

BeQut, un package R pour l'estimation bayésienne de modèles de régression quantile à effets mixtes via JAGS

Antoine Barbieri*

Christophe Tzourio[†]

Hélène Jacqmin-Gadda[‡]

Résumé

Le package `BeQut` permet l'estimation de modèles de régression quantile tels que la régression linéaire quantile simple ou à effets mixtes, ou encore les modèles conjoints pour données longitudinales et temps jusqu'à événement. Ce dernier combine un modèle de régression quantile à effets mixtes et un modèle à hasard proportionnel afin d'évaluer l'influence de l'évolution d'un quantile sur le risque de survenu d'un événement d'intérêt. Pour l'inférence statistique, il est classique de supposer un cadre paramétrique basé sur une distribution de probabilité. En régression quantile, la distribution naturelle est la distribution asymétrique de Laplace. Le package `BeQut` repose sur une procédure d'estimation bayésienne dont la vraisemblance du modèle découle d'une réécriture de la distribution asymétrique de Laplace. Chacun des modèles de régression quantile est défini comme un modèle hiérarchique bayésien et les échantillons *a posteriori* des paramètres sont obtenus via `JAGS`. La présentation de ce package permettra de : (1) discuter des avantages et inconvénients de la procédure d'estimation proposée, (2) illustrer son intérêt au travers différentes applications sur des données réelles, et (3) faire un retour d'expérience concernant la soumission au CRAN d'un package R utilisant `JAGS`.

Mots-clefs : Régression quantile – Effets mixtes – Modélisation conjointe – `JAGS` – Package R.

From motivations to BeQut

Le développement du package `BeQut` a été motivé par le fait qu'il n'existait pas de package permettant l'estimation de modèles conjoints pour les données longitudinales et temps jusqu'à événements basé sur la régression quantile. Les modèles conjoints pour les données longitudinales et les temps jusqu'à événements constituent un domaine de recherche très actif en biostatistique. Ces modèles sont nécessaires pour étudier les facteurs de risque variables dépendants du temps en tant que déterminants des événements de santé. Ils évitent les biais des analyses de survie standard telles que le modèle de Cox avec des variables explicatives dépendantes du temps. Les modèles conjoints combinent un modèle mixte pour l'évolution dans le temps du marqueur et un modèle de survie pour le risque d'événement incluant des fonctions des trajectoires des marqueurs comme variables explicatives [Rizopoulos, 2012]. Des packages R sont disponibles pour ajuster ces modèles (`JMBayes2`, `JoinerML`...), mais tous se focalisent sur l'évolution moyenne du marqueur et sur comment celle-ci impacte le risque de développer l'événement. Dans `BeQut`, nous proposons aux utilisateurs une alternative qui s'intéresse à l'évolution de quantile de la distribution du marqueur plutôt qu'à sa moyenne [Yang et al., 2019].

Par ailleurs, le package `BeQut` repose sur une procédure d'estimation bayésienne qui n'est pas utilisée dans les packages R traitant la régression linéaire quantile à effets mixtes (`lqmm` et `qrLMM`). Il permet également de proposer une alternative à ces packages basée sur une autre procédure d'estimation qui (dans certains cas) a montré de meilleurs résultats dans une étude de simulation. Par ailleurs, la procédure d'estimation pour la modélisation conjointe a été validée par simulation.

*Univ. Bordeaux, Inserm U1219, Bordeaux population health center, antoine.barbieri@u-bordeaux.fr

[†]Univ. Bordeaux, Inserm U1219, Bordeaux population health center, christophe.tzourio@u-bordeaux.fr

[‡]Univ. Bordeaux, Inserm U1219, Bordeaux population health center, helene.jacqmin-gadda@u-bordeaux.fr

Implémentation

Le package `BeQut` permet l'estimation de modèles de régression quantile tels que la régression linéaire quantile simple ou à effets mixtes, ou encore les modèles conjoints pour données longitudinales et temps jusqu'à événement [Barbieri and Jacqmin-Gadda]. L'estimation des modèles repose sur une procédure d'estimation bayésienne dont la vraisemblance découle de la distribution asymétrique de Laplace. Cette distribution est naturellement considérée pour l'inférence statistique en régression quantile [Koenker and Machado, 1999]. Chacun des modèles implémentés est défini comme un modèle hiérarchique bayésien et les échantillons *a posteriori* des paramètres sont obtenus en appelant le logiciel JAGS [Plummer, 2003] depuis R via le package `rjags`. JAGS est un programme permettant d'effectuer des simulations MCMC à partir des données et de la déclaration du modèle (définition de la vraisemblance et des distributions *a priori* des paramètres). Bien que l'utilisation de ce logiciel évite la programmation de l'algorithme MCMC pour d'obtenir les échantillons *a posteriori*, il présente tout de même quelques contraintes. Par exemple, les distributions de probabilité ne sont pas toutes définies dans JAGS. C'est notamment le cas pour la distribution asymétrique de Laplace ou celles permettant la définition des modèles de survie. Ainsi, une réécriture est nécessaire pour définir les modèles.

Illustrations sur données réelles

L'application du package sera présentée sur deux jeux de données réelles. Le premier jeu de données¹ permettra d'illustrer et de motiver l'utilisation de la régression linéaire quantile via la modélisation de la hauteur de vagues en fonction de la vitesse du vent. Le second, issu des données de l'essai clinique PROGRESS [Mahon et al., 2001], permettra d'illustrer la régression quantile pour la modélisation de la pression artérielle au cours du temps et aussi la modélisation conjointe entre les mesures répétées de la pression artérielle et le risque d'événements cardio- et cérébro-vasculaires.

Remarque : tous les packages R cités dans ce résumé n'apparaissent pas dans les références par souci de place. Cependant, ils sont tous disponibles sur le CRAN où toutes les informations sont disponibles.

Références

- Antoine Barbieri and Hélène Jacqmin-Gadda. *BeQut : Bayesian Estimation for Quantile Regression Mixed Models*. R package version 0.1.0.
- Roger Koenker and José A. F. Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448) :1296–1310, 1999. doi : 10.1080/01621459.1999.10473882.
- S. Mac Mahon, S. Neal, C Tzourio, A. Rodgers, M. Woodward, J Cutler, C Anderson, and J Chalmers. Randomised trial of a perindopril-based blood-pressure-lowering regimen among 6105 individuals with previous stroke or transient ischaemic attack. *The Lancet*, 358(9287) :1033–1041, 2001.
- M Plummer. Jags : A program for analysis of bayesian graphical models using gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 2003.
- Dimitris Rizopoulos. *Joint Models for Longitudinal and Time-to-Event Data : With Applications in R*. CRC Press, June 2012.
- Ming Yang, Sheng Luo, and Stacia DeSantis. Bayesian quantile regression joint models : Inference and dynamic predictions. *Statistical Methods in Medical Research*, 28(8) :2524–2537, 2019.

1. données `wave` du package `BeQut` regroupant des données issues de la bouée houle du Cap Ferret (concernant la campagne 03302 Cap Ferret), et des données météorologiques `Infoclimat2` mesurées sur le site de Bordeaux ;