HAL
open science

lcmm : un package R pour l'estimation des modèles mixtes à classes latentes et des modèles conjoints à classes latentes pour données répétées Gaussiennes, ordinales ou curvilinéaires et données de survie

Cécile Proust-Lima, Amadou Diakité, Benoit Liquet

To cite this version:

Cécile Proust-Lima, Amadou Diakité, Benoit Liquet. lcmm : un package R pour l'estimation des modèles mixtes à classes latentes et des modèles conjoints à classes latentes pour données répétées Gaussiennes, ordinales ou curvilinéaires et données de survie. 1ères Rencontres R, Jul 2012, Bordeaux, France. hal-00717553

HAL Id: hal-00717553

https://hal.science/hal-00717553v1

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

lcmm: un package *R* pour l'estimation des modèles mixtes à classes latentes et des modèles conjoints à classes latentes pour données répétées Gaussiennes, ordinales ou curvilinéaires et données de survie

C. Proust-Lima[a], **A. Diakité**[b] et **B. Liquet**[c]

[a,b,c]INSERM U897
ISPED, Univ. Bordeaux
F-33000 Bordeaux
[a]cecile.proust-lima@inserm.fr
[a]amadou.diakite@isped.u-bordeaux2.fr
[c]benoit.liquet@isped.u-bordeaux2.fr

Mots clefs : classification, données censurées, données hétérogènes, données longitudinales, modèles à effets aléatoires, modèles conjoints, modèles de mélange, processus latent.

Le modèle linéaire mixte [1], couramment utilisé pour décrire le changement au cours du temps d'une variable quantitative dans les études longitudinales, est fondé sur plusieurs hypothèses :

(i) normalité des erreurs et des effets aléatoires

(ii) linéarité des relations avec le marqueur longitudinal

(iii) continuité du marqueur longitudinal

(iv) homogénéité de la population d'étude

(v) données manquantes aléatoires

Le package *R* **lcmm** permet l'estimation de plusieurs extensions du modèle linéaire mixte:

- le modèle mixte à classes latentes [2-4] qui relache l'hypothèse (iv) pour analyser les changements au cours du temps d'un marqueur longitudinal en population hétérogène;
- le modèle conjoint à classes latentes [5] qui relache les hypothèses (iv) et (v) pour analyser conjointement un marqueur longitudinal et un délai d'événement censuré à droite et potentiellement tronqué à gauche;
- le modèle mixte à processus latent [6,7] qui relache les hypothèses (i),(ii) et (iii) (ainsi que l'hypothèse (iv) si besoin) pour analyser les changements au cours du temps de marqueurs quantitatifs potentiellement curvilinéaires, et de marqueurs ordinaux.

Les modèles mixtes à classes latentes consistent à explorer les profils latents de trajectoires qui peuvent exister dans une population hétérogène. Ces modèles combinent la théorie des modèles mixtes pour tenir compte de la corrélation individuelle entre les mesures répétées du marqueur et les modèles à classes latentes pour discriminer des groupes homogènes de sujets. Malgré leur intérêt en pratique, leur implémentation dans les logiciels gratuits est encore limitée. Au sein du package **lcmm**, la fonction `hlme` estime des modèles linéaires mixtes à classes latentes pour des données "gaussiennes" (i.e. sous les hypothèses (i),(ii) et (iii)), et la fonction `lcmm` estime des modèles mixtes à classes latentes pour des données continues curvilinéaires ou ordinales.

Les modèles pour données curvilinéaires et ordinales sont des modèles mixtes dits “à processus latent” qui font intervenir des fonctions de lien paramétrées pour lier les observations curvilinéaires ou ordinales à leur processus latent continu sous-jacent.

Il existe principalement deux types de modèles conjoints pour l'analyse jointe de données répétées et de données de survie : les modèles à effets aléatoires partagés dans lesquels des fonctions de la trajectoire du marqueur sont incluses dans le modèle de survie, et les modèles à classes latentes qui font l'hypothèse qu'une structure en classes latentes capture toute la corrélation entre les données répétées du marqueur et le risque d'événement. Nous avons implémenté ces derniers dans la fonction `Jointlcmm` du package **lcmm**. Notons que les modèles à effets aléatoires partagés peuvent aussi être estimés sous *R* avec le package **JM** [8].

La méthode d'estimation implémentée dans les fonctions `hlme`, `lcmm` et `Jointlcmm` est le maximum de vraisemblance par un algorithme de Marquardt [9] modifié avec des critères d'arrêt stricts (sur les dérivées premières et secondes) [3,6]. N'importe quelle forme de trajectoire peut être modélisée et des variables explicatives peuvent être incluses dans toutes les parties des modèles (avec ou sans effets spécifiques aux classes). Plusieurs fonctions de risque de base sont implémentées dans `Jointlcmm` (M-splines, Weibull, exponentiel par morceaux) et plusieurs fonctions paramétrées de lien sont implémentées dans la fonction `lcmm` pour tenir compte de la curvilinéarité (I-splines, FdR Beta, linéaire, modèle à seuils). Enfin, des fonctions sont disponibles pour évaluer l'adéquation des modèles, leur qualité prédictive, et les classifications a posteriori [5].

Références

[1] Laird N & Ware J (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-74.

[2] Verbeke G & Lesaffre E (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, **91**, 217-21.

[3] Muthén B & Shedden K (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, **55**, 463-9.

[4] Proust C & Jacqmin-Gadda H (2005). Estimation of linear mixed models with a mixture of distribution for the random-effects. *Computer Methods and Programs in Biomedicine*,**78**,165-73.

[5] Proust-Lima C, Séne M, Taylor JMG & Jacqmin-Gadda H (2012). Joint latent class models for longitudinal and time-to-event data: a review, *Statistical Methods in Medical Research*, in press.

[6] Proust C, Jacqmin-Gadda H, Taylor JMG, Ganiayre J & Commenges D (2006). A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*,**62**, 1014-24.

[7] Proust-Lima C, Dartigues J-F & Jacqmin-Gadda H (2011). Misuse of the linear mixed model when evaluating risk factors of cognitive decline. *American Journal of Epidemiology*, **174**(9),1077-88.

[8] Rizopoulos D (2010). JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* **35**(9), 1-33.

[9] Marquardt D (1963). An algorithm for leastsquares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, **11**, 431-41.