

Mise en oeuvre de méthodes semi-Bayésiennes de calcul des erreurs standards pour les données éparses dans le package `saemix`

Mélanie Guhl^{1,*}, Lucie Fayette¹, Julie Bertrand¹, and Emmanuelle Comets^{1,2}

*melanie.guhl@inserm.fr

¹Université Paris Cité, Inserm, IAME, F-75018 Paris, France

²Univ Rennes, Inserm, EHESP, Irset - UMR_S 1085, F-35000 Rennes, France

Résumé

Le package `saemix` implémente l'algorithme fréquentiste Stochastic Approximation Expectation Maximisation (SAEM) qui permet l'estimation de modèles non linéaires à effets mixtes. Ce type de modèle est très utilisé en biostatistiques pour la modélisation de données longitudinales, c'est-à-dire répétées dans le temps, par exemple pour étudier l'évolution clinique d'un patient, d'un biomarqueur ou la dynamique d'une réponse à un traitement. La méthode classique de calcul de l'incertitude autour de l'estimateur du maximum de vraisemblance (EMV) obtenu avec SAEM est basée sur la matrice d'information de Fisher (FIM), dont l'inverse est sa limite asymptotique. Ici, nous proposons une nouvelle approche dite semi-Bayésienne intégrée à l'algorithme SAEM pour échantillonner dans la distribution de l'EMV. Nous explorons deux algorithmes Bayésiens sur plusieurs jeux de simulations et des données issues d'un essai clinique sur des données éparses, c'est-à-dire éloignées de l'asymptotique, pour lesquelles la méthode basée sur la FIM a montré ses limites. Cette approche permet une meilleure couverture des paramètres que la méthode asymptotique, et sur le jeu de données réel, de fortes corrélations entre les paramètres donnent l'avantage à l'algorithme Approximate Bayesian Computation (ABC) n'utilisant pas les vraisemblances.

Mots-clés : Package `saemix` – Modèles non linéaires à effets mixtes – Erreurs standards – Inférence semi-Bayésienne – Données éparses

Introduction

Les modèles non linéaires à effets mixtes (NLMEM) sont utilisés pour modéliser des données longitudinales en biostatistiques capturant par exemple, la progression d'une maladie et/ou la réponse à un traitement. Les paramètres sont généralement estimés par le maximum de vraisemblance (EMV) avec l'algorithme Stochastic Approximation Expectation Maximization (SAEM), et leur incertitude via la matrice d'information de Fisher (FIM). Sur des petits échantillons (petit nombre de sujets N) et des designs épars (petit nombre d'observations par sujet n), la FIM peut sous-estimer l'incertitude [Loingeville et al., 2020], et échantillonner dans la distribution de l'EMV grâce à l'inférence Bayésienne a montré de meilleurs résultats dans ce cas, en utilisant l'algorithme Hamiltonian Monte Carlo (HMC) implémenté dans le logiciel *Stan* après estimation avec SAEM [Guhl et al., 2022]. Dans ce travail, nous proposons d'intégrer une méthode semi-Bayésienne dans le package R `saemix` [Comets et al., 2017] qui implémente SAEM.

Méthodes

Nous avons implémenté deux approches semi-Bayésiennes, utilisant l'algorithme de Metropolis-Hastings (SAEM_MH) [Guhl et al., in press] et l'Approximate Bayesian Computation (SAEM_ABC),

et testé différentes variations de ces méthodes. Nous les avons comparées avec des méthodes fréquentistes plus classiques (FIM, sampling importance resampling et bootstrap) et une autre méthode semi-Bayésienne utilisant l’algorithme HMC (plus complexe à réimplémenter) (Post). Nous avons comparé les taux de couverture (c’est-à-dire la proportion de jeux de données pour lesquels l’intervalle de confiance calculé sur un paramètre recouvre sa vraie valeur) et les erreurs standards (SE) relatives obtenues sur une étude de simulation de 500 jeux de données pharmacocinétiques (PK, $N=12$, $n=3$) dans deux scénarios, l’un étant plus complexe (fortes variabilités inter-individuelles, corrélations entre les effets aléatoires). Toutes ces méthodes ont été également appliquées sur des données réelles issues d’un essai de PK clinique comparant deux formes galéniques d’un anticorps monoclonal.

Résultats

SAEM_MH a donné de bons résultats sur le premier scénario de simulations mais présenté des limites face à des structures de variabilité complexes, les taux d’acceptation très bas indiquant le manque de fiabilité de la méthode. Utiliser un échantillonnage par blocs et une marche aléatoire a permis d’améliorer les taux d’acceptation, mais les SE et les taux de couverture étaient toujours sous-estimés. Post a également été mis en difficulté sur les corrélations entre effets aléatoires, contrairement à SAEM_ABC qui a présenté de meilleurs résultats et est également efficace en temps de calcul.

Sur les données PK de l’anticorps monoclonal ($N=24$ sujets par bras, $n=11$), toutes les méthodes ont donné des résultats concordants, par exemple des RSE entre 5 et 6% pour la clairance du médicament, suggérant que les conditions asymptotiques sont atteintes dans ce cas. Sur un sous-ensemble de ces données ($N=6$ sujets par bras, $n=11$), des différences ont été observées entre les méthodes, par exemple des RSE de 1 à 12% pour la clairance, illustrant le besoin d’identifier une méthode de calcul de l’incertitude fiable sur les petits jeux de données.

Conclusion

Les approches semi-Bayésiennes que nous avons implémentées dans l’algorithme SAEM semblent prometteuses pour le calcul de l’incertitude à distance finie. Cependant, l’algorithme MH présente des limites dues à la dimensionalité du vecteur à échantillonner. Ces difficultés sont partiellement surmontées par des variations de la méthode permettant de diminuer la dimension du vecteur de paramètres à échantillonner et ainsi de remonter les taux d’acceptation. L’algorithme ABC semble le plus prometteur, mais des travaux supplémentaires sont nécessaires pour fournir des recommandations plus complètes sur sa calibration.

Références

- E Comets, A Lavenu, and M Lavielle. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation. *Journal of Statistical Software*, 80 :1–41, 2017.
- M Guhl, F Mercier, C Hofmann, S Sharan, M Donnelly, K Feng, W Sun, G Sun, S Grosser, L Zhao, L Fang, F Mentré, E Comets, and J Bertrand. Impact of model misspecification on model-based tests in PK studies with parallel design : real case and simulation studies. *Journal of Pharmacokinetics and Pharmacodynamics*, 49(5) :557–577, 2022.
- M Guhl, J Bertrand, L Fayette, F Mercier, and E Comets. Uncertainty computation at finite distance in nonlinear mixed effects models - a new method based on metropolis hastings algorithm. *The AAPS Journal*, in press.
- F Loingeville, J Bertrand, TT Nguyen, S Sharan, K Feng, W Sun, J Han, S Grosser, L Zhao, L Fang, K Möllenhoff, H Dette, and F Mentré. New Model-Based Bioequivalence Statistical Approaches for Pharmacokinetic Studies with Sparse Sampling. *The AAPS Journal*, 22(6) :141, 2020.