The jagstargets R package: a reproducible workflow framework for Bayesian data analysis with JAGS

William Michael Landau[1]

1 Eli Lilly and Company

Statement of need

Researchers who perform Bayesian statistics regularly experiment with models to refine, compare, and understand them ([Gelman et al., 2020](#)). Each fitted model informs subsequent model-building choices, and even for experienced practitioners, the investigation often leads to final models that differ from the ones originally proposed. Fitting a model usually means applying Markov chain Monte Carlo (MCMC) or a similar method to approximate the full joint posterior distribution of the parameters ([Gelman et al., 2014](#)). Flexible probabilistic programming languages such as JAGS have made model specification quick and straightforward ([Plummer & others, 2003](#)), but computation time is still a bottleneck. A workflow can take several minutes or hours to run, and in subsequent iterations, researchers struggle to keep the results up to date with frequent changes to the models, code, and data.

Summary

The `jagstargets` R package ([Landau, n.d.](#)) is a workflow toolkit for Bayesian data analysis with JAGS. It helps users express a Bayesian statistical modeling exercise as a formal pipeline with dedicated steps for data generation, analysis, and summarization. Pipelines can be customized for popular use cases: for example, the analysis of a single dataset using multiple alternative models, or a large simulation study to test that a model is implemented correctly ([Carpenter, 2017](#)). These pipelines, which can be visualized and executed using `targets` ([Landau, 2021b](#)), support key features that increase the efficiency and reproducibility of Bayesian workflows. The steps of a pipeline are automatically orchestrated using optional distributed computing, and up-to-date tasks are automatically skipped if the upstream code and data did not change since the last run. Thus, researchers can quickly iterate on Bayesian workflows while maintaining agreement between the results and the underlying code, models, and datasets.

The `jagstargets` package is a bridge between packages `R2jags` ([Su & Masanao Yajima, 2020](#)) and `targets`. `targets` is a general-purpose pipeline toolkit for reproducible research and high-performance computing. It can be used on its own for Bayesian data analysis, but it does not natively support domain-specific functionality for Bayesian workflows or JAGS. `jagstargets`, on the other hand, leverages the existing capabilities of `targets` to more easily and concisely express computational pipelines for Bayesian data analysis, from single analyses or large-scale simulation studies. It abstracts away laborious error-prone tasks such as invoking JAGS, tidying posterior samples, and batching thousands of MCMC runs to improve computational efficiency, all of which reduce the user-side burden of writing custom R code for a `targets` pipeline. `jagstargets` is similar to the `stantargets` R package ([Landau, 2021a](#)), the latter of which streamlines pipeline construction of Bayesian data analysis pipelines with Stan ([Stan Development Team, 2012](#)).

References

Carpenter, B. (2017). *Bayesian Posteriors are Calibrated by Definition.* https://statmodeling. stat.columbia.edu/2017/04/12/bayesian-posteriors-calibrated/

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (F. Dominici, J. J. Faraway, M. Tanner, & J. idek, Eds.; Third). CRC Press.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). *Bayesian Workflow.* http://arxiv. org/abs/2011.01808

Landau, W. M. (n.d.). *jagstargets: Targets for JAGS Workflows.*

Landau, W. M. (2021a). The stantargets R package: a workflow framework for efficient reproducible Stan-powered Bayesian data analysis pipelines. *Journal of Open Source Software*, *6*(60), 3193. https://doi.org/10.21105/joss.03193

Landau, W. M. (2021b). The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959. https://doi.org/10.21105/joss.02959

Plummer, M., & others. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd international workshop on distributed statistical computing*, *124*, 1–10.

Stan Development Team. (2012). *Stan: A C++ library for probability and sampling.* https://mc-stan.org.

Su, Y.-S., & Masanao Yajima. (2020). *R2jags: Using R to Run 'JAGS'.* https://CRAN. R-project.org/package=R2jags