

N° 12-001-X au catalogue  
ISSN 1712-5685

## Techniques d'enquête

# Inférence bayésienne pour un modèle des composantes de la variance fondée sur la vraisemblance composite par paire à partir des données d'enquête

par Mary E. Thompson, Joseph Sedransk, Junhan Fang  
et Grace Y. Yi

Date de diffusion : le 21 juin 2022



Statistique  
Canada

Statistics  
Canada

Canada

---

## Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à [www.statcan.gc.ca](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

**Courriel** à [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Téléphone** entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- |   |                |
|---|----------------|
| • Service de renseignements statistiques                                    | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur   | 1-514-283-9350 |

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site [www.statcan.gc.ca](http://www.statcan.gc.ca) sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

## Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2022

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

*This publication is also available in English.*

---

# Inférence bayésienne pour un modèle des composantes de la variance fondée sur la vraisemblance composite par paire à partir des données d'enquête

Mary E. Thompson, Joseph Sedransk, Junhan Fang et Grace Y. Yi<sup>1</sup>

## Résumé

Nous considérons un modèle linéaire à effets aléatoires avec ordonnée à l'origine seulement pour l'analyse de données à partir d'un plan d'échantillonnage en grappes à deux degrés. Au premier degré, un échantillon aléatoire simple de grappes est tiré et au deuxième degré, un échantillon aléatoire simple d'unités élémentaires est prélevé dans chaque grappe sélectionnée. La variable de réponse est censée se composer d'un effet aléatoire au niveau de la grappe en plus d'un terme d'erreurs indépendant dont la variance est connue. Les objets de l'inférence sont la moyenne de la variable de résultat et la variance des effets aléatoires. Dans le cas d'un plan d'échantillonnage à deux degrés plus complexe, l'utilisation d'une approche fondée sur une fonction de vraisemblance composite par paire estimée présente des propriétés attrayantes. Notre objectif est d'utiliser notre cas plus simple pour comparer les résultats de l'inférence fondée sur la vraisemblance à ceux de l'inférence fondée sur une fonction de vraisemblance composite par paire qui est considérée comme une vraisemblance approximative, et en particulier comme une composante de la vraisemblance dans l'inférence bayésienne. Afin de fournir des intervalles de crédibilité ayant une couverture fréquentiste se rapprochant des valeurs nominales, la fonction de vraisemblance composite par paire et la densité *a posteriori* correspondante nécessitent une modification, comme un ajustement de la courbure. Au moyen d'études par simulations, nous étudions le rendement d'un ajustement proposé dans la littérature et concluons que ce dernier fonctionne bien pour la moyenne, mais fournit des intervalles de crédibilité pour la variance des effets aléatoires présentant un sous-couverture. Nous proposons des orientations futures possibles, y compris des prolongements du cas d'un plan d'échantillonnage complexe.

**Mots-clés :** Analyse d'échantillons en grappes; vraisemblance composite; ajustement de la courbure; modèle à effets aléatoires.

## 1. Introduction

Les plans d'échantillonnage à plusieurs degrés sont utilisés dans de nombreuses enquêtes menées à l'échelle de la population. De plus en plus, des modèles multiniveaux sont utilisés pour faire des inférences lorsque les données sont obtenues à partir d'une enquête à plusieurs degrés.

Dans le but d'améliorer ces inférences, Rao, Verret et Hidioglou (2013) ont proposé d'utiliser une approche fondée sur le logarithme du rapport de vraisemblance composite par paire pondéré. Il existe une abondante littérature sur les vraisemblances composites : voir les articles de Varin (2008), de Varin, Reid et Firth (2011) et de Yi (2017), ainsi que de nombreuses applications. À la section 4 de leur article, Rao, Verret et Hidioglou décrivent une approche unifiée applicable aux modèles linéaires et aux modèles linéaires généralisés. Des aspects importants de leurs travaux comprennent a) l'obtention d'estimations ponctuelles convergentes par rapport au plan d'échantillonnage des paramètres de moyenne et de régression et des composantes de la variance et b) l'utilisation de probabilités d'inclusion de premier ordre et de probabilités de deuxième ordre dans les grappes seulement. En particulier, les travaux de Rao, Verret

1. Mary E. Thompson, Département des statistiques et de l'actuariat, Université de Waterloo, Waterloo (Ontario) N2L 3G1, Canada. Courriel : methompson@uwaterloo.ca; Joseph Sedransk, Programme conjoint de méthodologie d'enquête, Université du Maryland, College Park, MD 20742, É.-U.; Junhan Fang, École de médecine, Université Yale, New Haven, CT 06520, É.-U.; Grace Y. Yi, Département des statistiques et de l'actuariat, Département d'informatique, Université Western, Londres (Ontario) N6A 5B7, Canada.

et Hidioglou à l'égard de a) sont importants en raison de la convergence par rapport au plan d'échantillonnage lorsque le nombre de grappes (unités de premier degré) augmente tandis que la taille des grappes demeure petite (Pfeffermann, Skinner, Holmes, Goldstein et Rasbash, 1998). Contrairement à l'approche de pseudo-vraisemblance couramment utilisée (Rabe-Hesketh et Skrondal, 2006), leur méthode permet de s'assurer que a) vaut pour les résultats des modèles linéaires généralisés. Les travaux de Rao, Verret et Hidioglou ont été approfondis par Yi, Rao et Li (2016), qui fournissent un cadre plus général, des considérations théoriques supplémentaires et des simulations exhaustives.

Deux phénomènes connexes ont donné lieu à nos travaux. D'une part, on s'intéresse de plus en plus à l'utilisation des méthodes bayésiennes pour faire des inférences à partir de données d'enquêtes. La section 5 présente une référence générale ainsi qu'une introduction à des articles décrivant une utilisation abondante de méthodes bayésiennes au Service national de la statistique agricole du ministère américain de l'Agriculture. D'autre part, il existe des écrits (bayésiens) démontrant la possibilité d'une précision surestimée en utilisant des vraisemblances composites *non ajustées*, par exemple Ribatet, Cooley et Davison (2012) et Stoehr et Friel (2018).

Notre approche consiste à utiliser d'abord une loi *a posteriori* considérée comme proportionnelle au produit d'une vraisemblance composite et d'une loi *a priori*. En comparant cette loi *a posteriori* approximative à une autre qui repose sur la vraisemblance complète, nous démontrons que les inférences fondées sur la loi *a posteriori* approximative affichent une précision surestimée. En apportant des ajustements à la loi *a posteriori* en fonction de la vraisemblance composite comme c'est le cas dans Ribatet, Cooley et Davison, nous utilisons ensuite des simulations pour comparer les trois façons de formuler une loi *a posteriori*, c'est-à-dire celles reposant sur la vraisemblance composite complète, composite et ajustée. Nous procédons, pour ce faire, à un examen visuel des graphiques des densités et des couvertures *a posteriori* (simulations répétées) de 95 % des intervalles de crédibilité pour les paramètres du modèle.

La méthodologie est décrite à la section 2.3. Les ajustements à la loi *a posteriori* approximative en fonction d'une vraisemblance composite sont dérivés d'une transformation du logarithme de la vraisemblance composite à son mode, conçue de sorte que la négative de l'inverse de la matrice de courbure de la densité *a posteriori* approximative à son mode soit proportionnelle à la matrice de variance-covariance *a posteriori* correspondante des paramètres. Cette approche est semblable à la propriété dans l'inférence fréquentiste faisant que l'inverse de la matrice d'information de Fisher observée (la négative de la matrice hessienne du logarithme du rapport de vraisemblance à son mode) permet d'estimer la matrice de variance-covariance des estimations du maximum de vraisemblance.

Afin de traiter de la question principale, nous utilisons une loi *a priori* « non informative » pour les paramètres de notre modèle, décrite ci-dessous. La densité *a posteriori* correspondante se rapproche alors de la vraisemblance normalisée, et les progrès démontrés dans un contexte bayésien seraient aussi observés dans une approche fréquentiste fondée sur des modèles.

Pour simplifier l'étude initiale, nous supposons un modèle de superpopulation linéaire à effets aléatoires simple (ordonnée à l'origine seulement). Examinons une population cible tirée de cette superpopulation et composée d'un grand nombre  $N$  de grappes, chacune ayant une taille commune de, disons,  $m$ . Soit  $Y_{ij}$  la variable de réponse continue pour l'unité élémentaire  $j$  dans la grappe  $i$  avec  $i = 1, \dots, N$  et  $j = 1, \dots, m$ . Nous utilisons la notation

$$Y_{ij} = \theta + u_i + e_{ij} \quad (1.1)$$

où  $u_i \sim N(0, \sigma_u^2)$ ,  $e_{ij} \sim N(0, \sigma_e^2)$ , tous les  $u_i$  et les  $e_{ij}$  sont indépendants et  $\theta$ ,  $\sigma_u$  et  $\sigma_e$  sont des paramètres.

Nous commençons aussi en supposant que le plan d'échantillonnage est un échantillon aléatoire simple de  $n$  grappes, où  $n$  est un nombre entier positif. Cela présente l'avantage que le modèle (1.1) vaut non seulement pour la superpopulation et la population finie, mais également (en remplaçant  $N$  par  $n$ ) pour l'échantillon lui-même, qui résulte de la génération de la population suivie de la sélection de l'échantillon à l'aide du plan d'échantillonnage. Cela permet de s'assurer que la fonction de vraisemblance à utiliser dans l'inférence bayésienne est bien définie. De plus, nous pouvons démontrer que l'inférence bayésienne de l'échantillon pour les paramètres du modèle (1.1) pourrait aussi être interprétée du point de vue de la théorie fréquentiste dans les utilisations analytiques des données d'enquête (Skinner, Holt et Smith, 1989).

Nos travaux sont utiles, car nous démontrons les dangers d'utiliser une vraisemblance composite par paire *non ajustée* pour former une loi *a posteriori* approximative d'inférence même dans ce cas très simple et direct. Des prolongements des plans d'échantillonnage avec probabilités inégales sont examinés à la section 4.

L'ajustement proposé donne lieu à d'excellentes propriétés fréquentistes pour l'inférence sur la moyenne  $\theta$ . La moyenne *a posteriori* de  $\theta$  est peu biaisée sur le plan de la fréquence, et la couverture fréquentiste des intervalles de crédibilité s'harmonise aux niveaux nominaux. Pour  $\sigma_u$ , cette approche donne lieu à une amélioration considérable par rapport à l'utilisation de la vraisemblance composite non ajustée. Cependant, la couverture est inférieure au niveau nominal, de sorte que d'autres travaux sur la manière de modifier l'ajustement sont nécessaires.

Le reste du présent document est structuré de la manière suivante. La section 2 fournit les définitions de la vraisemblance composite complète, composite et ajustée et des lois *a priori*. Une description de l'ajustement de la courbure et des raisons de son utilisation est ensuite présentée. Les études par simulations sont décrites à la section 3, y compris le modèle, les lois *a priori*, les tailles d'échantillon et leurs paramètres, le nombre de répliques, etc. Cette section décrit également un tableau de la manière dont les résultats sont exposés et présente un résumé de nos constatations. Des prolongements des cas d'échantillonnage avec probabilités inégales sont examinés à la section 4. Les conclusions sont présentées à la section 5.

## 2. Vraisemblance complète, vraisemblance par paire et mise en œuvre bayésienne

### 2.1 Modèle et formule

Comme à la section 1, supposons que  $Y_{ij}$  désigne la variable de réponse pour l'unité de deuxième degré  $j$  dans l'unité de premier degré  $i$  pour  $i=1, \dots, n$ , et  $j=1, \dots, m$ . Nous utilisons la lettre minuscule  $y_{ij}$  pour représenter les valeurs réalisées de  $Y_{ij}$ . Supposons que  $\mathbf{y}(n) = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  désigne les données d'échantillon avec  $\mathbf{y}_i = (y_{i1}, \dots, y_{im})^T$  pour  $i=1, \dots, n$ , où T désigne la transposée.

Dans un modèle à effets aléatoires plus général, nous pourrions supposer que, en fonction des effets aléatoires  $u_i$  pour  $i=1, \dots, n$ , les  $Y_{ij}$  sont distribuées indépendamment comme suit :

$$Y_{ij} \sim f_{y|u}(y_{ij} | u_i; \boldsymbol{\theta}_y) \quad \text{pour } j=1, \dots, m, \quad (2.1)$$

où  $f_{y|u}$  est une fonction de densité connue et  $\boldsymbol{\theta}_y$  est le vecteur de paramètres connexe. Ensuite, nous modélisons les effets aléatoires en supposant que les  $u_i$  sont indépendants et identiquement distribués comme suit :

$$u_i \sim f_u(u_i | \boldsymbol{\theta}_u) \quad \text{pour } i=1, \dots, n, \quad (2.2)$$

où  $f_u$  est une fonction de densité connue indexée par le vecteur de paramètres  $\boldsymbol{\theta}_u$ .

Soit  $\boldsymbol{\eta} = (\boldsymbol{\theta}_y^T, \boldsymbol{\theta}_u^T)^T$  le vecteur des paramètres du modèle d'intérêt. Dans le cadre fréquentiste, la méthode du maximum de vraisemblance est couramment utilisée pour faire des inférences au sujet de  $\boldsymbol{\eta}$  en maximisant la fonction de vraisemblance

$$L(\boldsymbol{\eta}) = \prod_{i=1}^n f(\mathbf{y}_i; \boldsymbol{\eta}),$$

où

$$f(\mathbf{y}_i; \boldsymbol{\eta}) = \int \left\{ \prod_{j=1}^{m_i} f_{y|u}(y_{ij} | u_i; \boldsymbol{\theta}_y) \right\} f_u(u_i | \boldsymbol{\theta}_u) du_i. \quad (2.3)$$

Une solution de rechange à la méthode de vraisemblance est l'approche fondée sur la vraisemblance composite (Lindsay, 1988). Plus particulièrement, la méthode fondée sur la vraisemblance par paire a souvent été employée. Soit  $L_{ij}(\boldsymbol{\eta}) = f(y_{ij}; \boldsymbol{\eta})$  la densité de  $Y_{ij}$ , déterminée au moyen de

$$f(y_{ij}; \boldsymbol{\eta}) = \int f_{y|u}(y_{ij} | u_i; \boldsymbol{\theta}_y) f_u(u_i | \boldsymbol{\theta}_u) du_i.$$

Pour  $j \neq k$ , soit  $L_{ijk}(\boldsymbol{\eta}) = f(y_{ij}, y_{ik}; \boldsymbol{\eta})$  la densité conjointe pour les réponses appariées  $(Y_{ij}, Y_{ik})$ , déterminée au moyen de

$$f(y_{ij}, y_{ik}; \boldsymbol{\eta}) = \int f_{y|u}(y_{ij} | u_i; \boldsymbol{\theta}_y) f_{y|u}(y_{ik} | u_i; \boldsymbol{\theta}_y) f_u(u_i | \boldsymbol{\theta}_u) du_i.$$

Une fonction de vraisemblance par paire marginale peut alors être formulée comme suit :

$$C(\boldsymbol{\eta}) = \prod_{i=1}^n \prod_{j < k} L_{ijk}^{d_{jk}}(\boldsymbol{\eta}) \times L_{ij}^{d_j}(\boldsymbol{\eta}) \times L_{ik}^{d_k}(\boldsymbol{\eta}),$$

où  $d_{jk}$ ,  $d_j$  et  $d_k$  sont des poids qui peuvent être précisés par l'utilisateur pour accroître l'efficacité ou faciliter certains aspects précis de la formulation. Une discussion portant sur le choix des poids figure dans Cox et Reid (2004), Lindsay, Yi et Sun (2011), Varin, Reid et Firth (2011), et Yi (2017). Pour limiter notre attention à l'utilisation de vraisemblances par paire marginales, conformément à l'approche de Rao, Verret et Hidiroglou, nous examinons ici le cas avec  $d_j = d_k = 0$  et  $d_{jk} = 1$ .

Si nous revenons au cas spécial du modèle (1.1), supposons que  $\sigma_e^2$  est connue et prenons  $\boldsymbol{\eta}$  qui est formé de  $\boldsymbol{\theta}_y = \theta$  et de  $\boldsymbol{\theta}_u = \sigma_u^2$ . Selon une approche bayésienne, il est nécessaire de choisir une loi *a priori* pour  $\boldsymbol{\eta}$ . Nous supposons une loi *a priori* dans laquelle  $\theta$  et  $\sigma_u^2$  sont indépendantes, avec une loi uniforme soutenant largement  $\theta$  et une loi pour  $\sigma_u$  qui est presque uniforme dans un intervalle qui est censé contenir le soutien de la fonction de vraisemblance complète pour  $\sigma_u^2$  avec une forte probabilité. Gelman (2006) présente un traitement rigoureux pour choisir une loi *a priori* de  $\sigma_u$  dans le modèle à effets aléatoires (1.1). Il recommande d'utiliser une loi *a priori* uniforme pour  $\sigma_u$  pour des valeurs modérées à grandes de  $n$ , mais une loi *a priori* demi-Cauchy pour de faibles valeurs de  $n$  (voir, en particulier, les sections 3.2 et 5.2 de Gelman, 2006). La loi *a priori* demi-Cauchy est appuyée sur  $(0, \infty)$  et est donnée par :

$$\pi(\sigma_u) \propto \left( 1 + \left( \frac{\sigma_u}{A} \right)^2 \right)^{-1}, \quad (2.4)$$

où  $A$  est un hyperparamètre d'échelle.

## 2.2 Vraisemblance composite par paire non ajustée

Prenons encore une fois le modèle (1.1) et, en supposant que  $\sigma_e^2$  est connue, soit  $\boldsymbol{\eta} = (\theta, \sigma_u^2)^T$  le vecteur des paramètres du modèle. Nous voulons comparer le rendement de la loi *a posteriori* de  $\boldsymbol{\eta}$  en fonction de l'utilisation de la vraisemblance complète (VC) ou de la vraisemblance par paire (VP), de même que de la loi *a posteriori* de la vraisemblance par paire ajustée décrite ci-dessous.

D'abord, considérons une situation simple où l'on suppose également que  $\sigma_u^2$  est connue et que seule  $\theta$  est inconnue. Soit  $\pi(\theta)$  la densité *a priori* de  $\theta$ . La densité *a posteriori* de  $\theta$  est donc

$$p_{VC}(\theta | \mathbf{y}(n)) \propto \pi(\theta) \prod_{i=1}^n f(\mathbf{y}_i; \theta), \quad (2.5)$$

où l'indice VC indique qu'elle repose sur la vraisemblance complète. En revanche, nous considérons

$$L_{i,VP}(\theta) = \prod_{1 \leq j < k \leq m} L_{ijk}(\theta),$$

où  $L_{ijk}(\theta) = \int f_{y|u}(y_{ij}|u_i; \theta) f_{y|u}(y_{ik}|u_i; \theta) f_u(u_i) du_i$ , puis nous définissons

$$p_{\text{VP}}(\theta|\mathbf{y}(n)) \propto \pi(\theta) \prod_{i=1}^n L_{i, \text{VP}}(\theta) \quad (2.6)$$

comme la densité *a posteriori* « par paire » de  $\theta$ . Nous voulons comparer les variances de  $\theta$  dérivées de  $p_{\text{VC}}(\theta|\mathbf{y}(n))$  et de  $p_{\text{VP}}(\theta|\mathbf{y}(n))$ , démontrées dans le théorème suivant, dont les calculs sont simples.

**Théorème :** *Supposons que  $\pi(\theta)$  est une loi a priori uniforme. Alors*

- (a)  $p_{\text{VC}}(\theta|\mathbf{y}(n))$  est une densité normale avec une moyenne  $\bar{y}$  et une variance  $\frac{\sigma_e^2 + m\sigma_u^2}{mn}$ ;
- (b)  $p_{\text{VP}}(\theta|\mathbf{y}(n))$  est une densité normale avec une moyenne  $\bar{y}$  et une variance  $\frac{\sigma_e^2 + 2\sigma_u^2}{m(m-1)n}$  où  $\bar{y} = (mn)^{-1} \sum_{i=1}^n \sum_{j=1}^m y_{ij}$ .

Le théorème démontre que, lorsque  $m$  est supérieure à 2, la variance dérivée de la densité *a posteriori* « par paire »  $p_{\text{VP}}(\theta|\mathbf{y}(n))$  est inférieure à celle de la densité *a posteriori*  $p_{\text{VC}}(\theta|\mathbf{y}(n))$ . Cette constatation semble raisonnable, car la vraisemblance par paire suppose dans les faits que toutes les paires d'observations  $m(m-1)/2$  dans chaque grappe sont indépendantes. Cela nous amène à nous pencher sur une version ajustée de  $p_{\text{PL}}(\theta|\mathbf{y}(n))$ , qui sera examinée par la suite.

Dans le cas où  $\sigma_u^2$  est également inconnue, on peut démontrer qu'un type d'ajustement semblable est nécessaire. En supposant des lois *a priori* uniformes indépendantes pour  $\theta$  et  $\sigma_u^2$ , il est simple de démontrer que

$$p_{\text{VC}}(\theta, \sigma_u^2 | \mathbf{y}(n)) \propto |\Sigma_m|^{-n/2} \exp[-0,5 \text{tr}(\Sigma_m^{-1} \mathbf{S}_0)] \quad (2.7)$$

où  $\mathbf{S}_0 = \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu}_m)(\mathbf{y}_i - \boldsymbol{\mu}_m)^T$ ,  $\boldsymbol{\mu}_m = \theta \mathbf{1}_m$ ,  $\Sigma_m = \sigma_e^2 \mathbf{I}_m + \sigma_u^2 \mathbf{1}_m \mathbf{1}_m^T$ ,  $\mathbf{1}_m$  représente le vecteur unitaire  $m \times 1$  et  $\mathbf{I}_m$  désigne la matrice d'identité  $m \times m$ .

Après quelques calculs algébriques, on peut démontrer que la loi *a posteriori* de la vraisemblance composite par paire (VP) est

$$p_{\text{VP}}(\theta, \sigma_u^2 | \mathbf{y}(n)) \propto |\Sigma_2|^{nm(m-1)/4} \exp[-0,5 \text{tr}(\Sigma_2^{-1} \mathbf{S}_{0\text{VP}})] \quad (2.8)$$

où, avec  $\mathbf{z}_{ijk} = (y_{ij} - \theta, y_{ik} - \theta)^T$ ,

$$\mathbf{S}_{0\text{VP}} = \sum_{i=1}^n \sum_{j < k} \mathbf{z}_{ijk} \mathbf{z}_{ijk}^T.$$

Il importe de souligner que  $\Sigma_2$  est définie dans l'équation (2.7) par  $m = 2$ .

En supposant des lois *a priori* uniformes indépendantes pour  $\theta$  et  $\sigma_u^2$ , nous considérons la densité *a posteriori* de  $\sigma_u^2$  avec  $\theta$  éliminée par intégration. Pour évaluer les précisions relatives de l'inférence bayésienne dans les deux cas, nous devons utiliser des approximations en raison de la complexité des deux densités *a posteriori*. Plus précisément, nous comparons la courbure de la log-densité *a posteriori* et de la



log-densité *a posteriori* par paire pour  $\sigma_u^2$  avec leurs modes. On peut démontrer que le rapport entre le dernier et le premier est égal pour un grand  $n$  à

$$\frac{2(m-1)(\sigma_e^2 + m\sigma_u^2)^2}{m(\sigma_e^2 + 2\sigma_u^2)^2},$$

ce qui laisse croire que la densité *a posteriori* par paire non ajustée pour  $m > 2$  surestimerait la précision de l'estimation de  $\sigma_u^2$ .

Ainsi, pour  $\theta$  et  $\sigma_u^2$  (ou  $\sigma_u$ ), le fait de fonder un logarithme du rapport de vraisemblance approximatif pour l'inférence bayésienne directement sur la vraisemblance composite par paire mènerait à des intervalles *a posteriori* qui sont trop étroits.

Note : À la section 3, le vecteur de paramètres  $\boldsymbol{\eta}$  correspond à  $(\theta, \sigma_u)^T$  (la variance  $\sigma_u^2$  étant remplacée par l'écart-type  $\sigma_u$ ), et une loi *a priori* demi-Cauchy est utilisée pour  $\sigma_u$ . Cependant, la comparaison de la densité *a posteriori* complète et de la log-densité *a posteriori* par paire demeurera semblable lorsque les transformations appropriées seront mises en application.

### 2.3 Ajustement de la courbure pour le logarithme du rapport de vraisemblance par paire

Dans la présente section, nous justifions l'ajustement de la courbure du logarithme du rapport de vraisemblance par paire du point de vue de la théorie des fonctions d'estimation, telle qu'elle est présentée, par exemple, par Jørgensen et Knudsen (2004).

D'abord, nous soulignons que si  $\mathbf{X}$  a une distribution normale  $q$ -variée comportant un vecteur moyen  $\boldsymbol{\mu}$  et une matrice de variance-covariance  $\boldsymbol{\Sigma}$ , le logarithme de la densité multivariée de  $\mathbf{X}$  prend la forme suivante :

$$-\frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad (2.9)$$

L'expression dans l'équation (2.9) sous forme de fonction de  $\mathbf{x}$  est à son maximum à  $\boldsymbol{\mu}$  et la courbure ou la matrice des dérivées secondes (hessienne) est au maximum égal à  $-\boldsymbol{\Sigma}^{-1}$ . Intuitivement, on peut s'attendre à ce que cette correspondance entre la courbure de la log-densité au maximum et l'inverse de la matrice de covariance soit vérifiée approximativement pour une densité multivariée qui est presque normale.

Considérons un modèle dans lequel la distribution de la variable d'observation  $\mathbf{Y}(n)$  dépend d'un paramètre vectoriel  $\boldsymbol{\eta}$ . Soit une observation  $\mathbf{Y}(n) = \mathbf{y}(n)$ , le logarithme du rapport de vraisemblance est désigné  $\ell(\boldsymbol{\eta}; \mathbf{y}(n)) = \log(f(\mathbf{y}(n); \boldsymbol{\eta}))$  où  $f$  est la densité de  $\mathbf{Y}(n)$ . Sous des conditions de régularité (par exemple Lehmann, 1999, chapitre 7), l'EMV  $\hat{\boldsymbol{\eta}}$  est calculée en résolvant le système

$$\mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n)) = \mathbf{0}, \quad (2.10)$$

où  $\mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n))$  désigne la fonction de score, le gradient de  $\ell(\boldsymbol{\eta}; \mathbf{y}(n))$ . Le système (2.10) est une équation d'estimation (vectorielle) sans biais et a une efficacité optimale, présentant une matrice de variance-covariance asymptotique minimale (du point de vue de la différence définie positive) parmi les solutions des systèmes d'équations d'estimation sans biais. Dans les cas ordinaires (par exemple Lehmann, 1999, chapitre 7), la fonction de score satisfait à la *deuxième identité de Bartlett* (par exemple Lindsay, 1988) :

$$\text{Var}_{\boldsymbol{\eta}}[\mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n))] = -E_{\boldsymbol{\eta}}[\nabla \mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n))] = -E_{\boldsymbol{\eta}}[\nabla^2 \ell(\boldsymbol{\eta}; \mathbf{y}(n))], \quad (2.11)$$

où  $\text{Var}$  désigne une matrice de variance-covariance et  $\nabla$  représente un gradient. De plus, asymptotiquement, au moyen d'une approximation par série de Taylor de  $\mathbf{s}(\hat{\boldsymbol{\eta}}; \mathbf{y}(n)) - \mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n)) = \mathbf{0} - \mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n))$ , nous avons :

$$\hat{\boldsymbol{\eta}} - \boldsymbol{\eta} \simeq -[\nabla \mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n))]^{-1} \mathbf{s}(\boldsymbol{\eta}; \mathbf{y}(n)). \quad (2.12)$$

Ainsi, l'inférence fondée sur la vraisemblance (fréquentiste) standard permet d'estimer la variance-covariance de  $\hat{\boldsymbol{\eta}}$  comme la réciproque de la matrice d'information de Fisher observée :

$$\mathbf{I} = -\frac{\partial^2}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \ell(\boldsymbol{\eta}; \mathbf{y}(n)) \Big|_{\hat{\boldsymbol{\eta}}} = -\nabla^2 \ell(\boldsymbol{\eta}; \mathbf{y}(n)) \Big|_{\hat{\boldsymbol{\eta}}}, \quad (2.13)$$

qui est la négative de la matrice hessienne (matrice de courbure) de la fonction du logarithme du rapport de vraisemblance à son maximum.

Dans le cas de l'inférence bayésienne, si  $\pi(\boldsymbol{\eta})$  est une densité *a priori* pour  $\boldsymbol{\eta}$ , le logarithme de la densité *a posteriori* pour  $\boldsymbol{\eta}$  est

$$\log \pi(\boldsymbol{\eta} | \mathbf{y}(n)) = \log \pi(\boldsymbol{\eta}) + \ell(\boldsymbol{\eta}; \mathbf{y}(n)) - K(\mathbf{y}(n)), \quad (2.14)$$

où

$$K(\mathbf{y}(n)) = \log \left\{ \int \pi(\boldsymbol{\eta}) f(\mathbf{y}(n); \boldsymbol{\eta}) d\boldsymbol{\eta} \right\}.$$

Si la densité *a priori* est plane dans les zones de vraisemblance appréciable, la densité *a posteriori* de  $\boldsymbol{\eta}$ , qui quantifie l'inférence au sujet de  $\boldsymbol{\eta}$ , correspond à une densité ayant un mode égal à  $\hat{\boldsymbol{\eta}}$  et la courbure de son logarithme est égale à la négative de la matrice d'information de Fisher, ce qui fait en sorte que la variance-covariance *a posteriori* de  $\hat{\boldsymbol{\eta}}$  est approximativement égale à la réciproque de  $\mathbf{I}$  dans l'équation (2.13). Ainsi, l'estimation bayésienne de  $\boldsymbol{\eta}$  est efficace du point de vue fréquentiste; autrement, l'inférence fréquentiste se rapproche de l'inférence bayésienne.

Supposons que, dans le contexte fréquentiste, la fonction de score est remplacée par une autre fonction d'estimation  $\mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})$  qui est sans biais dans le sens où elle a une espérance nulle. Voir, par exemple,

Lindsay, Yi et Sun (2011). L'estimateur  $\hat{\boldsymbol{\eta}}$  n'a alors plus une efficacité optimale. Cependant, il est convergent, et sa variance peut être estimée au moyen de la méthode delta ou de la linéarisation de la fonction  $\mathbf{g}$ . Nous pourrions vouloir considérer  $\mathbf{g}$  en remplacement d'un vecteur de score ou comme le gradient à l'égard de  $\boldsymbol{\eta}$  d'un substitut de la fonction de logarithme du rapport de vraisemblance. En particulier, on pourrait considérer les équations fondées sur la vraisemblance composite en remplacement des équations d'estimation de score.

Une question qui se pose est alors celle de savoir si un substitut de la fonction du logarithme du rapport de vraisemblance comportant le gradient  $\mathbf{g}$  pourrait jouer le rôle du logarithme du rapport de vraisemblance dans l'inférence bayésienne et mener à une loi *a posteriori* approximativement exacte dans l'équation (2.14) et, dans la négative, s'il existe des moyens fondés sur des principes de le corriger.

Ainsi, supposons que nous avons une solution de rechange à la fonction de score, à savoir la fonction d'estimation  $\mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})$ , qui est sans biais pour  $\boldsymbol{\eta}$  en ce sens que :

$$E_{\boldsymbol{\eta}}[\mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})] = \mathbf{0}.$$

Supposons que la solution  $\hat{\boldsymbol{\eta}}$  de l'équation  $\mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta}) = \mathbf{0}$  maximise une fonction  $h(\mathbf{y}(n); \boldsymbol{\eta})$  que nous voudrions considérer comme une solution de rechange à la fonction du logarithme du rapport de vraisemblance; par exemple,  $h(\mathbf{y}(n); \boldsymbol{\eta})$  pourrait être une fonction du logarithme du rapport de vraisemblance composite par paire et  $\mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta}) = \nabla h(\mathbf{y}(n); \boldsymbol{\eta})$ . Alors  $h(\mathbf{y}(n); \boldsymbol{\eta})$  serait approximativement égale à la valeur qu'aurait la log-densité *a posteriori* si la loi *a priori* était non informative et si nous considérions  $h(\mathbf{y}(n); \boldsymbol{\eta})$  comme étant un substitut de la fonction du logarithme du rapport de vraisemblance. La variance-covariance *a posteriori* substitut de  $\boldsymbol{\eta}$  serait approximativement l'inverse de la négative de la matrice de courbure de  $h(\mathbf{y}(n); \boldsymbol{\eta})$  à  $\hat{\boldsymbol{\eta}}$ . Selon la théorie des fonctions d'estimation (par exemple Heyde, 1997), si nous utilisons le même type d'approximation par série de Taylor que dans l'équation (2.12), la variance-covariance fréquentiste de  $\hat{\boldsymbol{\eta}}$  correspond à :

$$\text{Var}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\eta}}^T) \simeq \left\{ E_{\boldsymbol{\eta}}[\nabla \mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})] \right\}^{-1} \text{Var}_{\boldsymbol{\eta}}[\mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})] \left\{ E_{\boldsymbol{\eta}}[\nabla \mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})]^T \right\}^{-1}. \quad (2.15)$$

Si  $h(\mathbf{y}; \boldsymbol{\eta})$  était la fonction du logarithme du rapport de vraisemblance composite par paire, nous obtiendrions, selon la notation de Ribatet, Cooley et Davison :

$$\text{Var}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\eta}}^T) \simeq \frac{1}{n} \left[ \mathbf{H}(\boldsymbol{\eta}_0) \mathbf{J}(\boldsymbol{\eta}_0)^{-1} \mathbf{H}(\boldsymbol{\eta}_0) \right]^{-1}, \quad (2.16)$$

où  $\boldsymbol{\eta}_0$  est la valeur réelle de  $\boldsymbol{\eta}$ ,  $n\mathbf{H}(\boldsymbol{\eta}_0)$  est inférieure à l'espérance de  $\nabla h$  et  $n\mathbf{J}(\boldsymbol{\eta}_0)$  est égale à la matrice de variance-covariance de  $\mathbf{g}$ , le gradient de  $h$ .

Si  $\mathbf{g}$  avait la propriété (analogue à l'équation [2.11]) suivante :

$$\text{Var}_{\boldsymbol{\eta}}[\mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})] = -E_{\boldsymbol{\eta}}[\nabla \mathbf{g}(\mathbf{y}(n); \boldsymbol{\eta})], \quad (2.17)$$

de sorte que  $\mathbf{J}(\boldsymbol{\eta}_0) = -\mathbf{H}(\boldsymbol{\eta}_0)$ , le côté droit de l'équation (2.15) ou de l'équation (2.16) serait alors approximativement le même que la variance-covariance *a posteriori* substitut de  $\boldsymbol{\eta}$ .

La propriété (2.17) est appelée *l'absence de biais d'information* d'une fonction d'estimation (Lindsay, 1982). Soit un  $\mathbf{g}$  qui ne satisfait pas l'équation (2.17), pour produire un  $\mathbf{g}^*$  qui correspond approximativement à l'équation (2.17), nous pourrions alors établir

$$h^*(\mathbf{y}(n); \boldsymbol{\eta}) = h(\mathbf{y}(n); \hat{\boldsymbol{\eta}} + \mathbf{C}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}})) = h(\mathbf{y}(n); \boldsymbol{\eta}^*) \quad (2.18)$$

pour une matrice constante  $\mathbf{C}$ , de sorte que le gradient de  $h^*$  soit  $\mathbf{C}^T$  fois le gradient de  $h$ , tandis que l'estimation ponctuelle de  $\boldsymbol{\eta}$  qui maximise  $h^*$  et sa variance-covariance approximative demeurent inchangées.

Nous voulons obtenir  $\text{Var}_{\boldsymbol{\eta}}(\mathbf{g}^*) = -E_{\boldsymbol{\eta}} \nabla \mathbf{g}^*$ , et il peut être démontré que cela équivaut à

$$\mathbf{H}(\boldsymbol{\eta}_0) \mathbf{J}(\boldsymbol{\eta}_0)^{-1} \mathbf{H}(\boldsymbol{\eta}_0) = \mathbf{C}^T \mathbf{H}(\boldsymbol{\eta}_0) \mathbf{C}, \quad (2.19)$$

qui est un *ajustement de la courbure* comme celui présenté dans l'étude de Ribatet, Cooley et Davison, où il est suggéré de prendre la solution de l'équation (2.19) qui établit  $\mathbf{C} = \mathbf{M}^{-1} \mathbf{M}_A$ , où  $\mathbf{M}_A^T \mathbf{M}_A = \mathbf{H}(\boldsymbol{\eta}_0) \mathbf{J}(\boldsymbol{\eta}_0)^{-1} \mathbf{H}(\boldsymbol{\eta}_0)$  et  $\mathbf{M}^T \mathbf{M} = \mathbf{H}(\boldsymbol{\eta}_0)$ .

### 3. Études par simulations

#### 3.1 Plan de simulation

Au moyen d'études par simulations, nous avons évalué le rendement de la méthode proposée, c'est-à-dire la vraisemblance composite par paire avec un ajustement de la courbure, et nous l'avons comparé à celui de la vraisemblance complète et de la vraisemblance composite par paire. Nous avons utilisé le modèle de l'équation (1.1) pour générer nos données, c'est-à-dire que pour  $i=1, \dots, n$  et  $j=1, \dots, m$ , nous avons simulé les valeurs de  $Y_{ij}$  à partir de

$$Y_{ij} = \theta + u_i + e_{ij}, \quad (3.1)$$

où  $\theta = 1$ ,  $u_i \stackrel{\text{iid}}{\sim} N(0, \sigma_u^2)$  et  $e_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_e^2)$ . Cela équivaut à avoir appliqué la génération de la superpopulation et l'échantillonnage décrits au paragraphe se rapportant à l'équation (1.1).

Notre première étude, qui ne fait pas l'objet du présent article, s'est penchée sur l'inférence au sujet de  $\theta$  avec  $\sigma_u$  connue et  $\sigma_e$ . Elle a démontré que l'utilisation de la vraisemblance composite par paire pour l'inférence au sujet de  $\theta$  surestimait grandement la précision et que l'ajustement de la courbure était efficace. Nous avons donc procédé à une étude plus approfondie visant à examiner l'inférence à la fois pour  $\theta$  et  $\sigma_u$ . Par souci de simplicité, nous avons pris  $\sigma_e = 0,5$  et avons considéré  $n \in \{20; 40\}$  et

$m \in \{5; 10\}$ . Pour la loi *a priori* demi-Cauchy définie dans l'équation (2.4), nous avons pris  $A \in \{5; 10; 15\}$ . Il y avait 500 ensembles de données qui se répliquaient pour chaque paramètre.

Nous avons examiné trois scénarios : 1)  $\sigma_u \in \{0,1; 0,5\}$  et la loi *a priori* demi-Cauchy sur  $\sigma_u$ ; 2) rapport signal-bruit,  $\text{RSB} \in \{0,25; 0,75\}$  et la loi *a priori* demi-Cauchy sur  $\sigma_u$ , où le  $\text{RSB} = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$ ; et 3)  $\sigma_u \in \{0,1; 0,5\}$  et une loi *a priori* uniforme  $\sigma_u$ . Pour l'ensemble des scénarios, nous avons utilisé une loi *a priori* uniforme pour  $\theta$ .

À la section 3.2, nous décrivons les algorithmes utilisés dans les études par simulations.

### 3.2 Algorithmes

Comme nous l'avons fait aux sections 2.1 et 2.2, définissons  $\mathbf{y}(n) = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  avec  $\mathbf{y}_i = (y_{i1}, \dots, y_{im})^T$  et  $\bar{y} = \sum_{i=1}^n \sum_{j=1}^m y_{ij} / (mn)$ . De plus,  $\boldsymbol{\eta}^{(t)}$  désigne la valeur de  $\boldsymbol{\eta}$  à la  $t^{\text{e}}$  itération où  $\boldsymbol{\eta} = (\theta, \sigma_u)^T$ . La vraisemblance complète est

$$L_{\text{VC}}(\theta, \sigma_u | \mathbf{y}(n)) \propto |\boldsymbol{\Sigma}_m|^{-n/2} \exp\left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_m^{-1} \mathbf{S}_0)\right], \quad (3.2)$$

comme dans l'équation (2.7).

L'utilisation de l'équation (3.2) de même que de la loi *a priori*,  $\pi(\boldsymbol{\eta})$ , permet d'obtenir la densité *a posteriori*,

$$p_{\text{VC}}(\boldsymbol{\eta} | \mathbf{y}(n)) \propto L_{\text{VC}}(\boldsymbol{\eta} | \mathbf{y}(n)) \pi(\boldsymbol{\eta}).$$

L'échantillonnage de  $\theta$  et de  $\sigma_u$  est réalisé en deux étapes :

**Étape 1.** Échantillonner  $\theta^{(t)}$  à partir de  $p_{\text{VC}}(\theta | \mathbf{y}(n), \sigma_u^{(t-1)})$  où

$$\theta | (\mathbf{y}(n), \sigma_u) \sim N\left(\bar{y}, \frac{\sigma_e^2 + m\sigma_u^2}{mn}\right).$$

Nous établissons la valeur de départ,  $\sigma_u^{(0)}$ , comme l'estimation du maximum de vraisemblance de  $\sigma_u$ .

**Étape 2.** Utiliser l'algorithme de Metropolis-Hastings (MH) pour tirer  $\sigma_u^{(t)}$  à partir de  $p_{\text{VC}}(\sigma_u | \mathbf{y}(n), \theta^{(t)})$ . Cette dernière valeur est facilement obtenue à partir de  $p_{\text{VC}}(\boldsymbol{\eta} | \mathbf{y}(n))$ . Soit  $s > 0$ , le candidat  $\sigma_u$ , désigné  $\sigma_u^*$ , est tiré de la distribution des sauts,  $N(\sigma_u^{(t-1)}, s^2)$ . Si  $\sigma_u^* < 0$ ,  $\sigma_u^{(t)} = \sigma_u^{(t-1)}$ . Sinon, la procédure est normale et comporte un rapport d'acceptation-rejet  $p_{\text{VC}}(\boldsymbol{\eta}_{\text{VC}}^* | \mathbf{y}(n)) / p_{\text{VC}}(\boldsymbol{\eta}_{\text{VC}}^{(t-1)} | \mathbf{y}(n))$  où  $\boldsymbol{\eta}_{\text{VC}}^* = (\theta^{(t)}, \sigma_u^*)^T$  et  $\boldsymbol{\eta}_{\text{VC}}^{(t-1)} = (\theta^{(t-1)}, \sigma_u^{(t-1)})^T$ .

**Étape 3.** Répéter les étapes 1 et 2 pour  $K = 1\,000$  fois avec les 200 premiers échantillons utilisés comme rodage.

La vraisemblance composite par paire (VP) est

$$L_{\text{VP}}(\theta, \sigma_u | \mathbf{y}(n)) \propto |\boldsymbol{\Sigma}_2|^{nm(m-1)/4} \exp\left[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}_2^{-1} \mathbf{S}_{0\text{VP}})\right], \quad (3.3)$$

comme dans l'équation (2.8).

L'utilisation de l'équation (3.3) de même que de la loi *a priori* choisie,  $\pi(\boldsymbol{\eta})$ , permet d'obtenir la densité *a posteriori*  $p_{VP}(\boldsymbol{\eta}|\mathbf{y}(n))$ .

L'échantillonnage de  $\theta$  et de  $\sigma_u$  est réalisé en trois étapes :

**Étape 1.** Échantillonner  $\theta^{(t)}$  à partir de  $p_{VP}(\theta|\mathbf{y}(n), \sigma_u^{(t-1)})$  où

$$\theta | (\mathbf{y}(n), \sigma_u) \sim N\left(\bar{y}, \frac{\sigma_e^2 + 2\sigma_u^2}{nm(m-1)}\right).$$

**Étape 2.** Utiliser l'algorithme de Metropolis-Hastings (MH) pour tirer  $\sigma_u^{(t)}$  à partir de  $p_{VP}(\sigma_u|\mathbf{y}(n), \theta^{(t)})$ , tel qu'il est décrit à l'étape 2 ci-dessus pour la VP (en remplaçant la VP par la VC dans toutes les formules).

**Étape 3.** Répéter les étapes 1 et 2 pour  $K=1000$  fois avec les 200 premiers échantillons utilisés comme rodage.

La dernière étape consiste à obtenir la vraisemblance composite par paire ajustée (VPA) (courbure), tel qu'elle est décrit à la section 2.3. Cette dérivation, fondée sur l'approche de Ribatet, Cooley et Davison, exploite  $\hat{\boldsymbol{\eta}}_{VPA}$ , les moyennes *a posteriori* estimées de  $\theta$  et de  $\sigma_u$ .

**Étape 1.** Soit  $(s_\theta, s_\sigma)$ , tirer le candidat  $\boldsymbol{\eta}^* = (\theta^*, \sigma_u^*)^T$  à partir de la distribution des sauts normale à deux variables,  $N_2(\boldsymbol{\eta}^{(t-1)}, \boldsymbol{\Sigma})$  où  $\boldsymbol{\Sigma} = \text{diag}(s_\theta^2, s_\sigma^2)$ . Si  $\sigma_u^* < 0$ ,  $\boldsymbol{\eta}^{(t)} = \boldsymbol{\eta}^{(t-1)}$ . Sinon, passer à l'étape 2.

**Étape 2.** Définissons  $\ell_{VP}(\mathbf{y}(n)|\theta, \sigma_u)$  comme le logarithme du rapport de vraisemblance composite par paire obtenu en prenant le logarithme de l'équation (3.3) et  $\ell_{VP}(\mathbf{y}_i|\theta, \sigma_u)$  comme le logarithme du rapport de vraisemblance composite par paire correspondant aux données de la grappe  $i$ , c'est-à-dire  $\mathbf{y}_i$ .

**Étape 3.** Obtenir numériquement  $\hat{\mathbf{H}} = \nabla^2 \ell_{VP}(\mathbf{y}(n)|\hat{\theta}_{VP}, \hat{\sigma}_{uVP})$  et

$$\hat{\mathbf{J}} = \sum_{i=1}^n \left[ \nabla \ell_{VP}(\mathbf{y}_i|\hat{\theta}_{VP}, \hat{\sigma}_{uVP}) \left\{ \nabla \ell(\mathbf{y}_i|\hat{\theta}_{VP}, \hat{\sigma}_{uVP}) \right\}^T \right],$$

où  $\hat{\theta}_{VP}$  et  $\hat{\sigma}_{uVP}$  sont les moyennes *a posteriori* estimées  $\theta$  et  $\sigma_u$ .

**Étape 4.** D'après l'approche de Ribatet, Cooley et Davison, et au moyen de la décomposition en valeurs singulières, nous écrivons  $\hat{\mathbf{H}} = \mathbf{M}^T \mathbf{M}$  et  $\hat{\mathbf{H}} \hat{\mathbf{J}}^{-1} \hat{\mathbf{H}} = \mathbf{M}_A^T \mathbf{M}_A$  pour certaines matrices  $\mathbf{M}$  et  $\mathbf{M}_A$ . Définir ensuite  $\mathbf{C} = \mathbf{M}^{-1} \mathbf{M}_A$ . Dans notre cas,  $\mathbf{C}$  est une matrice  $2 \times 2$ .

**Étape 5.** Selon l'approche de Ribatet, Cooley et Davison, le logarithme du rapport de vraisemblance composite par paire ajusté,  $\ell_{VPA}$ , est

$$\ell_{VPA}(\mathbf{y}(n)|\boldsymbol{\eta}) = \ell_{VP}(\mathbf{y}(n)|\boldsymbol{\eta}^*)$$

où

$$\boldsymbol{\eta}^* = \hat{\boldsymbol{\eta}}_{VP} + \mathbf{C}(\boldsymbol{\eta} - \hat{\boldsymbol{\eta}}_{VP}).$$

**Étape 6.** Définir la densité *a posteriori* par paire ajustée comme

$$p_{VPA}(\boldsymbol{\eta} | \mathbf{y}(n)) \propto L_{VPA}(\mathbf{y}(n) | \boldsymbol{\eta}) \pi(\theta, \sigma_u)$$

où  $L_{VPA}(\mathbf{y}(n) | \boldsymbol{\eta}) = \exp(\ell_{VPA}(\mathbf{y}(n) | \boldsymbol{\eta}))$ , ce dernier étant défini à l'étape 5.

À l'aide de la valeur candidate,  $\boldsymbol{\eta}^*$ , obtenue à l'étape 1, définir la valeur candidate ajustée  $\boldsymbol{\eta}_c^* = \hat{\boldsymbol{\eta}}_{VP} + \mathbf{C}(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}}_{VP})$ . Le rapport d'acceptation-rejet est alors

$$p_{VPA}(\boldsymbol{\eta}_c^* | \mathbf{y}(n)) / p_{VPA}(\boldsymbol{\eta}^{(t)} | \mathbf{y}(n)).$$

Les autres étapes sont les étapes normales pour l'algorithme de Metropolis-Hastings.

### 3.3 Résultats des simulations

Pour chaque méthode (VC, VP, VPA), chaque paramètre du plan  $(m, n)$  et chaque loi *a priori*, nous résumons les résultats des simulations en utilisant a) les taux de couverture des intervalles de crédibilité sous échantillonnage répété et b) les moyennes des points 0,025; 0,25; 0,50; 0,75 et 0,975 des lois *a posteriori* de  $\theta$  et de  $\sigma_u$ .

Des résumés graphiques sont également présentés, c'est-à-dire des estimations de la densité *a posteriori moyenne* pour chacune des lois *a posteriori*, c'est-à-dire  $p_{VC}(\boldsymbol{\eta} | \mathbf{y}(n))$ ,  $p_{VP}(\boldsymbol{\eta} | \mathbf{y}(n))$  et  $p_{VPA}(\boldsymbol{\eta} | \mathbf{y}(n))$ . D'abord, considérons un intervalle, disons,  $[a, b]$ , qui soutient la plus grande partie de la masse (par exemple 95 %) des densités *a posteriori*. Divisons-le ensuite en  $M = 50$  sous-intervalles échelonnés également comportant les points de découpage  $a = c_0 < c_1 < \dots < c_{M-1} < c_M = b$ . Pour  $t = 1, \dots, T$ , soit  $\hat{f}_P^{(t)}(\cdot)$  l'estimation de la densité *a posteriori*  $f_P(\cdot)$ , dérivée de la simulation  $t^e$ , où  $P$  désigne la VC, la VP ou la VPA et  $T$  correspond au nombre de simulations. Définissons ensuite, pour  $r = 1, \dots, M$ ,

$$\hat{f}_P(c_r) = \frac{1}{T} \sum_{t=1}^T \hat{f}_P^{(t)}(c_r).$$

Prenons ensuite une courbe reliant les points  $\{c_r, \hat{f}_P(c_r)\}$  pour  $a = c_0 < c_1 < \dots < c_M = b$  comme estimation de la densité *a posteriori moyenne* pour  $f_P(\cdot)$ .

Le tableau 3.1 présente les taux de couverture pour  $\theta$  et  $\sigma_u$  pour  $A = 15$ ,  $n \in \{20; 40\}$ ,  $m \in \{5; 10\}$  et  $\sigma_u \in \{0,1; 0,289; 0,5; 0,866\}$ . La figure 3.1 présente les estimations moyennes de la densité *a posteriori* pour  $\theta$  et  $\sigma_u$  pour  $A = 15$ ,  $\sigma_u \in \{0,1; 0,5\}$ ,  $n = 40$  et  $m = 10$ . Au tableau 3.1 et à la figure 3.1, les résumés sont présentés pour la vraisemblance complète (VC), la vraisemblance composite par paire (VP) et la vraisemblance composite par paire ajustée (VPA).

**Tableau 3.1**  
**Taux de couverture (en pourcentage) pour les 95 % des intervalles de crédibilité de  $\theta$  et  $\sigma_u$  avec  $A = 15$**

		$\sigma_u = 0,1$		$\sigma_u = 0,289$		$\sigma_u = 0,5$		$\sigma_u = 0,866$	
		$n = 20$	$n = 40$	$n = 20$	$n = 40$	$n = 20$	$n = 40$	$n = 20$	$n = 40$
		$\theta$							
$m = 5$	$\hat{\theta}_{VC}$	97,40	95,80	94,84	94,60	94,80	94,40	94,80	95,00
	$\hat{\theta}_{VP}$	68,20	66,60	58,45	58,40	53,60	51,40	50,00	50,20
	$\hat{\theta}_{VPA}$	92,40	93,00	92,96	93,60	92,20	92,20	91,60	93,00
$m = 10$	$\hat{\theta}_{VC}$	94,80	95,00	95,00	94,00	94,80	94,20	95,00	93,80
	$\hat{\theta}_{VP}$	43,80	42,80	35,40	31,80	30,40	29,60	27,40	26,40
	$\hat{\theta}_{VPA}$	90,60	91,80	92,20	93,40	92,80	92,60	91,80	93,00
$m = 5$	$\hat{\sigma}_{u,VC}$	97,20	99,00	91,55	95,40	93,00	94,80	92,60	95,00
	$\hat{\sigma}_{u,VP}$	92,80	85,60	59,62	61,80	52,40	54,20	46,20	48,20
	$\hat{\sigma}_{u,VPA}$	88,40	83,40	86,85	92,20	84,40	91,20	82,00	89,60
$m = 10$	$\hat{\sigma}_{u,VC}$	99,00	97,20	93,60	92,80	93,80	93,80	93,00	93,60
	$\hat{\sigma}_{u,VP}$	63,60	56,80	33,40	38,00	27,00	29,60	24,40	26,60
	$\hat{\sigma}_{u,VPA}$	82,80	84,40	85,20	89,00	80,80	86,60	79,00	87,00

Le résumé suivant présente les résultats pour la loi *a priori* demi-Cauchy seulement avec  $A \in \{5; 10; 15\}$ ,  $m \in \{5; 10\}$ ,  $n \in \{20; 40\}$  et  $\sigma_u \in \{0,1; 0,289; 0,5; 0,866\}$ , les deuxième et quatrième valeurs correspondant au RSB = 0,25 et au RSB = 0,75, respectivement. Les résultats sont semblables pour les trois choix de  $A$  et pour la loi *a priori* uniforme.

Sans ajustement, les couvertures de la VP sont considérablement différentes du niveau nominal de 0,95. Par exemple (tableau 3.1), pour  $A = 15$ ,  $n = 40$ ,  $m = 10$  et  $\sigma_u = 0,5$ , la couverture pour  $\theta$  est inférieure à 0,30. Si nous considérons *toutes* les valeurs des paramètres du plan, la plus grande couverture est de 0,70. Dans la plupart des cas, la couverture pour  $\theta$  est bien inférieure à 0,70.

Lorsque l'ajustement de la courbure est pris en compte, la couverture pour  $\theta$  est excellente. Des 48 cas (trois choix de  $A$ , deux choix de  $m$ , deux choix de  $n$ , quatre choix de  $\sigma_u$ ), treize avaient une couverture entre 0,93 et 0,95, vingt-deux entre 0,92 et 0,93, onze entre 0,91 et 0,92, et deux de moins de 0,91, avec le dernier cas de figure pour  $\sigma_u = 0,1$ ,  $n = 20$ ,  $m = 10$  et  $A = 5$  et 15.

Lorsque l'ajustement de la courbure est pris en compte, la couverture pour  $\sigma_u$  varie considérablement, mais il y a, dans la plupart des cas, une très grande amélioration de la couverture comparativement à l'utilisation de la vraisemblance composite par paire non ajustée.

Les tracés (figure 3.1) montrent que, pour  $\theta$ , la loi *a posteriori* correspondant à la vraisemblance ajustée se rapproche beaucoup de la loi *a posteriori* fondée sur la vraisemblance complète. Pour  $\sigma_u$ , il existe des différences entre les lois *a posteriori* correspondant à la vraisemblance complète et à la



vraisemblance ajustée, en particulier une transition vers des valeurs plus faibles dans le cas de cette dernière.

Pour étudier les effets de l'augmentation de  $m$  et de  $n$ , considérons la différence  $\delta = C_{VC} - C_{VPA}$ , où  $C$  désigne la couverture et VC et VPA renvoient aux lois *a posteriori* correspondantes.

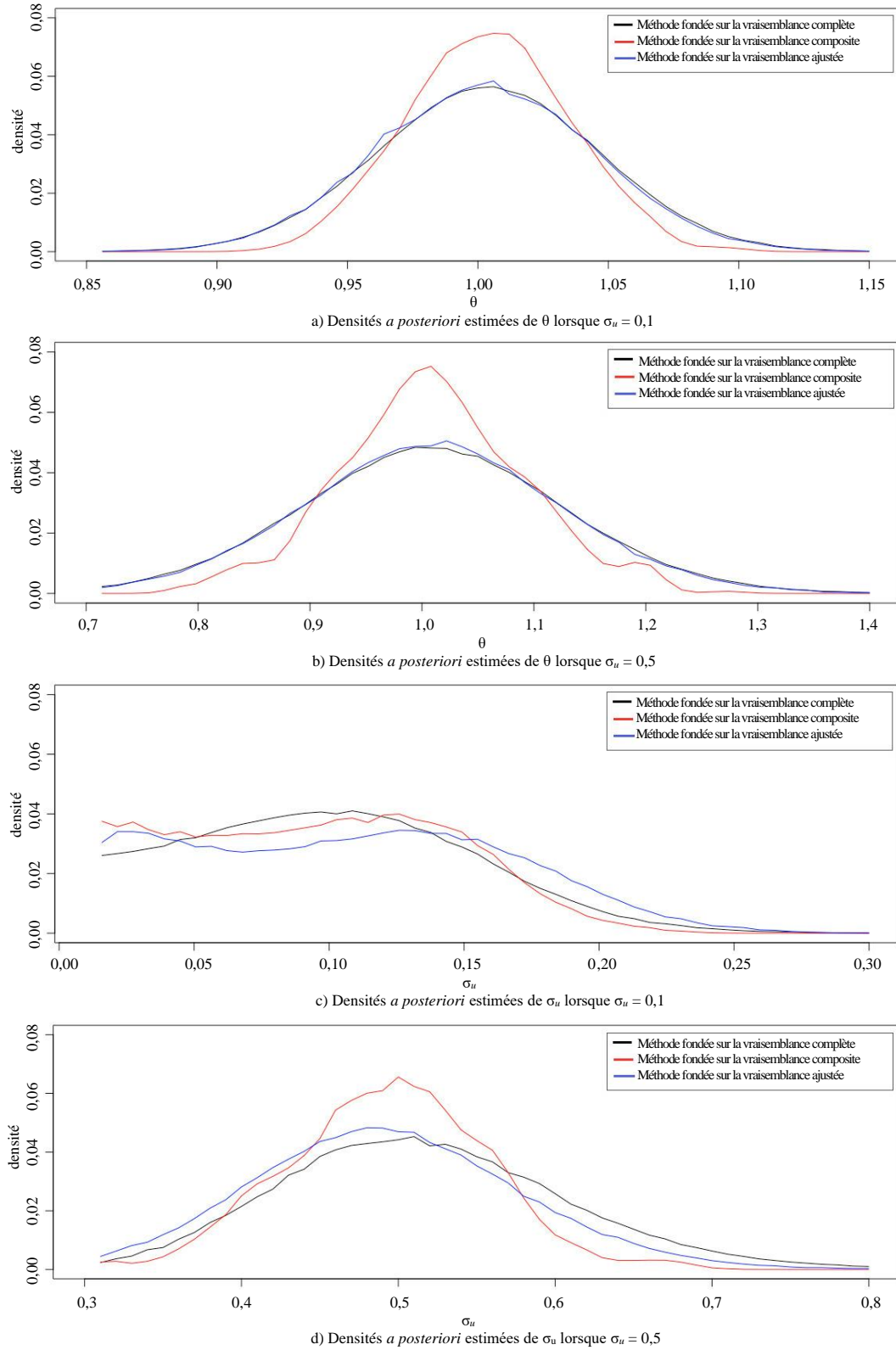
Dans l'ensemble, lorsque nous considérons tous les  $m, n, A$  et  $\sigma_u$ , pour  $\theta$ ,  $\delta$  diminue à mesure que  $n$  augmente. Pour les valeurs plus grandes de  $\sigma_u$ ,  $\delta$  diminue à mesure que  $m$  augmente, tandis que pour les valeurs plus faibles de  $\sigma_u$ ,  $\delta$  a tendance à augmenter à mesure que  $m$  augmente. Dans l'ensemble, pour  $\sigma_u$ ,  $\delta$  diminue à mesure que  $n$  augmente, sauf dans le cas où  $\sigma_u = 0,1$ , tandis que  $\delta$  augmente à mesure que  $m$  augmente.

La raison de la détérioration de l'ajustement à mesure que  $m$  augmente pourrait être que le nombre de paires par grappe est  $m(m-1)/2$  et augmente plus rapidement, de sorte que la vraisemblance par paire se concentre plus rapidement autour de son mode; l'ajustement de la courbure peut ne pas être suffisant pour compenser un changement de forme du logarithme du rapport de vraisemblance composite par paire, par exemple une augmentation de l'aplatissement.

Le tableau 3.2 présente les taux de non-couverture unilatéraux pour les 95 % des intervalles de crédibilité pour  $\theta$  et  $\sigma_u$  au moyen de  $A = 15$ ,  $n \in \{20; 40\}$ ,  $m \in \{5; 10\}$  et  $\sigma_u \in \{0,1; 0,289; 0,5; 0,866\}$ . Nous observons ce qui suit :

- i) Pour  $\theta$ , la non-couverture pour les intervalles de vraisemblance complète semble symétrique. La vraisemblance par paire ajustée présente un sous-dénombrement pour  $\theta$  et, sauf lorsque  $\sigma_u$  est égale à 0,1, le sous-dénombrement est symétrique. Nous observons seulement une dépendance de la couverture sur  $m$  dans le cas où  $\sigma_u = 0,1$ .
- ii) Pour  $\sigma_u$ , l'intervalle de vraisemblance complète présente un sous-dénombrement qui se rapproche du niveau nominal et n'est pas très asymétrique, sauf dans le cas où  $\sigma_u = 0,1$ , pour lequel nous observons un sous-dénombrement marqué. Pour  $\sigma_u > 0,1$  et  $m = 5$ , la couverture s'améliore à mesure que  $n$  passe de 20 à 40, mais pour  $\sigma_u > 0,1$  et  $m = 10$ , la couverture présente peu de différence pour les deux valeurs de  $n$ .
- iii) Pour  $\sigma_u$ , la vraisemblance par paire ajustée présente une couverture asymétrique. Sauf dans le cas où  $\sigma_u = 0,1$ , l'ampleur de la non-couverture à gauche a tendance à être semblable à celle de la vraisemblance complète, mais beaucoup plus grande à droite, et la couverture s'améliore à mesure que  $n$  passe de 20 à 40.

**Figure 3.1** Densités *a posteriori* estimées de  $\theta$  et de  $\sigma_u$  au moyen de trois méthodes lorsque  $A = 15$ ,  $n = 40$ ,  $m = 10$  et  $\sigma_u = (0,1; 0,5)$  utilisant une loi *a priori* demi-Cauchy pour  $\sigma_u$ .



Sachant que le logarithme du rapport de vraisemblance par paire ajusté n'est pas explicitement construit pour estimer le logarithme du rapport de vraisemblance complet, nous constatons que, dans la figure 3.1, le logarithme du rapport de vraisemblance par paire ajusté diminue plus rapidement dans les queues.

Nous avons également tenté de centrer l'ajustement de la courbure sur le mode *a posteriori* du logarithme du rapport de vraisemblance par paire plutôt que sur la moyenne *a posteriori* du logarithme du rapport de vraisemblance par paire, et nous avons constaté que le sous-dénombrement augmentait, même si l'asymétrie de la couverture était moins marquée, pour les intervalles de crédibilité qui en découlent.

Tableau 3.2

Taux de non-couverture unilatéraux (en pourcentage) pour les 95 % des intervalles de crédibilité (IC) de  $\theta$  et  $\sigma_u$  avec  $A = 15$

		TNC-G	TNC-D	TNC-G	TNC-D	TNC-G	TNC-D	TNC-G	TNC-D
		$\sigma_u = 0,1$				$\sigma_u = 0,289$			
		$n = 20$		$n = 40$		$n = 20$		$n = 40$	
		$\theta$							
$m = 5$	$\hat{\theta}_{VC}$	1,40	1,20	1,60	2,60	3,05	2,11	3,20	2,20
	$\hat{\theta}_{VP}$	16,60	15,20	16,40	17,00	21,13	20,42	20,40	21,20
	$\hat{\theta}_{VPA}$	2,60	5,00	2,20	4,80	3,76	3,29	3,80	2,60
$m = 10$	$\hat{\theta}_{VC}$	2,80	2,40	2,40	2,60	3,00	2,00	3,20	2,80
	$\hat{\theta}_{VP}$	26,80	29,40	28,60	28,60	33,00	31,60	33,80	34,40
	$\hat{\theta}_{VPA}$	3,60	5,80	4,60	3,60	4,00	3,80	3,80	2,80
		$\sigma_u$							
$m = 5$	$\hat{\sigma}_{u,VC}$	2,80	0,00	1,00	0,00	3,05	5,40	1,80	2,80
	$\hat{\sigma}_{u,VP}$	7,20	0,00	9,20	5,20	14,79	25,59	14,80	23,40
	$\hat{\sigma}_{u,VPA}$	4,60	7,00	3,60	13,00	3,05	10,09	2,40	5,40
$m = 10$	$\hat{\sigma}_{u,VC}$	1,00	0,00	2,00	0,80	3,00	3,40	3,80	3,40
	$\hat{\sigma}_{u,VP}$	14,60	21,80	17,80	25,40	22,80	43,80	24,80	37,20
	$\hat{\sigma}_{u,VPA}$	3,40	13,80	3,80	11,80	3,00	11,80	2,80	8,20
		$\sigma_u = 0,5$				$\sigma_u = 0,866$			
		$n = 20$		$n = 40$		$n = 20$		$n = 40$	
		$\theta$							
$m = 5$	$\hat{\theta}_{VC}$	3,20	2,00	3,00	2,60	3,40	1,80	3,00	2,00
	$\hat{\theta}_{VP}$	24,40	22,00	24,60	24,00	26,40	23,60	25,80	24,00
	$\hat{\theta}_{VPA}$	4,40	3,40	4,20	3,60	4,20	4,20	3,80	3,20
$m = 10$	$\hat{\theta}_{VC}$	3,00	2,20	3,40	2,40	3,00	2,00	3,40	2,80
	$\hat{\theta}_{VP}$	34,60	35,00	35,60	34,80	36,80	35,80	37,60	36,00
	$\hat{\theta}_{VPA}$	4,00	3,20	4,20	3,20	4,80	3,40	3,40	3,60
		$\sigma_u$							
$m = 5$	$\hat{\sigma}_{u,VC}$	3,00	4,00	2,20	3,00	3,40	4,00	2,00	3,00
	$\hat{\sigma}_{u,VP}$	16,00	31,60	18,20	27,60	19,20	34,60	21,00	30,80
	$\hat{\sigma}_{u,VPA}$	1,20	14,40	1,80	7,00	1,40	16,60	2,20	8,20
$m = 10$	$\hat{\sigma}_{u,VC}$	3,20	3,00	2,80	3,40	3,80	3,20	3,20	3,20
	$\hat{\sigma}_{u,VP}$	24,00	49,00	28,00	42,40	25,40	50,20	29,80	43,60
	$\hat{\sigma}_{u,VPA}$	2,60	16,60	2,40	11,00	3,20	17,80	2,00	11,00

Note : TNC-G désigne les taux de non-couverture à gauche (en pourcentage) pour les 95 % des IC de  $\theta$  et de  $\sigma_u$ ; TNC-D désigne les taux de non-couverture à droite (en pourcentage) pour les 95 % des IC de  $\theta$  et de  $\sigma_u$ .

## 4. Prolongement des plans d'échantillonnage avec probabilités inégales

Un prolongement important des conditions que nous avons établies se rapporte à un cadre d'échantillonnage complexe, pour lequel l'estimation fréquentiste des paramètres par l'estimation d'une vraisemblance composite par paire à l'échelle de la population est maintenant assez couramment utilisée. Rao, Verret et Hidioglou et Yi, Rao et Li ont démontré qu'une approche fondée sur l'application d'une vraisemblance composite par paire fréquentiste fonctionne bien pour estimer les composantes de la variance d'un modèle multiniveau dans le cas de certains plans d'échantillonnage avec probabilités inégales et permet d'éviter le problème d'incohérence lorsque les tailles de l'échantillon de deuxième degré sont petites. Dans le cadre de cette approche, l'estimation de l'incertitude est fondée sur la théorie des fonctions d'estimation et peut ne pas nécessiter les ajustements que nous prenons en compte dans le présent article. Cependant, il serait souhaitable de formuler un équivalent bayésien de cette méthode. Si l'on convenait d'une formulation bayésienne, les résultats de l'étude permettraient de prévoir la nécessité de l'ajustement du pseudo-logarithme du rapport de vraisemblance composite par paire pour l'harmoniser à une fonction de logarithme du rapport de vraisemblance complet adéquat.

Supposons que l'objectif est toujours analytique, que le modèle pour  $Y_{ij}$  est l'équation (1.1) et que les objets de l'inférence sont la moyenne  $\theta$  et la composante de la variance  $\sigma_u^2$  ou sa racine carrée. La population cible comporte  $N$  unités de premier degré présentant des tailles de  $M_i$ ,  $i=1, \dots, N$ , et l'échantillon de premier degré se compose de  $n$  de ces unités, sélectionnées au moyen d'un plan d'échantillonnage avec probabilités inégales. Au deuxième degré,  $m_i$  unités élémentaires sont sélectionnées par échantillonnage aléatoire simple à partir de la  $i^{\text{e}}$  unité de premier degré, si cette unité a été échantillonnée au premier degré. Si les tailles  $M_i$  et  $m_i$  et les probabilités liées au plan d'échantillonnage  $p(s)$  (où  $s$  traverse les sous-ensembles de la population de deuxième degré satisfaisant aux spécifications de tailles d'échantillon) ne dépendent pas des valeurs de  $u_i$  ou de  $e_{ij}$ , la fonction de vraisemblance peut être considérée comme prenant la forme de l'équation (2.3),  $m$  étant remplacée par  $m_i$ , et le prolongement de nos travaux est en principe simple. Cependant, si les tailles ou les probabilités liées au plan d'échantillonnage dépendent des valeurs de  $u_i$  ou de  $e_{ij}$ , elles nous renseigneront sur les paramètres d'intérêt. La fonction de vraisemblance à l'échelle de l'échantillon issue de la combinaison du modèle multiniveau et du plan d'échantillonnage peut être mal définie ou impossible à traiter. D'un point de vue bayésien, nous devons alors considérer ce qui peut raisonnablement remplacer la vraisemblance véritable et dans quelle mesure ce substitut peut être estimé avec précision au moyen d'une vraisemblance composite par paire ajustée. Les réponses peuvent dépendre de la méthode privilégiée pour utiliser les probabilités liées au plan d'échantillonnage dans l'inférence, et il existe plusieurs possibilités. Il serait intéressant d'étudier ces possibilités dans les travaux à venir.

Une méthode, dont l'applicabilité est limitée, serait fondée sur l'approche de Léon-Novelo et de Savitsky (2019). En supposant un échantillonnage de Bernoulli à un degré (de sorte que les probabilités d'échantillonnage soient entièrement déterminées par les probabilités d'inclusion), ceux-ci modélisent la

distribution conjointe de la variable de résultat,  $Y$ , et la probabilité d'inclusion,  $\pi$ , en utilisant le modèle générant  $Y$  à partir de  $\mathbf{x}$  dans la population et un modèle générant  $\pi$  à partir de  $\mathbf{x}$  et de  $Y$ . Pour rendre les calculs possibles, des restrictions doivent être imposées à la forme de ce modèle; voir le théorème 1 et, en particulier, le cas spécial présenté à la section 2.1 de leur article.

Nous pouvons étendre le modèle présenté à la section 2.1 de Léon-Novelo et Savitsky (2019) à l'échantillonnage en grappes à deux degrés. On peut faire un autre prolongement, c'est-à-dire remplacer la densité d'échantillonnage de  $Y$  par une vraisemblance composite par paire analogue à la partie de vraisemblance du modèle (2.6). Ainsi, sous réserve des limites du théorème 1 de Léon-Novelo et Savitsky (2019), il existe des équivalents aux densités *a posteriori*, les équations (2.5) et (2.6), qui incluent les probabilités d'inclusion.

Une autre méthode, partiellement bayésienne, mais peut-être le prolongement le plus largement applicable de notre approche, consiste à considérer comme exacte la fonction de logarithme du rapport de vraisemblance de la population (complète) (équations [2.5] et [2.6] de Rao, Verret et Hidiroglou) et à formuler une fonction correspondante de logarithme du rapport de vraisemblance composite par paire pour la population complète comme dans notre section 2. Nous tenterions ensuite d'estimer cette dernière à partir de l'échantillon en utilisant des poids d'échantillonnage (équation [4.2] de Rao, Verret et Hidiroglou) et nous ferions des ajustements, comme la normalisation adéquate des poids ou le « rééchelonnage » comme dans Pfeffermann, Skinner, Holmes, Goldstein et Rasbash (1998), et des ajustements de la courbure à la fonction de logarithme du rapport de vraisemblance composite par paire estimé qui en résultent. Cela produirait une fonction de pseudo-logarithme du rapport de vraisemblance par paire qui pourrait être utilisée comme une fonction de logarithme du rapport de vraisemblance approximatif dans l'inférence bayésienne. Cela permettrait d'obtenir un équivalent bayésien à la méthode fréquentiste proposée par Rao, Verret et Hidiroglou et par Yi, Rao et Li et d'étendre la méthode exposée dans le présent article à la situation d'échantillonnage avec probabilités inégales.

Nous avons obtenu des renseignements préliminaires pour cette deuxième approche. C'est-à-dire que si  $\sigma_u^2$  est connue, les expressions analytiques pour la vraisemblance complète et la vraisemblance composite par paire sont disponibles pour  $\theta$  à l'échelle de la population complète. Pour la vraisemblance partielle, nous modifions l'équation (2.8) en prenant  $\sigma_u$  maintenue constante et en ajoutant les poids  $w_i$  et  $w_{jk|i}$  comme dans l'équation (4.2) de Rao, Verret et Hidiroglou. Avec une loi *a priori* uniforme localement pour  $\theta$ ,

$$p_{VP}(\theta|\mathbf{y}(n)) \propto \exp\left\{-0,5 \sum_{i=1}^n \sum_{j < k} w_i w_{jk|i} (y_{ij} - \theta \quad y_{ik} - \theta) \Sigma_2^{-1} (y_{ij} - \theta \quad y_{ik} - \theta)^T\right\}$$

où

$$\Sigma_2^{-1} = \begin{bmatrix} \sigma^{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{bmatrix}$$

avec

$$\sigma^{11} = \sigma^{22} = \sigma_e^{-2} \left( 1 - \frac{\sigma_u^2}{\sigma_e^2 + 2\sigma_u^2} \right)$$

et

$$\sigma^{12} = \sigma^{21} = -\frac{\sigma_u^2}{\sigma_e^2 (\sigma_e^2 + 2\sigma_u^2)}.$$

Après quelques calculs algébriques,

$$p_{VP}(\theta | \mathbf{y}(n)) \propto \exp \left\{ -0,5 \frac{2 \sum_{i=1}^n \sum_{j < k} w_i w_{jk|i}}{\sigma_e^2 + 2\sigma_u^2} \left[ \theta - \frac{\sum_{i=1}^n \sum_{j < k} w_i w_{jk|i} (y_{ij} + y_{ik}) / 2}{\sum_{i=1}^n \sum_{j < k} w_i w_{jk|i}} \right]^2 \right\}.$$

De même, nous modifions l'équation (2.7) en maintenant  $\sigma_u$  constante et en ajoutant les poids. Avec une loi *a priori* uniforme localement pour  $\theta$ ,

$$p_{VC}(\theta | \mathbf{y}(n)) \propto \exp \left\{ -0,5 \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^m w_i w_{jk|i} \sigma^{jk} (y_{ij} - \theta) (y_{ik} - \theta) \right\}.$$

Après quelques calculs algébriques,

$$p_{VC}(\theta | \mathbf{y}(n)) \propto \exp \left\{ -0,5 \sum_{i=1}^n w_i \left\{ \sum_{j=1}^m w_{j|i} a^{(1)} + \sum_{j \neq k} w_{jk|i} a^{(2)} \right\} (\theta - \hat{\theta})^2 \right\}$$

où

$$a^{(1)} = \sigma_e^{-2} \left( 1 - \frac{\sigma_u^2}{\sigma_e^2 + m\sigma_u^2} \right),$$

$$a^{(2)} = -\frac{\sigma_u^2}{\sigma_e^2 (\sigma_e^2 + m\sigma_u^2)}$$

et

$$\hat{\theta} = \frac{\sum_{i=1}^n w_i \left[ \sum_{j=1}^m a^{(1)} w_{j|i} y_{ij} + \sum_{j \neq k} a^{(2)} w_{jk|i} (y_{ij} + y_{ik}) / 2 \right]}{\sum_{i=1}^n w_i \left[ \sum_{j=1}^m a^{(1)} w_{j|i} + \sum_{j \neq k} a^{(2)} w_{jk|i} \right]}.$$

Le choix du rééchantillonnage des poids sera important. Pour quantifier la précision surestimée dans la loi *a posteriori* du logarithme du rapport de vraisemblance composite par paire, il peut être nécessaire de faire une évaluation numérique.

Un avantage de procéder à d'autres prolongements de cette approche bayésienne dans les travaux à venir serait qu'elle est axée sur l'inférence pour les paramètres du modèle, plutôt sur que les quantités dans la population finie, et il ne serait donc pas nécessaire de considérer des probabilités d'inclusion de troisième ou de quatrième ordre dans l'estimation de l'incertitude pour  $\sigma_u^2$  ou  $\sigma_u$ .

## 5. Conclusion

Il existe des raisons philosophiques et théoriques bien connues d'envisager des approches bayésiennes pour réaliser un échantillonnage, et des travaux de recherche dans ce domaine sont réalisés depuis longtemps. Voir, par exemple, Sedransk (2008). Il existe aussi des avantages pratiques. L'utilisation d'une approche bayésienne plutôt qu'une approche fréquentiste s'appuie beaucoup moins sur des approximations, en remplaçant des calculs par des expressions asymptotiques. Dans le contexte des modèles à effets aléatoires, un avantage important est la capacité de contraindre les composantes de la variance à des valeurs non négatives dans la loi *a priori*, sans masquer les lacunes des données.

Le Service national de la statistique agricole (NASS) du ministère américain de l'Agriculture est l'un des endroits où des méthodes bayésiennes sont largement utilisées. Au NASS, on emploie des méthodes bayésiennes pour produire des statistiques officielles à l'échelle du pays et des États pour des variables comme la superficie des cultures et le rendement des cultures. Souvent, ces inférences reposent sur plusieurs sources de données. On accorde une attention particulière à l'estimation cohérente dans l'ensemble de la hiérarchie des régions géographiques d'intérêt aux fins de l'inférence. Voir Nandram, Berg et Barboza (2014); Erciulescu, Cruze et Nandram (2020, 2019, 2018); et Cruze, Erciulescu, Nandram, Barboza et Young (2019) pour obtenir de plus amples précisions.

Nous nous sommes penchés sur une utilisation de la vraisemblance composite par paire dans l'inférence bayésienne pour les données d'enquête, c'est-à-dire que nous avons élaboré une loi *a posteriori* pour la moyenne  $\theta$  et le paramètre d'écart-type  $\sigma_u$  d'un modèle à effets aléatoires simple. Nous avons évalué la loi *a posteriori* en ce qui a trait aux propriétés de couverture fréquentistes des intervalles de crédibilité pour les paramètres et constaté qu'elle fonctionne bien pour  $\theta$ , mais qu'elle n'est pas pleinement satisfaisante pour l'inférence au sujet de  $\sigma_u$  dans les conditions considérées. Des observations correspondantes pourraient être faites pour l'inférence fréquentiste de la vraisemblance composite par paire, considérée comme une fonction de vraisemblance approximative. Il est possible que de meilleurs résultats puissent être obtenus en appliquant une transformation adéquate à  $\sigma_u$ ; le sujet pourrait faire l'objet de futurs travaux de recherche.

Une situation idéale pour l'utilisation de la vraisemblance composite dans l'inférence bayésienne en est une dans laquelle a) un modèle de génération des données est entièrement spécifié, de sorte qu'une fonction de vraisemblance véritable existe, et b) on peut raisonnablement estimer la vraisemblance véritable au moyen de la vraisemblance composite, de sorte que les lois *a posteriori* correspondantes aient une bonne concordance. Par exemple, pour Stoehr et Friel (2018), l'objectif est l'utilisation de l'inférence

bayésienne d'une pseudo-vraisemblance pour les données issues d'un champ aléatoire de Gibbs. Ils établissent des identités qui lient le gradient et la matrice hessienne de la loi *a posteriori* du log pour un paramètre à des moments de statistiques suffisantes du champ aléatoire et les utilisent pour accroître la capacité de la log-densité *a posteriori* composite par paire à estimer la fonction de log-densité *a posteriori*. L'ajustement de la courbure de Ribatet, Cooley et Davison, sur lequel repose notre approche, porte plutôt sur le logarithme du rapport de vraisemblance composite par paire de sorte que son gradient (que nous pourrions appeler le « vecteur de score par paire ») présente la propriété d'absence de biais de l'information menant à des intervalles de crédibilité dont les probabilités de couverture fréquentistes se rapprochent des valeurs nominales. Intuitivement, compte tenu de l'augmentation du nombre  $n$  de grappes,  $m$  demeurant constant, cette approximation devrait s'améliorer et son calcul ne nécessite pas d'utiliser les propriétés de la vraisemblance elle-même. Dans le présent article, nous avons tiré profit de la disponibilité de la vraisemblance complète dans le cas simple pour évaluer dans quelle mesure l'inférence bayésienne fondée sur la vraisemblance composite par paire ajustée se rapproche de l'inférence bayésienne complète.

## Remerciements

Les travaux de recherche sont financés en partie grâce à des subventions du Conseil de recherches en sciences naturelles et en génie du Canada accordées à Mary E. Thompson et à Grace Y. Yi. Grace Y. Yi est titulaire de la Chaire de recherche du Canada en sciences des données (niveau 1). Ses travaux de recherche ont été entrepris, en partie, grâce au financement du Programme des chaires de recherche du Canada.

## Bibliographie

- Cox, D.R., et Reid, N. (2004). A note on pseudo likelihood constructed from marginal densities. *Biometrika*, 91, 729-737.
- Cruze, N., Erciulescu, A., Nandram, B., Barboza, W. et Young, L. (2019). Producing official county-level agricultural estimates in the United States: Needs and challenges. *Statistical Science*, 34, 301-316.
- Erciulescu, A., Cruze, N. et Nandram, B. (2018). Benchmarking a triplet of official estimates. *Environmental and Ecological Statistics*, 23, 523-547.
- Erciulescu, A., Cruze, N. et Nandram, B. (2019). Model-based county level crop estimates incorporating auxiliary sources of information. *Journal of the Royal Statistical Society, Series A*, 182, 283-303.
- Erciulescu, A., Cruze, N. et Nandram, B. (2020). Statistical challenges in combining survey statistics and auxiliary data to produce official statistics. *Journal of Official Statistics*, 36, 63-88.



- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 3, 515-533.
- Heyde, C.C. (1997). *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation*. New York: Springer-Verlag.
- Jørgensen, B., et Knudsen, S.J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scandinavian Journal of Statistics*, 31, 93-114.
- Lehmann, E.L. (1999). *Elements of Large-Sample Theory*. New York: Springer-Verlag.
- Léon-Novelo, L., et Savitsky, T. (2019). Fully Bayesian estimation under informative sampling. *Electronic Journal of Statistics*, 13, 1608-1645.
- Lindsay, B.G. (1982). Conditional score functions: Some optimality results. *Biometrika*, 69, 505-512.
- Lindsay, B.G. (1988). Composite likelihood methods. *Contemporary Mathematics*, 80, 220-239.
- Lindsay, B.G., Yi, G.Y. et Sun, J. (2011). Issues and strategies in the selection of composite likelihoods. *Statistica Sinica*, 21, 71-105.
- Nandram, B., Berg, E. et Barboza, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. *Journal of Environmental and Ecological Statistics*, 21, 507-530.
- Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. et Rasbash, J. (1998). Weighting for unequal selection probabilities in multi-level models. *Journal of the Royal Statistical Society, Series B*, 60, 23-56.
- Rao, J.N.K., Verret, F. et Hidiroglou, M.A. (2013). Une approche d'inférence fondée sur la vraisemblance composite pondérée pour des modèles à deux niveaux issus de données d'enquête. *Techniques d'enquête*, 39, 2, 263-282. Article accessible à l'adresse <https://www150.statcan.gc.ca/n1/pub/12-001-x/2013002/article/11887-fra.pdf>.
- Rabe-Hesketh, S., et Skrondal, A. (2006). Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society, Series A*, 169, 805-827.
- Ribatet, M., Cooley, D. et Davison, A.C.D. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, 22, 813-845.

Sedransk, J. (2008). Assessing the value of Bayesian methods for inference about finite population quantities. *Journal of Official Statistics*, 24, 495-506.

Skinner, C.J., Holt, D. et Smith, T.F.M. (1989). *Analysis of Complex Surveys*. Wiley.

Stoehr, J., et Friel, N. (2018). Calibration of conditional composite likelihood for Bayesian inference on Gibbs random fields. *Artificial Intelligence and Statistics*, 921-929. arXiv:150201997v2.

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis*, 92, 1. <https://doi.org/10.1007/s10182-008-0060-7> (consulté le 1<sup>er</sup> juillet 2020).

Varin, C., Reid, N. et Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 21, 5-24.

Yi, G.Y. (2017). Composite likelihood/pseudolikelihood. *Wiley StatsRef: Statistics Reference Online*. 1-14.

Yi, G.Y., Rao, J.N.K. et Li, H. (2016). A weighted composite likelihood approach for analysis of survey data under two-level models. *Statistica Sinica*, 26, 569-587.