Un nouvel estimateur pour l'information de Fisher dans les modèles à variables latentes

Estelle Kuhn, Maud Delattre

Une nouvelle méthode d'estimation de la matrice d'information de Fisher dans les modèles à variables latentes

Maud Delattre [1] & Estelle Kuhn [2]

[1] *UMR518 AgroParisTech/INRA, AgroParisTech, 16 rue Claude Bernard, 75231 Paris Cedex 05*
maud.delattre@agroparistech.fr
[2] *INRA, MaIAGE, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex*
estelle.kuhn@inra.fr

Résumé. La matrice d'information de Fisher (FIM) joue un rôle fondamental en statistiques. Elle est très largement utilisée, pour évaluer la précision d'estimateurs, mettre en œuvre certains tests statistiques ou encore planifier des expériences. Dans la plupart des modèles à variables latentes, les variables non observées rendent difficile l'évaluation de la FIM. Plusieurs méthodes ont été proposées pour évaluer la FIM observée, parmi lesquelles des approches de type Monte-Carlo ou des algorithmes itératifs basés sur le "missing information principle". Celles-ci requièrent l'existence et le calcul des dérivées secondes de la log-vraisemblance complète. Cependant, ces dérivées n'existent pas toujours et les calculer peut présenter de nombreux désavantages computationnels. Nous considérons la matrice de covariance empirique du score comme estimateur de la FIM observée. Cet estimateur, très peu utilisé en pratique, a l'avantage de ne nécessiter que le calcul des dérivées premières de la log-vraisemblance complète. Lorsque celui-ci n'admet pas d'expression explicite, nous proposons un algorithme d'approximation stochastique afin de l'évaluer. Nous étudions les propriétés asymptotiques de cet estimateur et celles de l'algorithme stochastique proposé. Nous évaluons les propriétés de convergence de l'estimateur à distance finie sur des données simulées selon un modèle linéaire à effets mixtes d'une part, et selon un modèle de mélange d'autre part. Dans ces deux modèles, l'estimateur peut être calculé explicitement, ainsi que la FIM observée. Nous illustrons les propriétés de convergence de l'algorithme d'estimation via des données simulées selon un modèle non linéaire à effets mixtes. Nous calculons également la FIM observée pour un jeu de données de théophylline ajusté avec ce modèle.

Mots-clés. Matrice d'information de Fisher, approximation stochastique, modèle à variables latentes, estimation

Abstract. The Fisher information matrix (FIM) plays a key role in statistics. It is crucial in practice not only for evaluating the precision of parameter estimates, but also for statistical testing or experimental designs. In many latent variable models, the evaluation of the FIM is non-elementary due to the unobserved variables. Several methods have been proposed to estimate the observed FIM. Among the most frequently used approaches are Monte-Carlo methods or iterative algorithms derived from the missing information principle. These methods require to compute second derivatives of the complete data log-likelihood. However these derivatives do not always exist or their evaluations have some disadvantages from a computational point of view. We consider the empirical first order moment estimate of the covariance matrix of the score. This estimator of the observed FIM only requires to compute the first derivatives of the complete data log-likelihood with respect to the parameters. When it does not admit an analytical expression, we derive a stochastic approximation algorithm for computing its value. We establish the asymptotic properties of the proposed estimator and of the corresponding stochastic approximation algorithm. Some numerical experiments are performed in linear mixed effects models and mixture models to illustrate the finite sample size properties of the estimator. We also carry out a simulation study in a nonlinear mixed effects model to highlight the convergence of the proposed estimation algorithm. Finally we evaluate the observed FIM for a real dataset of theophylline adjusted with a nonlinear mixed effects model.

Keywords. Fisher information matrix, stochastic approximation, latent variables, estimation

1 Introduction

La matrice d'information de Fisher (FIM) est une quantité fondamentale en statistique. À titre d'exemple, son inverse, égale à la borne de Cramer-Rao, est la matrice de variance-covariance asymptotique de l'estimateur du maximum de vraisemblance dans les modèles paramétriques réguliers. La FIM permet en particulier d'évaluer très généralement la précision des estimateurs. Par ailleurs, la FIM intervient dans plusieurs tests statistiques d'utilisation courante tels que le test de Wald et le test du rapport de vraisemblance, soit dans l'expression de la statistique de test, soit dans la distribution asymptotique de cette statistique. La FIM apparaît également dans des résultats asymptotiques post-sélection de modèles. Il est donc primordial de pouvoir l'évaluer efficacement.

2 Définition et estimateurs de la matrice d'information de Fisher

Considérons une variable aléatoire Y dont la densité g dépend d'un paramètre $\theta \in \Theta \subset \mathbb{R}^d$ pour $d \in \mathbb{N}^*$. Dans ce cadre paramétrique très général, la matrice d'information de Fisher quantifie l'information relative à θ portée par la variable Y. Sous des conditions générales sur le modèle, la FIM se définit originellement comme la matrice de covariance du score :

$$\mathcal{I}(\theta) = \mathbb{E}_\theta \left[\frac{\partial}{\partial\theta}(\log g(Y;\theta)) \frac{\partial}{\partial\theta}(\log g(Y;\theta))^t \right]. \tag{1}$$

Lorsque la fonction de log-vraisemblance $\log g$ est deux fois différentiable sur Θ, on dispose également de l'expression suivante :

$$\mathcal{I}(\theta) = -\mathbb{E}_\theta \left[\frac{\partial^2}{\partial\theta^2} (\log g(Y;\theta)) \right]. \tag{2}$$

Cependant les expressions (1) et (2) ne sont pas toujours simples à calculer. Ainsi, en pratique, on s'intéresse généralement plutôt à la FIM observée, qui fournit une estimation de la matrice d'information de Fisher théorique à partir d'un échantillon $(y_1, \ldots, y_n)$ de Y. Aux deux expressions de la FIM théorique présentées ci-dessus correspondent deux versions empiriques de la FIM observée données pour tout $\theta \in \Theta$ par les deux expressions $V_n(\theta)$ et $W_n(\theta)$ ci-dessous :

$$\begin{aligned} V_n(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}(\log g(y_i;\theta)) \frac{\partial}{\partial\theta}(\log g(y_i;\theta))^t \\ W_n(\theta) &= -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}(\log g(y_i;\theta)) \end{aligned}$$

Ces deux M-estimateurs de la FIM sont consistants et les vecteurs de leurs composantes sont asymptotiquement normaux. L'estimateur $W_n(\theta)$ est le plus utilisé en pratique. Cependant il recquiert l'existence des dérivées secondes de la log-vraisemblance complète, et nécessite leurs calculs. Or certains modèles ne vérifient pas cette hypothèse de régularité. En particulier certains modèles mécanistes, comme par exemple des modèles de croissance des plantes, sont peu réguliers en les paramètres et ne satisfont pas cette hypothèse. L'estimateur $V_n(\theta)$ ne nécessite lui que l'existence et le calcul des dérivées premières de la log-vraisemblance complète.

3 Estimation dans un modèle à variables latentes

On considère dans la suite un modèle à variables latentes dans lequel les données complètes sont notées (Y, Z), où Y désigne la variable observée et Z la variable latente. Le modèle est caractérisé par la distribution jointe de (Y, Z). On suppose que cette distribution admet une densité paramétrique, notée f et caractérisée par le paramètre θ, par rapport à une mesure de référence μ. Dans ce cas, il est fréquent que la vraisemblance des observations Y ne soit pas explicite car elle s'exprime comme une intégrale qui a rarement une forme close :

$$g(Y; \theta) = \int f(Y, Z; \theta)\mu(dZ).$$

Lorsque les dérivées premières et secondes de la log-vraisemblance n'admettent pas de forme explicite, les estimateurs $V_n(\theta)$ et $W_n(\theta)$ définis plus haut ne sont pas directement calculables. La formule de Louis (1982) permet dans ce cas d'exprimer la FIM observée au moyen des moments conditionnels des dérivées premières et secondes de la log-vraisemblance des données complètes. De cette formule dérive en particulier une méthode d'évaluation de $W_n(\theta)$ par approximation stochastique combinant un algorithme de type Robbins Monro à la simulation de réalisations des variables latentes sous leur loi conditionnelle par rapport aux observations (Delyon et al., 1999). Cette approche est très largement utilisée en pratique, notamment dans le cadre des modèles non linéaires à effets mixtes (Lavielle, 2014). Cette méthode nécessite néanmoins de dériver deux fois la log-vraisemblance des données complètes, ce qui ne permet pas de l'appliquer dans les modèles qui ne satisfont pas cette hypothèse.

Nous considérons dans la suite l'estimateur $V_n(\theta)$ qui existe dès que la FIM existe. En remarquant que

$$\frac{\partial}{\partial\theta}(\log g(Y;\theta)) = \mathbb{E}_{Z|Y;\theta}\left[\frac{\partial}{\partial\theta}(\log f(Z,Y;\theta))\right]$$

nous obtenons l'expression suivante pour l'estimateur $V_n(\theta)$:

$$V_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{Z_i|Y_i;\theta}\left[\frac{\partial}{\partial\theta}(\log f(Z_i,Y_i;\theta))\right] \mathbb{E}_{Z_i|Y_i;\theta}\left[\frac{\partial}{\partial\theta}(\log f(Z_i,Y_i;\theta))\right]^t$$

Dans certains modèles comme les modèles linéaires à effets mixtes ou les modèles de mélange, cette expression admet une forme explicite. Lorsque cette expression ne peut être calculée, nous proposons une nouvelle méthode itérative pour évaluer simultanément l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ et l'estimateur $V_n(\hat{\theta})$, inspirée de celle de Delyon et al. (1999). Cette approche ne nécessite que le calcul des dérivées premières de la log-vraisemblance des données complètes. Pour des valeurs initiales θ_0, Q_0, Δ_0 et un entier K, on répète pour k variant de 1 à K les étapes suivantes à chaque itération de l'algorithme :

- **Simulation des variables latentes :** pour $1 \leq i \leq n$, $Z_i^k \sim p(\cdot|Y_i;\theta_{k-1})$.
- **Approximation stochastique de la log-vraisemblance complète et de sa dérivée :** pour $1 \leq i \leq n$, pour $\theta \in \Theta$, calcul des quantités

$$\begin{aligned} Q^k(\theta) &= (1-\gamma_k)Q^{k-1}(\theta) + \gamma_k \log f(Z_i^k, Y_i; \theta) \\ \Delta_i^k &= (1-\gamma_k)\Delta_i^{k-1} + \gamma_k \partial_\theta \log f(Z_i^k, Y_i; \theta_{k-1}) \end{aligned}$$

  où (γ_k) est une suite de pas positive décroissante telle que pour tout $k \in \mathbb{N}$, $0 \leq \gamma_k \leq 1$, $\sum \gamma_k = \infty$, $\sum \gamma_k^2 < \infty$.
- **Mise à jour du paramètre :**

$$\theta_k = \arg\max_\theta Q^k(\theta)$$

Pour K assez grand, on calcule ensuite la quantité :

$$V_n^K = \frac{1}{n}\sum_{i=1}^{n} \Delta_i^K (\Delta_i^K)^t \tag{3}$$

qui est un estimateur de $V_n(\hat{\theta})$. Nous montrons la convergence presque sûre de cet algorithme sous des hypothèses générales de régularité du modèle.

4 Expériences numériques

Nous illustrons sur des données simulées les performances à distance finie de l'estimateur $V_n(\theta)$ dans un modèle linéaire à effets mixtes et dans un modèle de mélange. Nous réalisons ensuite une étude numérique dans un modèle non linéaire à effets mixtes pour évaluer la convergence de l'algorithme stochastique d'estimation. Nous ajustons finalement un jeu de données réelles de théophylline par un modèle non linéaire mixte et estimons la FIM observée.

Bibliographie

[1] Louis, T.A. (1982) *Finding the Observed Information Matrix when Using the EM Algorithm*, *JRSS B*, 44 (2), 226–233.
[2] Delyon, B., Lavielle, M. and Moulines, E. (1999) *Convergence of a stochastic approximation version of the EM algorithm*, *Ann. Statist.*, 1999, 27 (1), 94–128.
[3] Lavielle, M. (2014) *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*, Chapman and Hall.