

Une bibliothèque Julia pour des modèles de régression bayésiens : comparaison avec les pratiques liées à R

Simon Devauchelle* Lucas Ondel Yang† David Doukhan‡ Nicolas Denier§
Albert Rilliard¶

Résumé

Nous présentons le module YABLR.jl de l'infrastructure logicielle *PTAL*¹, développée en Julia, qui permet de prototyper et d'appliquer des modèles de régression bayésienne. *PTAL* est un outil en développement qui fournit à la communauté des outils de traitement automatique de la parole libres et accessibles.

Les principales solutions pour appliquer des régressions bayésiennes à des données de sciences expérimentales (McElreath [2020]) reposent sur des outils tels que *Stan* (Team [2026]), interfacé en R notamment par les bibliothèques *rstan* (Stan Development Team [2025]), *rethinking* (McElreath [2024]) ou *brms* (Bürkner [2017]). La grande force de ces approches est de mettre à disposition des algorithmes permettant d'échantillonner (via la méthode de Monte Carlo hamiltonien) la distribution a posteriori de modèles complexes et flexibles.

Toutefois, un bon nombre de modèles n'a pas besoin de recourir à des outils aussi complexes et gourmands en ressources. En choisissant un prior dit "conjugué", il est souvent possible de trouver une forme analytique à la distribution postérieure. Cela simplifie à la fois l'estimation statistique et l'étude du modèle et permet d'enlever le bruit inhérent aux méthodes d'échantillonnage. En outre, cela permet de traiter des données en grande quantité, lorsque les approches à base d'échantillonnage sont limitées en mémoire et en temps de calcul. Le choix du prior "conjugué" est pleinement justifié d'un point de vue théorique car, pour un modèle donné, il est le prior d'entropie maximale. C'est cette possibilité que nous cherchons à rendre plus accessible grâce à la bibliothèque YABLR.jl de *PTAL*.

Pour illustrer son fonctionnement, nous montrerons un modèle de l'évolution diachronique de variables acoustiques liées à la parole (la fréquence fondamentale et certaines résonances du conduit vocal, voir Devauchelle et al. [2024, 2026]). Modéliser ces paramètres phonétiques dans le temps permet de questionner l'évolution de notre langue dans le cadre de son intégration sociale.

Mots-clés : Régression bayésienne – modélisation statistique – phonétique acoustique – diachronie – Julia

Références

Paul-Christian Bürkner. *brms* : An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1) :1–28, 2017. doi : 10.18637/jss.v080.i01.

Simon Devauchelle, Albert Rilliard, David Doukhan, and Lucas Ondel Yang. *Variation of Perceived Voice Pitch Across Time Periods, Gender, and Age in French Media Archives*, volume 12 of *Studi AISV*, page 47–71. Officinaventuno, IT, 2024. ISBN 978-88-97657-73-6. URL <https://doi.org/10.17469/02112AISV000004>.

*Université Paris Saclay, CNRS, LISN, Orsay & Institut National de l'Audiovisuel, Paris
simon.devauchelle@universite-paris-saclay.fr

†Université Paris Saclay, CNRS, LISN, Orsay lucas.ondel@cnrs.fr

‡Institut National de l'Audiovisuel, Paris ddoukhan@ina.fr

§Université Paris Saclay, CNRS, LISN, Orsay nicolas.denier@lisn.fr

¶Université Paris Saclay, CNRS, LISN, Orsay albert.rilliard@lisn.fr

1. <https://ptal.lisn.upsaclay.fr/>

Simon Devauchelle, Lucas Ondel-Yang, David Doukhan, Rémi Uro, Nicolas Denier, and Albert Rilliard. Une approche bayésienne pour modéliser l'évolution diachronique de la voix comme marqueur du genre en France. In *Journées d'Études sur la Parole 2026*, page soumis. AFCEP, 2026.

Richard McElreath. *Statistical Rethinking : A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC, 2 edition, March 2020. ISBN 978-0-429-02960-8. doi : 10.1201/9780429029608. URL <https://www.taylorfrancis.com/books/9780429642319>.

Richard McElreath. *rethinking : Statistical Rethinking book package*, 2024. URL <https://github.com/rmcelreath/rethinking>. R package version 2.42, commit ac1b3b2cda83f3e14096e2d997a6e30ad109eeee.

Stan Development Team. RStan : the R interface to Stan, 2025. URL <https://mc-stan.org/>. R package version 2.32.7.

Stan Development Team. *Stan Reference Manual*, 2026. URL <https://mc-stan.org>. version v2.38.0.