

Modélisation conjointe de données longitudinales et de temps d'événements sous R

Cécile Proust-Lima*

Résumé

Les études en santé impliquent généralement la collecte de variables mesurées de manière répétée au fil du temps. Cela inclut des expositions (e.g., traitement, pression artérielle, nutrition), des marqueurs de progression (e.g., volumes cérébraux, marqueurs sanguins, taille des tumeurs, score de qualité de vie) et des délais jusqu'à des événements cliniques (e.g., décès, diagnostic, rechute). L'analyse jointe de ces données longitudinales et de ces temps d'événement se fait par des modèles dits conjoints qui prennent en compte la corrélation entre les processus en jeu¹. Les modèles conjoints sont devenus au fil des ans un outil essentiel en biostatistique car ils permettent d'aborder diverses questions prédictives, descriptives et analytiques. Cela inclut la prédiction du risque d'événement basée sur des marqueurs ou expositions mesurés de manière répétée au fil du temps, la modélisation de la progression de marqueurs tout en tenant compte d'une sortie d'étude informative, ou la description et compréhension de la structure d'interdépendance qui peut exister entre plusieurs processus.

Dans cette présentation, je vais introduire le principe de la modélisation jointe et décrire différentes approches proposées dans la littérature en m'appuyant sur les solutions R associées. Seront abordés les modèles à effets aléatoires partagés pour lesquels divers packages R existent (e.g., JM¹, JM-bayes2, INLAjoint²), les modèles à classes latentes avec le package lcmm^{3,4}, ainsi qu'une alternative aux modèles conjoints via des forêts aléatoires de survie avec le package DynForest^{5,6}.

Mots-clés

biostatistique, modèles mixtes, classes latentes, forêts aléatoires, modèles conjoints

Références

1. Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*. (CRC Press, Boca Raton, 2012).
2. Rustand, D., van Niekerk, J., Krainski, E. T., Rue, H. & Proust-Lima, C. Fast and flexible inference for joint models of multivariate longitudinal and survival data using integrated nested Laplace approximations. *Biostatistics* kxad019 (2023) doi:10.1093/biostatistics/kxad019.
3. Proust-Lima, C., Philipps, V. & Liqueur, B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. *Journal of Statistical Software, Articles* **78**, 1–56 (2017).
4. Proust-Lima, C. *et al.* Describing complex disease progression using joint latent class models for multivariate longitudinal markers and clinical endpoints. *Stat Med* **42**, 3996–4014 (2023).
5. Devaux, A., Proust-Lima, C. & Genuer, R. Random Forests for time-fixed and time-dependent predictors: The DynForest R package. *arXiv* (2023) doi:10.48550/arXiv.2302.02670.
6. Devaux, A., Helmer, C., Genuer, R. & Proust-Lima, C. Random survival forests with multivariate longitudinal endogenous covariates. *Stat Methods Med Res* **32**, 2331–2346 (2023).

* Univ. Bordeaux, Inserm, Bordeaux Population Research Center, Bordeaux, France, cecile.proust-lima@inserm.fr