Quand  $\theta$  a le prior uniforme, la moyenne pondérée des moyennes a posteriori est 0,132, et pour  $\theta$  la MP est 0,131, l'ETP est 0,011, et un intervalle DPM à 95 % pour  $\theta$  est (0,110 ; 0,151). Cela montre les déficiences de la loi a priori uniforme que nous utilisons uniquement aux fins de comparaison. Il convient de souligner que  $\mu_1, \dots, \mu_{\ell-1}$  et  $\theta$  sont calculés pour commencer. Puis  $\mu_\ell$  est obtenu par soustraction. Cela est fait à chaque itération de l'échantillonneur de Gibbs. Ensuite, les valeurs sommaires a posteriori pour  $\sum_{i=1}^{\ell} \omega_i \pi_i$  et  $\theta$  sont calculées. Donc, il y aura de légères discordances qui sont dues à l'arrondissement.

Enfin, nous avons sélectionné les quatre plus petits domaines pour comparer les densités a posteriori des probabilités. Nous avons utilisé l'estimateur à noyau de la densité de probabilité de Parzen-Rosenblatt pour estimer les densités a posteriori ; voir Silverman (1986) pour des détails. La figure 1 compare les densités a posteriori estimées pour les quatre modèles. Il est intéressant de noter que, à mesure que les tailles des domaines augmentent, les quatre modèles se rapprochent. En outre, dans tous les cas, les queues de distribution dans chaque panneau sont très semblables ; les différences entre ces distributions tiennent aux intervalles modaux (c'est-à-dire intervalle contenant le mode), et leurs hauteurs. Comme prévu, la densité a posteriori correspondant au modèle non restreint est la plus courte, simplement parce qu'elle a plus de variabilité. Le modèle 4 voit sa densité a posteriori déplacée vers la gauche et est légèrement bimodal pour le domaine le plus petit. Donc, l'inférence au sujet des modes de ces distributions sera différente. Mais l'inférence faisant intervenir les queues ne sera pas si différente ; sauf pour le modèle 4, les intervalles de crédibilité à 95 % seront similaires.

### 3.2 Étude en simulation

Nous utilisons une étude en simulation pour évaluer les propriétés statistiques de notre méthode. Nous voulons voir si le gain de précision persiste et comment les estimateurs des probabilités sont déplacés. Nous étudions aussi les propriétés fréquentistes des estimateurs des probabilités. Dans la description de la simulation, il est commode d'utiliser les noms abrégés des modèles qui sont NR (modèle 1, non restreint), FI (modèle 2,  $\theta$  fixe), IN (modèle 3,

prior informatif pour  $\theta$ ) et UN (modèle 4, prior uniforme pour  $\theta$ ).

Nous fixons  $\theta_0 = 0,15$ ,  $\mu_0 = \theta_0$  et  $\tau_0 = 100$ . Nous avons choisi trois valeurs de  $\ell = 12, 24, 36$ , 12 étant le nombre de domaines dans les données de la NHANES. Nous avons tiré les tailles d'échantillon à partir d'une densité de probabilité uniforme dans (25 ; 150), de nouveau pour refléter les données de la NHANES. Premièrement, nous avons généré

$$\pi_i \stackrel{\text{iid}}{\sim} \text{Bêta} \{ \mu_0 \tau_0, (1 - \mu_0) \tau_0 \}, \quad i = 1, \dots, \ell.$$

Pour exécuter cette tâche, nous avons tiré des ensembles de  $\ell$   $\pi_i$  jusqu'à  $\theta_0 - w_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \theta_0$  ; ensemble  $\pi_\ell = (\theta_0 - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / w_\ell$ . Puis, nous avons généré

$$s_i \stackrel{\text{ind}}{\sim} \text{Binomiale}(n_i, \pi_i).$$

Nous avons généré 1 000 ensembles de données de cette manière pour chacun des  $\ell = 12, 24, 36$ . Puis, nous avons ajusté les quatre modèles (un modèle non restreint et trois modèles restreints). Le processus est très rapide (c'est-à-dire, pour les tailles d'échantillon de 12, 24, 30, il y a eu respectivement 22, 90 et 153 rejets dans les 1 000 échantillons). Nous avons ajusté chaque ensemble de données en utilisant des échantillons aléatoires pour le modèle non restreint, et l'échantillonneur de Gibbs « griddy » pour les modèles restreints. Nous avons ajusté les 1 000 ensembles de données en une ou deux heures sur notre ordinateur Alpha 2 × 833 MHz.

Pour ces 1 000 simulations, nous étudions la MP, la couverture (C), le biais (B), l'ETP, la REQM et la largeur (L) des intervalles de crédibilité à 95 %. Pour chaque domaine, nous calculons le biais  $MP - \pi$ , puis nous calculons la moyenne de ces valeurs sur tous les domaines et toutes les exécutions des simulations, et nous appelons maintenant cette quantité  $B$ . Associé à  $B$ , nous avons aussi calculé  $AB$ , la moyenne de  $|MP - \pi|$ . De même, nous avons calculé

$$\text{REQM} = \sqrt{(MP - \pi)^2 + \text{ETP}^2}$$

pour chaque domaine et chaque exécution de la simulation et nous avons calculé la moyenne de ces valeurs sur tous les domaines et toutes les exécutions de simulation. Notons que les probabilités réelles,  $\pi_i$ , sont connues par conception. Nous obtenons la couverture (C) en calculant la proportion de tous les intervalles contenant la valeur réelle de  $\pi_i$  sur tous les domaines et toutes les exécutions de simulation. Nous obtenons aussi la moyenne des largeurs des intervalles de crédibilité à 95 %. Les erreurs-types numériques sont obtenues pour toutes les quantités.

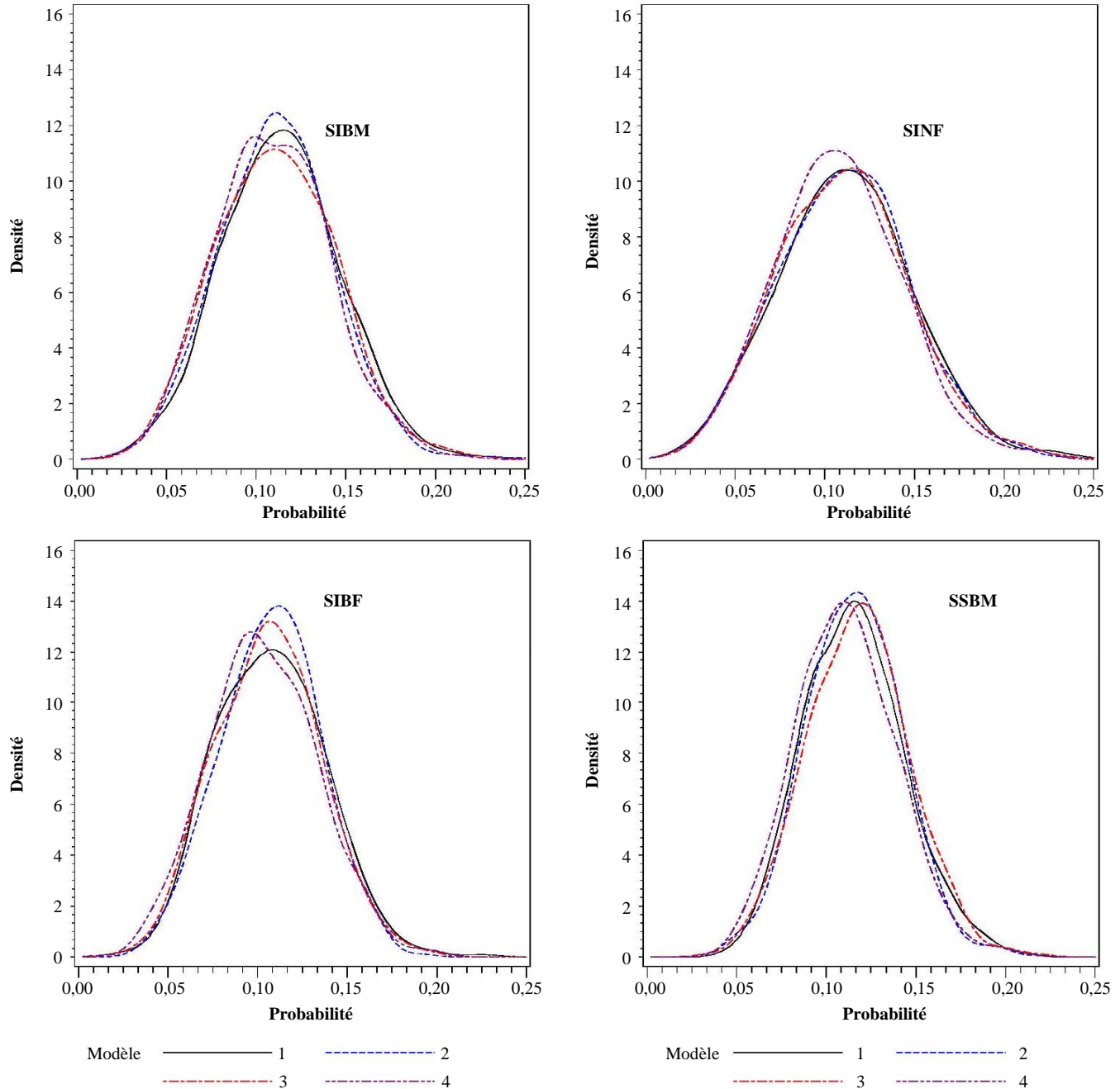


Figure 1 Tracés des densités a posteriori estimées pour  $\pi_1$ ,  $\pi_2$ ,  $\pi_4$ , et  $\pi_7$  pour les quatre modèles et les données de la NHANES

Dans le tableau 2, nous étudions les estimations des probabilités de petits domaines. Il est commode d'utiliser des noms abrégés des quatre modèles pour nos discussions. Pour IN, les MP sont proches de la valeur nominale de 0,15, mais pour UN, les MP sont plus petites que la valeur nominale, particulièrement pour NR à  $\ell = 12$ . Nous observons que la couverture pour tous les modèles NR, FI et UN est toujours plus grande que la valeur nominale de 95 %, mais pour le modèle IN, la couverture est plus faible que la valeur nominale de 95 %. Une différence similaire existe pour le biais ; bien que le biais soit faible pour tous les modèles, les

modèles NR, FI (la valeur spécifiée de  $\theta$  est 0,15) et UN ont un biais négatif, mais IN a un biais positif. Sauf pour  $\ell = 36$ , IN possède le plus grand AB. Les ETP sont pour la plupart principalement semblables et les REQM ont les mêmes caractéristiques ; il existe quelques différences à  $\ell = 12$ . Les quatre modèles deviennent similaires à mesure que  $\ell$  augmente ; quand  $\ell$  est grand, il semble que notre méthode ne soit plus nécessaire. Cependant, de nouveau, le gain de précision semble être par ordre croissant FI, IN, UN et NR.

**Tableau 2**

**Simulation : Comparaison des quatre modèles en utilisant la couverture (C), le biais et la moyenne de la valeur absolue du biais (B et AB), l'écart-type a posteriori (ETP), la racine carrée de l'erreur quadratique moyenne (REQM) et la largeur (L) de l'intervalle de crédibilité à 95 % de  $\pi_i$**

$\ell$	Modèle	C	B	AB	ETP	REQM	L
12	NR	0,960 <sub>0,0018</sub>	-0,002 <sub>0,0003</sub>	0,0231 <sub>0,00016</sub>	0,033 <sub>0,0001</sub>	0,043 <sub>0,0001</sub>	0,125 <sub>0,0003</sub>
	FI	0,961 <sub>0,0018</sub>	-0,000 <sub>0,0003</sub>	0,0219 <sub>0,00020</sub>	0,031 <sub>0,0001</sub>	0,040 <sub>0,0001</sub>	0,118 <sub>0,0003</sub>
	IN	0,946 <sub>0,0021</sub>	0,005 <sub>0,0003</sub>	0,0275 <sub>0,00066</sub>	0,032 <sub>0,0001</sub>	0,043 <sub>0,0001</sub>	0,122 <sub>0,0002</sub>
24	NR	0,956 <sub>0,0019</sub>	-0,000 <sub>0,0003</sub>	0,0261 <sub>0,00019</sub>	0,032 <sub>0,0001</sub>	0,042 <sub>0,0001</sub>	0,122 <sub>0,0003</sub>
	FI	0,957 <sub>0,0013</sub>	-0,001 <sub>0,0002</sub>	0,0229 <sub>0,00012</sub>	0,031 <sub>0,0000</sub>	0,041 <sub>0,0001</sub>	0,119 <sub>0,0002</sub>
	IN	0,943 <sub>0,0015</sub>	0,006 <sub>0,0002</sub>	0,0252 <sub>0,00058</sub>	0,030 <sub>0,0000</sub>	0,041 <sub>0,0001</sub>	0,116 <sub>0,0001</sub>
36	NR	0,952 <sub>0,0014</sub>	-0,000 <sub>0,0002</sub>	0,0236 <sub>0,00012</sub>	0,031 <sub>0,0002</sub>	0,041 <sub>0,0002</sub>	0,118 <sub>0,0005</sub>
	FI	0,960 <sub>0,0010</sub>	-0,001 <sub>0,0001</sub>	0,0224 <sub>0,00009</sub>	0,030 <sub>0,0000</sub>	0,040 <sub>0,0001</sub>	0,117 <sub>0,0001</sub>
	IN	0,961 <sub>0,0010</sub>	-0,000 <sub>0,0001</sub>	0,0218 <sub>0,00009</sub>	0,030 <sub>0,0000</sub>	0,039 <sub>0,0001</sub>	0,115 <sub>0,0001</sub>
	NR	0,948 <sub>0,0012</sub>	0,005 <sub>0,0002</sub>	0,0224 <sub>0,00009</sub>	0,030 <sub>0,0000</sub>	0,040 <sub>0,0001</sub>	0,114 <sub>0,0001</sub>
	NR	0,957 <sub>0,0011</sub>	-0,000 <sub>0,0001</sub>	0,0228 <sub>0,00010</sub>	0,030 <sub>0,0000</sub>	0,040 <sub>0,0001</sub>	0,116 <sub>0,0001</sub>

Nota : Quatre modèles sont : modèle 1 – non restreint (NR) ; modèle 2 –  $\theta$  fixe (FI) ; modèle 3 – prior informatif pour  $\theta$  (IN) ; modèle 4 – prior uniforme pour  $\theta$  (UN). REQM =  $\sqrt{(\pi - MP)^2 + ETP^2}$ . La notation  $a_b$  signifie que  $a$  est une estimation et que  $b$  est l'erreur-type.

Dans la plupart des applications, la valeur exacte de  $\theta$  est inconnue. Par conséquent, les ETP des  $\pi_i$ , dans la situation où  $\theta$  est supposé connu, sous-estime vraisemblablement les ETP réels. Donc, nous étudions les écarts des ETP de IN et UN par rapport à ceux de FI, et nous calculons les ratios,  $R_1 = ETP_{IN} / ETP_{FI}$  et  $R_2 = ETP_{UN} / ETP_{FI}$ . Dans le tableau 3, nous présentons les sommaires à cinq chiffres de ces ratios selon la taille d'échantillon. La plupart des ratios sont situés autour de 1 (c'est-à-dire intervalle interquartile) avec une certaine tendance à être plus grands que 1. (Notons que les maxima à  $\ell = 12$  et  $\ell = 24$  sont des valeurs aberrantes, peut-être dues à de mauvais échantillons simulés.) Donc, dans l'ensemble, les ETP sous IN et UN ne sont pas beaucoup plus grands que sous FI.

**Tableau 3**

**Simulation : Une étude de l'écart-type a posteriori (ETP) de  $\pi_i$  en utilisant les grandeurs sommaires à cinq chiffres des ratios,  $R_1$  et  $R_2$ , selon la taille d'échantillon**

$\ell$	Ratio	Min	$Q_1$	Méd	$Q_3$	Max
12	$R_1$	0,673	0,972	1,032	1,091	5,329
	$R_2$	0,022	0,984	1,034	1,086	85,370
24	$R_1$	0,019	0,965	1,005	1,047	16,017
	$R_2$	0,024	0,979	1,014	1,049	486,960
36	$R_1$	0,690	0,962	0,998	1,034	1,236
	$R_2$	0,837	0,979	1,011	1,044	1,243

Nota :  $R_1 = ETP_{IN} / ETP_{FI}$  et  $R_2 = ETP_{UN} / ETP_{FI}$ . Les cinq grandeurs sommaires sont le minimum (min), le premier quartile ( $Q_1$ ), la médiane (méd), le troisième quartile ( $Q_3$ ) et le maximum (max).

Dans le tableau 4, nous étudions l'estimation de  $\theta$  pour deux modèles pertinents IN et UN. Pour les deux modèles, les probabilités de couverture sont plus faibles que la valeur nominale, et la couverture pour UN est plus faible que l'intervalle pour IN. Le biais est faible pour les deux modèles, positif pour IN et négatif pour UN. Sauf pour

$\ell = 36$ , IN possède de loin le plus grand AB. Les ETP et les REQM sont généralement plus petits pour IN, et les largeurs des intervalles de crédibilité à 95 % sont significativement plus faibles pour IN. Il semble qu'il est difficile d'estimer  $\theta$  sous UN, mais IN semble donner de légèrement meilleurs résultats.

Dans le tableau 5, nous présentons des résultats plus détaillés (c'est-à-dire par domaine) pour le cas où le nombre de domaines est 12. Pour montrer des gains supplémentaires de précision, nous avons réduit la taille d'échantillon de moitié [c'est-à-dire nous avons tiré les tailles d'échantillon uniformément dans l'intervalle (12 ; 75)]. Nous présentons l'écart-type a posteriori et la racine carrée de l'erreur quadratique moyenne a posteriori, moyennés sur les exécutions de simulation. De nouveau, les erreurs-types sont présentées. Nous notons que tous les contenus de probabilité (non présentés) sont au moins égaux à la valeur nominale de 95 %. Les erreurs-types numériques sont faibles dans tous les cas. Les ETP et les REQM sont dans le bon ordre. Notons que, parce que les tailles d'échantillon sont classées par ordre de la plus petite à la plus grande, les ETP et les REQM diminuent à mesure que le nombre de domaines augmente.

Nous étudions la densité a posteriori de  $\pi_1$  pour  $\ell = 12$ , et nous comparons les quatre modèles. De nouveau, nous utilisons l'estimateur de la densité de Parzen-Rosenblatt. À la figure 2, nous présentons les densités a posteriori estimées (Parzen-Rosenblatt) moyennées sur les 1 000 exécutions pour  $\ell = 12$ . Nous obtenons les mêmes résultats que pour les données sur l'IMC. De nouveau, les queues sont similaires. FI est la densité la plus élevée et UN la plus courte. UN est déplacé légèrement vers la gauche de IN. Dans la figure 3, nous présentons un échantillon systématique de dix densités pour les 1 000 simulations exécutées par modèle. Nous pouvons voir une variation importante entre les

dix densités a posteriori estimées. De nouveau, nous constatons que FI est la plus élevée ; UR, FI et UN montrent une variation comparable, avec IN légèrement plus grande ; il est important de prendre la moyenne pour les comparaisons comme à la figure 2.

#### 4. Conclusion

Nous avons étendu le modèle bêta-binomial de l'estimation sur petits domaines pour accommoder une spécification a priori d'une moyenne pondérée des probabilités de domaine. Nous avons utilisé l'approche bayésienne, qui est particulièrement séduisante pour les problèmes avec fonction de probabilité difficile à utiliser, comme dans notre application avec la contrainte de la moyenne pondérée du modèle bêta-binomial. Nous avons considéré la contrainte comme une connaissance a priori qui peut être précise ou moins informative. L'échantillonneur de Gibbs « griddy » est utilisé pour ajuster les modèles, ce qui évite de recourir à l'échantillonneur de Metropolis-Hastings plus complexe. Nous avons élaboré une théorie qui permet un échantillonnage à partir d'une fonction de densité proportionnelle au produit d'une densité bêta-binomial tronquée et d'une densité bêta généralisée. Nous avons constaté que, dans

l'ensemble, notre algorithme complet formant l'échantillonneur de Gibbs « griddy » s'exécute efficacement et rapidement.

Nous avons montré qu'il pourrait y avoir des gains de précision quand de l'information supplémentaire est intégrée dans le modèle bêta-binomial. Nous avons considéré trois scénarios dans lesquels un praticien des sondages (a) ne peut spécifier aucune contrainte (modèle bêta-binomial standard pour petits domaines), (b) peut spécifier une contrainte et le paramètre complètement et (c) peut spécifier une contrainte et l'information qui peut être utilisée pour construire une distribution a priori pour le paramètre. Notre exemple d'obésité des enfants dans la National Health and Nutrition Examination Survey et l'étude par simulation ont montré que le gain de précision au-delà de (a) est dans un ordre tel que (b) est plus grand que (c). Comme les arguments algébriques exacts sont difficiles, nous avons obtenu une approximation analytique qui montre qu'il pourrait en effet y avoir un gain de précision de (b) par rapport à (a). Aux fins de comparaison, nous avons considéré un quatrième scénario dans lequel  $\theta$  possède de l'information vague et, comme prévu, il s'est avéré plutôt non intéressant et inefficace.

Tableau 4

**Simulation : Comparaison des modèles informatif (IN) et uniforme (UN) en utilisant la moyenne a posteriori (MP), la couverture (C), le biais et la moyenne de la valeur absolue du biais (B et AB), l'écart-type a posteriori (ETP), la racine carrée de l'erreur quadratique moyenne (REQM) et la largeur (L) de l'intervalle de crédibilité à 95 % de  $\pi_i$**

$\ell$	Modèle	MP	C	B	AB	ETP	REQM	L
12	IN	0,149 <sub>0,0012</sub>	0,853 <sub>0,0112</sub>	0,000 <sub>0,0003</sub>	0,00152 <sub>0,00081</sub>	0,008 <sub>0,0000</sub>	0,012 <sub>0,0002</sub>	0,030 <sub>0,0001</sub>
	UN	0,138 <sub>0,0005</sub>	0,881 <sub>0,0102</sub>	-0,012 <sub>0,0004</sub>	0,00038 <sub>0,00003</sub>	0,011 <sub>0,0001</sub>	0,016 <sub>0,0002</sub>	0,042 <sub>0,0002</sub>
24	IN	0,153 <sub>0,0015</sub>	0,833 <sub>0,0118</sub>	0,003 <sub>0,0015</sub>	0,00212 <sub>0,00103</sub>	0,007 <sub>0,0006</sub>	0,012 <sub>0,0015</sub>	0,024 <sub>0,0015</sub>
	UN	0,145 <sub>0,0029</sub>	0,842 <sub>0,0115</sub>	-0,005 <sub>0,0003</sub>	0,00012 <sub>0,00006</sub>	0,008 <sub>0,0001</sub>	0,012 <sub>0,0002</sub>	0,030 <sub>0,0002</sub>
36	IN	0,150 <sub>0,0002</sub>	0,828 <sub>0,0119</sub>	0,000 <sub>0,0002</sub>	0,00004 <sub>0,00000</sub>	0,004 <sub>0,0000</sub>	0,007 <sub>0,0001</sub>	0,017 <sub>0,0001</sub>
	UN	0,145 <sub>0,0003</sub>	0,794 <sub>0,0128</sub>	-0,005 <sub>0,0002</sub>	0,00009 <sub>0,00000</sub>	0,006 <sub>0,0000</sub>	0,010 <sub>0,0001</sub>	0,024 <sub>0,0001</sub>

Nota : Les deux modèles considérés sont : modèle 3 – prior informatif pour  $\theta$  (IN) et modèle 4 – prior uniforme pour  $\theta$  (UN).  $REQM = \sqrt{(\pi - MP)^2 + ETP^2}$ . La notation  $a_b$  signifie que  $a$  est une estimation et que  $b$  est l'erreur-type.

Tableau 5

**Simulation : Comparaison des quatre modèles en utilisant l'écart-type a posteriori et la racine carrée de l'erreur quadratique moyenne (REQM) des  $\pi_i$  par domaine (D)**

D	Non restreint		Fixe		Informatif		Uniforme	
	ETP	REQM	ETP	REQM	ETP	REQM	ETP	REQM
1	0,048 <sub>0,0003</sub>	0,057 <sub>0,0004</sub>	0,046 <sub>0,0003</sub>	0,054 <sub>0,0004</sub>	0,045 <sub>0,0002</sub>	0,056 <sub>0,0005</sub>	0,047 <sub>0,0004</sub>	0,056 <sub>0,0005</sub>
2	0,046 <sub>0,0003</sub>	0,055 <sub>0,0004</sub>	0,044 <sub>0,0003</sub>	0,053 <sub>0,0004</sub>	0,044 <sub>0,0002</sub>	0,054 <sub>0,0005</sub>	0,045 <sub>0,0004</sub>	0,054 <sub>0,0005</sub>
3	0,044 <sub>0,0002</sub>	0,053 <sub>0,0004</sub>	0,042 <sub>0,0002</sub>	0,050 <sub>0,0004</sub>	0,042 <sub>0,0002</sub>	0,052 <sub>0,0005</sub>	0,043 <sub>0,0003</sub>	0,051 <sub>0,0004</sub>
4	0,042 <sub>0,0002</sub>	0,050 <sub>0,0004</sub>	0,040 <sub>0,0002</sub>	0,047 <sub>0,0004</sub>	0,040 <sub>0,0002</sub>	0,050 <sub>0,0004</sub>	0,041 <sub>0,0002</sub>	0,049 <sub>0,0004</sub>
5	0,041 <sub>0,0002</sub>	0,049 <sub>0,0004</sub>	0,038 <sub>0,0002</sub>	0,046 <sub>0,0004</sub>	0,039 <sub>0,0002</sub>	0,048 <sub>0,0004</sub>	0,039 <sub>0,0003</sub>	0,048 <sub>0,0005</sub>
6	0,040 <sub>0,0002</sub>	0,048 <sub>0,0004</sub>	0,037 <sub>0,0002</sub>	0,045 <sub>0,0004</sub>	0,037 <sub>0,0002</sub>	0,048 <sub>0,0004</sub>	0,038 <sub>0,0003</sub>	0,047 <sub>0,0005</sub>
7	0,038 <sub>0,0002</sub>	0,046 <sub>0,0004</sub>	0,035 <sub>0,0002</sub>	0,043 <sub>0,0003</sub>	0,036 <sub>0,0002</sub>	0,046 <sub>0,0004</sub>	0,037 <sub>0,0003</sub>	0,045 <sub>0,0004</sub>
8	0,037 <sub>0,0002</sub>	0,045 <sub>0,0003</sub>	0,034 <sub>0,0002</sub>	0,041 <sub>0,0003</sub>	0,036 <sub>0,0002</sub>	0,046 <sub>0,0004</sub>	0,036 <sub>0,0003</sub>	0,044 <sub>0,0004</sub>
9	0,036 <sub>0,0002</sub>	0,044 <sub>0,0003</sub>	0,033 <sub>0,0002</sub>	0,040 <sub>0,0004</sub>	0,034 <sub>0,0001</sub>	0,044 <sub>0,0004</sub>	0,035 <sub>0,0003</sub>	0,042 <sub>0,0004</sub>
10	0,035 <sub>0,0002</sub>	0,043 <sub>0,0003</sub>	0,032 <sub>0,0002</sub>	0,039 <sub>0,0003</sub>	0,034 <sub>0,0001</sub>	0,044 <sub>0,0004</sub>	0,034 <sub>0,0003</sub>	0,042 <sub>0,0004</sub>
11	0,034 <sub>0,0001</sub>	0,042 <sub>0,0003</sub>	0,031 <sub>0,0002</sub>	0,038 <sub>0,0003</sub>	0,033 <sub>0,0001</sub>	0,042 <sub>0,0004</sub>	0,033 <sub>0,0003</sub>	0,041 <sub>0,0004</sub>
12	0,035 <sub>0,0002</sub>	0,047 <sub>0,0005</sub>	0,031 <sub>0,0002</sub>	0,042 <sub>0,0004</sub>	0,034 <sub>0,0003</sub>	0,047 <sub>0,0006</sub>	0,034 <sub>0,0007</sub>	0,046 <sub>0,0008</sub>

Note: Quatre modèles sont : modèle 1 – non restreint (NR) ; modèle 2 –  $\theta$  fixe (FI) ; modèle 3 – prior informatif pour  $\theta$  (IN) ; modèle 4 – prior uniforme pour  $\theta$  (UN).  $REQM = \sqrt{(\pi - MP)^2 + ETP^2}$ . La notation  $a_b$  signifie que  $a$  est une estimation et que  $b$  est l'erreur-type. Ici, 12 domaines sont utilisés et les tailles d'échantillon originales sont divisées par 2.

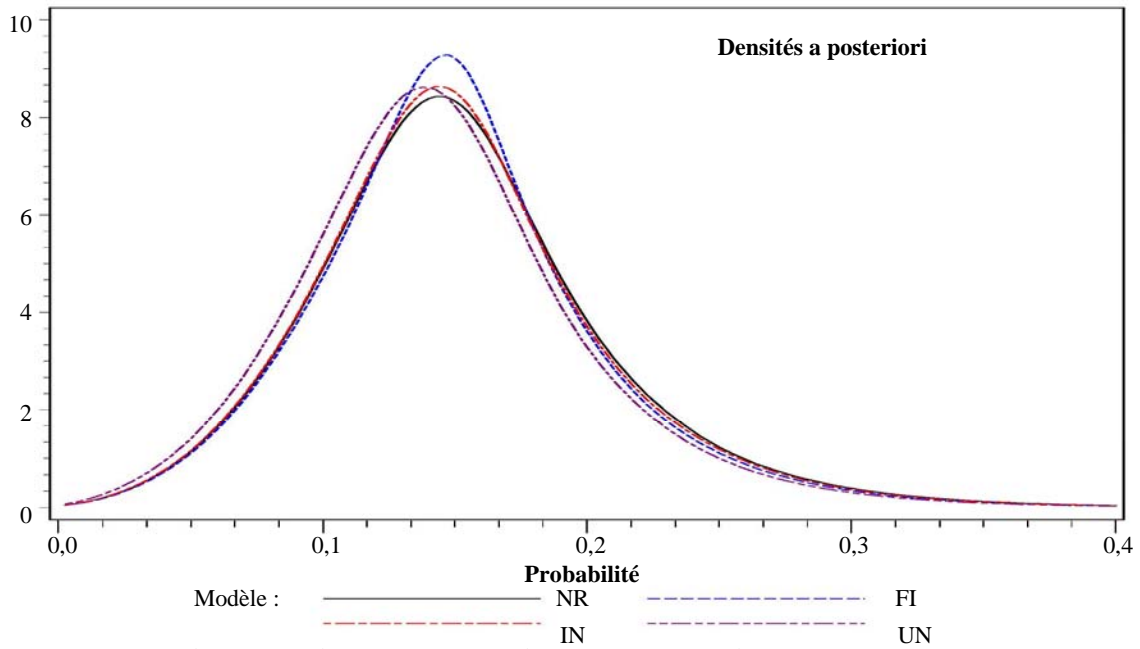


Figure 2 Tracés des densités a posteriori estimées pour  $\pi_1$  par modèle quand il existe 12 domaines

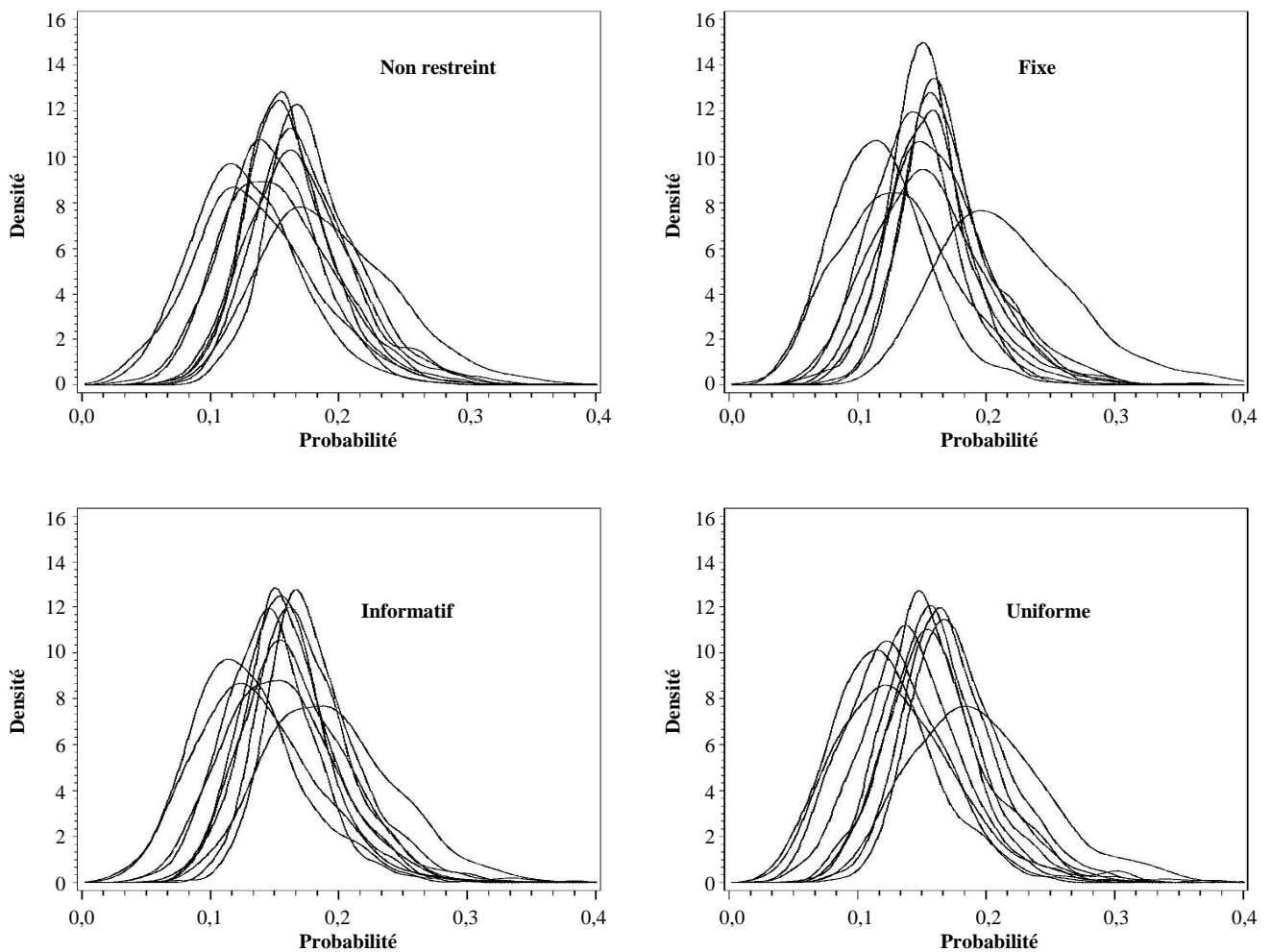


Figure 3 Tracés des densités a posteriori estimées de  $\pi_1$  pour un échantillon systématique de taille 10 pour les 1 000 exécutions par modèle quand il existe 12 domaines

Il est simple de faire une inférence prédictive bayésienne au sujet de la moyenne de population finie de chaque petit domaine. Soit  $P_i = T_i / N_i$  la proportion de la population finie pour le  $i^e$  domaine, où  $T_i = \sum_{j=1}^{N_i} y_{ij}$ ,  $y_{ij}$  sont les réponses binaires, et  $N_i$ , le nombre d'individus dans le  $i^e$  domaine, est supposé connu. Maintenant  $T_i = t_i^{(s)} + t_i^{(ns)}$ , où  $t_i^{(s)}$  et  $t_i^{(ns)}$  sont respectivement le total d'échantillon et le total de non-échantillon. Maintenant, sous n'importe lequel des modèles,  $t_i^{(ns)} | \pi_i \sim$  Bêta-binomiale( $n_i, \pi_i$ ) et  $p(t_i^{(ns)} | \mathbf{y}_s) = \int p(t_i^{(ns)} | \pi_i) p(\pi | \mathbf{y}_s) d\pi_i$ , où  $\mathbf{y}_s = (y_1, \dots, y_\ell)'$ . Donc, il est facile d'obtenir la densité a posteriori empirique de  $P_i$  en utilisant la méthode fondée sur l'échantillonnage. Nandram et Sedransk (1993) ont obtenu certaines caractéristiques analytiques de  $P_i$  quand  $\tau$  est connu, mais non avec la contrainte ; voir aussi Nandram (1998).

Nous mentionnons une généralisation de notre modèle bayésien hiérarchique bêta-binomial restreint au modèle multinomial de Dirichlet (par exemple, Nandram 1998). Soit  $\mathbf{y}_i$  le  $c$ -vecteur de fréquence de cellule (c'est-à-dire le nombre de personnes possédant l'un des  $c$  traits), et soit  $n_i$  la taille de l'échantillon dans le  $i^e$  domaine,  $i = 1, \dots, \ell$ . Nous supposons que

$$\mathbf{y}_i | \pi_i \sim \text{Multinomiale}(n_i, \pi_i),$$

$$\pi_i | \boldsymbol{\mu}, \tau, \boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\mu}\tau)$$

avec  $\sum_{i=1}^{\ell} w_i \pi_i = \boldsymbol{\theta}$ . Enfin  $\boldsymbol{\theta} \sim \text{Dirichlet}(\boldsymbol{\mu}_0 \tau_0)$ , où  $\boldsymbol{\mu}_0$  et  $\tau_0$  doivent être spécifiés, et indépendamment  $p(\boldsymbol{\mu}, \tau) = (k-1)! / (1+\tau)^2$ ,  $0 < \mu_k < 1$ ,  $k = 1, \dots, c$ ,  $\sum_{k=1}^c \mu_k = 1$ . Avec  $k$  contraintes, ce problème est beaucoup plus complexe, mais nous prévoyons nous y atteler. D'autres extensions à la non-réponse non ignorable (Nandram et Choi 2002) et aux tables de contingence à deux variables sont possibles.

## Remerciements

Les auteurs remercient le rédacteur associé et les deux examinateurs qui ont contribué considérablement à améliorer la qualité de la présentation.

## Annexe A

### Preuves des lemmes 1, 2 et des théorèmes 1, 2

#### Preuve du lemme 1

Il s'agit d'un cas particulier d'un résultat général. En utilisant la règle de multiplication et parce que le prior est approprié, il est clair que la densité conjointe de  $\boldsymbol{\pi}, \boldsymbol{\mu}, \tau, \mathbf{s}$  « s'intègre » pour donner la valeur un. Par conséquent, la densité a posteriori conjointe de  $\boldsymbol{\pi}, \boldsymbol{\mu}, \tau$  sachant  $\mathbf{s}$  est appropriée.

#### Preuve du théorème 1

Soit  $\mathcal{T} = \{(\boldsymbol{\pi}, \boldsymbol{\mu}, \tau, \boldsymbol{\theta}) : 0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0, \boldsymbol{\theta} - \omega_\ell \leq \sum_{i=1}^{\ell-1} \omega_i \pi_i \leq \boldsymbol{\theta}, 0 < \boldsymbol{\theta} < 1, \pi_\ell = (\boldsymbol{\theta} - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell\}$  et  $\mathcal{T}^* = \{(\boldsymbol{\pi}, \boldsymbol{\mu}, \tau) : 0 < \pi_i < 1, i = 1, \dots, \ell, 0 < \mu < 1, \tau > 0\}$ ; notons que  $\mathcal{T} \subset \mathcal{T}^*$ .

Soit  $\tilde{g}(\boldsymbol{\pi}, \boldsymbol{\mu}, \tau | \mathbf{s})$  le deuxième membre de la densité a posteriori non contrainte dans (7) et  $\tilde{p}(\boldsymbol{\pi}_{(\ell)}, \boldsymbol{\mu}, \tau, \boldsymbol{\theta} | \mathbf{s}, \phi = 0)$  le deuxième membre de la densité a posteriori contrainte dans (9). En notant que  $\pi_\ell = (\boldsymbol{\theta} - \sum_{i=1}^{\ell-1} \omega_i \pi_i) / \omega_\ell$ , nous observons que

$$\tilde{p}(\boldsymbol{\pi}_{(\ell)}, \boldsymbol{\mu}, \tau, \boldsymbol{\theta} | \mathbf{s}, \phi = 0) =$$

$$\tilde{g}(\boldsymbol{\pi}, \boldsymbol{\mu}, \tau | \mathbf{s}) \times \theta^{\mu_0 \tau_0 - 1} (1 - \theta)^{(1 - \mu_0) \tau_0 - 1}, (\boldsymbol{\pi}, \boldsymbol{\mu}, \tau, \boldsymbol{\theta}) \in \mathcal{T}.$$

Comme  $\theta^{\mu_0 \tau_0 - 1} (1 - \theta)^{(1 - \mu_0) \tau_0 - 1}$  est proportionnel à la fonction de densité de la variable aléatoire bêta, nous avons

$$\int_{\mathcal{T}} \tilde{p}(\boldsymbol{\pi}_{(\ell)}, \boldsymbol{\mu}, \tau, \boldsymbol{\theta} | \mathbf{s}, \phi = 0) d\boldsymbol{\pi} d\boldsymbol{\mu} d\tau d\boldsymbol{\theta} =$$

$$A \int_{\mathcal{T}} \tilde{g}(\boldsymbol{\pi}, \boldsymbol{\mu}, \tau | \mathbf{s}) d\boldsymbol{\pi} d\boldsymbol{\mu} d\tau \leq A \int_{\mathcal{T}^*} \tilde{g}(\boldsymbol{\pi}, \boldsymbol{\mu}, \tau | \mathbf{s}) d\boldsymbol{\pi} d\boldsymbol{\mu} d\tau,$$

où  $A = B\{\mu_0 \tau_0, (1 - \mu_0) \tau_0\}$  est la fonction bêta. En vertu du lemme 1,  $\int_{\mathcal{T}^*} \tilde{g}(\boldsymbol{\pi}, \boldsymbol{\mu}, \tau | \mathbf{s}) d\boldsymbol{\pi} d\boldsymbol{\mu} d\tau < \infty$ . Donc,  $p(\boldsymbol{\pi}_{(\ell)}, \boldsymbol{\mu}, \tau, \boldsymbol{\theta} | \mathbf{s}, \phi = 0)$  est approprié.

#### Preuve du lemme 2 (a)

Celui-ci peut être prouvé de deux façons. La dérivée seconde de  $\log\{f_2(x)\}$  est négative dans  $(c, d)$ , et donc la dérivée première, quand elle est fixée à zéro, fournit un mode unique qui est  $\delta d + (1 - \delta)c$ . Alternativement, comme  $(X - c)/(d - c) \sim$  Bêta( $a, b$ ) avec  $a, b > 1$ , il existe un mode unique pour  $(X - c)/(d - c)$ , et cela se traduit par  $\delta d + (1 - \delta)c$ ; notons que  $\delta d + (1 - \delta)c$  est un point dans  $(c, d)$ . Donc, en introduisant  $\delta d + (1 - \delta)c$  par substitution dans  $f_2(x)$ , nous avons

$$\sup_{c < x < d} f_2(x) = \delta^{a-1} (1 - \delta)^{b-1} / (d - c) B(a, b).$$

#### Preuve du lemme 2 (b)

Comme  $a, b > 1$ ,  $x \geq x - c$  et  $1 - x \geq d - x$ , il est vrai que

$$A^{-1} \geq D^{-1} \int_c^d (x - c)^{a+g-2} (d - x)^{b+h-2} dx,$$

où  $D = (d - c)^{a+b-1} B(a, b) B(g, h) \{F_{g,h}(d) - F_{g,h}(c)\}$  et  $F_{g,h}(x)$  est la fonction de répartition d'une variable aléatoire bêta standard dans  $(0, 1)$ . Notons que, comme  $c < d$  (strictement) et  $F_{g,h}(x)$  est croissante de manière monotone dans  $(0, 1)$ ,  $F_{g,h}(d) - F_{g,h}(c) > 0$  (strictement). Par comparaison avec la densité bêta généralisée [c'est-à-dire Beta( $a + g - 1, b + h - 1, c, d$ )], l'intégrale est  $(d - c)^{a+b+g+h-3} B(a + g - 1, b + h - 1)$ . Donc,



$$A^{-1} \geq \frac{(d-c)^{g+h-2} B(a+g-1, b+h-1)}{B(a, b)B(g, h)\{F_{g,h}(d) - F_{g,h}(c)\}} = H_1 > 0.$$

En outre, nous avons

$$A^{-1} \leq \int_c^d f_1(x) \sup_{c < x < d} f_2(x) dx.$$

Donc, en vertu du lemme 2 (a),

$$\begin{aligned} A^{-1} &\leq \frac{\delta^{a-1}(1-\delta)^{b-1}}{(d-c)B(a, b)} \int_c^d f_1(x) dx = \frac{\delta^{a-1}(1-\delta)^{b-1}}{(d-c)B(a, b)} \\ &= H_2 < \infty. \end{aligned}$$

## Preuve du théorème 2

Pour prouver cette allégation, nous calculons la fonction de répartition  $F_X(\cdot)$  de la variable aléatoire  $X$  définie dans le théorème. Nous avons

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P[F_{g,h}^{-1}\{UF_{g,h}(d) + (1-U)F_{g,h}(c)\} \leq x] \\ &= P[UF_{g,h}(d) + (1-U)F_{g,h}(c) \leq F_{g,h}(x)] \\ &= P[U\{F_{g,h}(d) - F_{g,h}(c)\} \leq F_{g,h}(x) - F_{g,h}(c)] \\ &= P\left[U \leq \frac{F_{g,h}(x) - F_{g,h}(c)}{F_{g,h}(d) - F_{g,h}(c)}\right]. \end{aligned}$$

Maintenant, puisque  $U \sim \text{Uniforme}(0, 1)$ , d'après l'expression susmentionnée pour  $F_X(\cdot)$ , nous avons  $F_X(x) = 1$  si  $x \geq d$  et  $F_X(x) = 0$  si  $x \leq c$ . Quand  $c \leq x \leq d$ , nous avons

$$F_X(x) = \frac{F_{g,h}(x) - F_{g,h}(c)}{F_{g,h}(d) - F_{g,h}(c)}.$$

Cela montre que  $X$  a la densité bêta tronquée  $f_1(x)$  par (20).

Maintenant, si nous voulons utiliser l'algorithme d'acceptation-rejet, considérons

$$\frac{f(x)}{f_1(x)} = Af_2(x).$$

En vertu du lemme 2, nous avons

$$\sup_{c < \pi < d} \left\{ \frac{f(x)}{f_1(x)} \right\} = A \sup_{c < \pi < d} f_2(x) = A \frac{\delta^{a-1}(1-\delta)^{b-1}}{(d-c)B(a, b)} < \infty.$$

Donc, en vertu de l'algorithme d'acceptation-rejet, si

$$V \leq \frac{1}{(d-c)^{a+b-2}} \left( \frac{X-c}{\delta} \right)^{a-1} \left( \frac{d-X}{1-\delta} \right)^{b-1},$$

alors  $X$  a la densité  $f(x)$  par (19).

## Bibliographie

- Cochran, W.G. (1977). *Sampling Techniques*, troisième édition. New York : John Wiley & Sons, Inc.
- Gilks, W.R., et Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society, Séries C*, 41, 337-348.
- Gelman, A. (2006). Prior distribution for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515-533.
- Ghosh, M., Natarajan, K., Stroud, T.W.F. et Carlin, B.P. (1998). Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, 93, 273-282.
- Hillmer, S.C., et Trabelsi, A. (1987). Benchmarking of economic time series. *Journal of the American Statistical Association*, 82, 1064-1071.
- Lazar, R., Meeden, G. et Nelson, D. (2008). Une approche bayésienne non informative de l'échantillonnage d'une population finie en utilisant des variables auxiliaires. *Techniques d'enquête*, 34, 55-70.
- Nandram, B. (1998). A Bayesian analysis of the three-stage hierarchical multinomial model. *Journal of Statistical Computation and Simulation*, 61, 97-126.
- Nandram, B., et Choi, J.W. (2002). Hierarchical Bayesian nonresponse models for binary data from small areas with uncertainty about ignorability. *Journal of the American Statistical Association*, 97, 381-388.
- Nandram, B., et Choi, J.W. (2002). A Bayesian analysis of a proportion under non-ignorable non-response. *Statistics in Medicine*, 21, 9, 1189-1212.
- Nandram, B., et Sedransk, J. (1993). Bayesian predictive inference for a finite population proportion: Two-stage cluster sampling. *Journal of the Royal Statistical Society, Séries B*, 55, 399-408.
- Nandram, B., Toto, M.C.S. et Choi, J.W. (2011). A Bayesian benchmarking of the Scott-Smith model for small areas. *Journal of Statistical Computation and Simulation* (en cours d'impression, préimprimez).
- Rao, J.N.K. (2003). *Small Area Estimation*. New York : John Wiley & Sons, Inc.
- Ritter, C., et Tanner, M.A. (1992). The gibbs sampler and the griddy gibbs sampler. *Journal of the American Statistical Association*, 87, 861-868.
- Robert, C.P., et Casella, G. (1999). *Monte Carlo Statistical Methods*. New York : Springer-Verlag.
- Silvapulle, M.J., et Sen, P.K. (2006). *Constrained Statistical Inference: Inequality, Order and Shape Restrictions*. New York : John Wiley & Sons, Inc.
- Silverman, B.W. (1986). *Density Estimation*. Londres : Chapman and Hall.