Journal of Statistical Software

November 2021, Volume 100, Issue 1. *doi: 10.18637/jss.v100.i01*

Software for Bayesian Statistics

Michela Cameletti
Università degli Studi di Bergamo

Virgilio Gómez-Rubio
Universidad de Castilla-La Mancha

Abstract

In this summary we introduce the papers published in the special issue on Bayesian statistics. This special issue comprises 20 papers on Bayesian statistics and Bayesian inference on different topics such as general packages for hierarchical linear model fitting, survival models, clinical trials, missing values, time series, hypothesis testing, priors, approximate Bayesian computation, and others.

Keywords: Bayesian statistics, special issue, R, **JAGS**, **INLA**, Stan, **nimble**.

1. Introduction

Bayesian computation has experienced a surge in recent years (see, for example, Martin, Frazier, and Robert 2020, for a thorough historical review of Bayesian computation), driven by both methodological and software developments. In particular, the development of new computational methods such as the integrated nested Laplace approximation (INLA), Hamiltonian Monte Carlo, sequential Monte Carlo, or variational Bayes (to mention a few) has paved the way to new software tools.

For these reasons, in December 2018 a call for papers for contributions on a special issue on Bayesian statistics was announced worldwide. In the end, 20 papers have been accepted to be included in this special issue. Contributed papers span a wide range of topics, including software for hierarchical linear model fitting, time series, survival models, epidemiology, hypothesis testing, priors, approximate Bayesian computation, and many others.

All but one paper use the R software (R Core Team 2021) and the R packages described in this special issue contribute to the already rich task view dedicated to "Bayesian Inference" (Park 2021) available from the Comprehensive R Archive Network at `https://CRAN.R-project.org/view=Bayesian`.

2. Model fitting

The special issue includes a number of packages for fitting Bayesian hierarchical models using different types of software. Van Niekerk, Bakka, Rue, and Schenk (2021) present new developments for the **INLA** package (Lindgren and Rue 2015) about complex joint survival models, non-separable space-time models and high performance computing to fit very large models faster. Similarly, Michaud, De Valpine, Turek, Paciorek, and Nguyen (2021) describe the implementation of algorithms for state-space model analysis using sequential Monte Carlo methods for the **nimble** software which have been included in the **nimbleSMC** package. The **bamlss** package is described in Umlauf, Klein, Simon, and Zeileis (2021) to fit Bayesian generalized additive models for location, scale and shape (GAMLSS models). Bürkner (2021) describes how to fit Bayesian item response models using Stan (Carpenter *et al.* 2017) and the R package **brms** (Bürkner 2017). Similarly, Merkle, Fitzsimmons, Uanhoro, and Goodrich (2021) describe efficient Bayesian structural equation modeling using Stan and compare the new implementation with the already previously available one based on **JAGS** (Plummer 2003) in the R package **blavaan** (Merkle and Rosseel 2018). Dutta *et al.* (2021) describe the **ABCpy** scientific library for approximate Bayesian computation (ABC) that has been implemented using the Python programming language (Van Rossum *et al.* 2021).

3. Hierarchical linear models

Certain papers describe fitting particular types of Bayesian models. Fasiolo, Wood, Zaffran, Nedellec, and Goude (2021) focus on fitting Bayesian nonparametric quantile regression models using R package **qgam**, which is an extension of the well-known **mgcv** package for fitting generalized additive models (Wood 2017). Bonner, Kim, Westneat, Mutzel, Wright, and Schofield (2021) show how to fit double hierarchical linear models with the **dalmatian** package, which relies on **JAGS** and **nimble** (de Valpine, Turek, Paciorek, Anderson-Bergman, Temple Lang, and Bodik 2017) for model fitting. The **BayesSUR** R package for high-dimensional multivariate Bayesian variable and covariance selection for seemingly unrelated regression models is described in Zhao, Banterle, Bottolo, Richardson, Lewin, and Zucknick (2021).

4. Time series

A couple of papers in this special issue focus on time series analysis. Hosszejni and Kastner (2021) present an update of R package **stochvol** (Kastner 2016) to fit univariate stochastic volatility (SV) models (that now can handle linear mean models, conditionally heavy tails, and the leverage effect in combination with SV) and describe the **factorstochvol** R package for multivariate SV models. Time-varying parameter (TVP) models using global-local shrinkage priors to avoid overfitting with the **shrinkTVP** R package are described in Knaus, Bitto-Nemling, Cadonna, and Frühwirth-Schnatter (2021). Kuschnig and Vashold (2021) fit vector auto-regression (VAR) models for multivariate time series with hierarchical prior selection using the R package **BVAR**.

5. Survival models

Mayrink, Duarte, and Demarqui (2021) describe the implementation of a new **JAGS** module that implements the piece-wise exponential distribution so that it can be used when defining models using the BUGS language (Lunn, Spiegelhalter, Thomas, and Best 2009). This distribution if often employed to model the baseline hazard function when fitting Cox proportional hazards models in survival analysis.

6. Mixture models and clustering

Corradin, Canale, and Nipoti (2021) present the **BNPmix** R package for efficient Bayesian inference on nonparametric mixture models used for density estimation and clustering. Methods for model-based clustering of binary dissimilarity matrices implemented in the **dmbc** package are described in Venturini and Piccarreta (2021).

7. Hypothesis testing

Two papers in this special issue tackle the problem of hypothesis testing. Gronau, Raj K. N., and Wagenmakers (2021) show how to perform Bayesian inference for A/B tests using R package **abtest**, which allows for the incorporation of expert knowledge in the priors. Mulder *et al.* (2021) describe a general framework for testing hypotheses using Bayes factors for many different types of models.

8. Clinical trials

Weber, Li, Seaman, Kakizume, and Schmidli (2021) discuss the use of the meta-analytic predictive (MAP) approach to derive informative priors from historical data for clinical trials that has been implemented in the **RBesT** R package. Similarly, Eggleston, Ibrahim, McNeil, and Catellier (2021) describe the **BayesCTDesign** for the R programming language for two-arm randomized Bayesian clinical trials design, that can take advantage of historical control data when available.

9. Missing values

Erler, Rizopoulos, and Lesaffre (2021) deal with the problem of missing data and describe the **JointAI** R package to perform simultaneous analysis and imputation in regression models with incomplete covariates.

10. Discussion

Most of the papers in the special issue perform Bayesian inference by using Markov chain Monte Carlo (MCMC) algorithms. For simulating values from the posterior distributions, they use the BUGS language via **JAGS** (see e.g., Bonner *et al.* 2021; Erler *et al.* 2021; Mayrink *et al.* 2021; Weber *et al.* 2021), Stan (see e.g., Bürkner 2021; Merkle *et al.* 2021; Weber *et al.*

2021), or **nimble** (Michaud *et al.* 2021; Bonner *et al.* 2021), interfaced with R by means of the corresponding packages **rjags** (Plummer, Stukalov, and Denwood 2021), **rstan** (Stan Development Team 2021) and **nimble** (de Valpine *et al.* 2021). Alternatively, other papers (e.g., Corradin *et al.* 2021; Hosszejni and Kastner 2021; Knaus *et al.* 2021; Venturini and Piccarreta 2021) write sampling functions in C++ which are then integrated into R by using the **Rcpp** (Eddelbuettel and François 2011) and **RcppArmadillo** (Eddelbuettel and Sanderson 2014) packages. Finally, some papers do not use MCMC but numerical approximations such as Eggleston *et al.* (2021), Fasiolo *et al.* (2021) and Van Niekerk *et al.* (2021). Two papers (Kuschnig and Vashold 2021; Weber *et al.* 2021) also implement priors for specific Bayesian models.

The applications presented in the papers to illustrate the software encompass a wide range of disciplines including biology, ecology, medicine, economics, and sociology. This shows that Bayesian methods are extremely versatile and can be applied to any real data problem when it is necessary to make inference on unknown quantities by taking properly into account all the uncertainty sources. We hope that the papers included in the special issue will inspire researchers new to the Bayesian approach but also more expert users who are looking for new (and hopefully more efficient) software.

Acknowledgments

We would like to thank the editorial team of the Journal of Statistical Software for having welcomed our idea of a special issue on Bayesian statistics and for their continuous help and support along the entire publication process.

V. Gómez-Rubio has been supported by grant SBPLY/17/180501/000491, funded by Consejería de Educación, Cultura y Deportes (JCCM, Spain) and FEDER, and grant PID2019-106341GB-I00, funded by Ministerio de Ciencia e Innovación (Spain).

References

Bonner S, Kim HN, Westneat D, Mutzel A, Wright J, Schofield M (2021). "**dalmatian**: A Package for Fitting Double Hierarchical Linear Models in R via **JAGS** and **nimble**." *Journal of Statistical Software*, **100**(10), 1–25. doi:10.18637/jss.v100.i10.

Bürkner PC (2017). "**brms**: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software*, **80**(1), 1–28. doi:10.18637/jss.v080.i01.

Bürkner PC (2021). "Bayesian Item Response Modeling in R with **brms** and Stan." *Journal of Statistical Software*, **100**(5), 1–54. doi:10.18637/jss.v100.i05.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). "Stan: A Probabilistic Programming Language." *Journal of Statistical Software*, **76**(1), 1–32. doi:10.18637/jss.v076.i01.

Corradin R, Canale A, Nipoti B (2021). "**BNPmix**: An R Package for Bayesian Nonparametric Modeling via Pitman-Yor Mixtures." *Journal of Statistical Software*, **100**(15), 1–33. doi:10.18637/jss.v100.i15.

de Valpine P, Paciorek C, Turek D, Michaud N, Anderson-Bergman C, Obermeyer F, Wehrhahn Cortes C, Rodríguez A, Temple Lang D, Paganin S, Hug J (2021). ***nimble**: MCMC, Particle Filtering, and Programmable Hierarchical Modeling*. R package version 0.12.1, URL https://CRAN.R-project.org/package=nimble.

de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Temple Lang D, Bodik R (2017). "Programming with Models: Writing Statistical Algorithms for General Model Structures with **nimble**." *Journal of Computational and Graphical Statistics*, **26**(2), 403–413. doi:10.1080/10618600.2016.1172487.

Dutta R, Schoengens M, Pacchiardi L, Ummadisingu A, Widmer N, Künzli P, Onnela JP, Mira A (2021). "**ABCpy**: A High-Performance Computing Perspective to Approximate Bayesian Computation." *Journal of Statistical Software*, **100**(7), 1–38. doi:10.18637/jss.v100.i07.

Eddelbuettel D, François R (2011). "**Rcpp**: Seamless R and C++ Integration." *Journal of Statistical Software*, **40**(8), 1–18. doi:10.18637/jss.v040.i08.

Eddelbuettel D, Sanderson C (2014). "**RcppArmadillo**: Accelerating R with High-Performance C++ Linear Algebra." *Computational Statistics & Data Analysis*, **71**, 1054–1063. doi:10.1016/j.csda.2013.02.005.

Eggleston BS, Ibrahim JG, McNeil B, Catellier D (2021). "**BayesCTDesign**: An R Package for Bayesian Trial Design Using Historical Control Data." *Journal of Statistical Software*, **100**(21), 1–51. doi:10.18637/jss.v100.i21.

Erler NS, Rizopoulos D, Lesaffre EMEH (2021). "**JointAI**: Joint Analysis and Imputation of Incomplete Data in R." *Journal of Statistical Software*, **100**(20), 1–56. doi:10.18637/jss.v100.i20.

Fasiolo M, Wood SN, Zaffran M, Nedellec R, Goude Y (2021). "**qgam**: Bayesian Nonparametric Quantile Regression Modeling in R." *Journal of Statistical Software*, **100**(9), 1–31. doi:10.18637/jss.v100.i09.

Gronau QF, Raj K N A, Wagenmakers EJ (2021). "Informed Bayesian Inference for the A/B Test." *Journal of Statistical Software*, **100**(17), 1–39. doi:10.18637/jss.v100.i17.

Hosszejni D, Kastner G (2021). "Modeling Univariate and Multivariate Stochastic Volatility in R with **stochvol** and **factorstochvol**." *Journal of Statistical Software*, **100**(12), 1–34. doi:10.18637/jss.v100.i12.

Kastner G (2016). "Dealing with Stochastic Volatility in Time Series Using the R Package **stochvol**." *Journal of Statistical Software*, **69**(5), 1–30. doi:10.18637/jss.v069.i05.

Knaus P, Bitto-Nemling A, Cadonna A, Frühwirth-Schnatter S (2021). "Shrinkage in the Time-Varying Parameter Model Framework Using the R Package **shrinkTVP**." *Journal of Statistical Software*, **100**(13), 1–32. doi:10.18637/jss.v100.i13.

Kuschnig N, Vashold L (2021). "**BVAR**: Bayesian Vector Autoregressions with Hierarchical Prior Selection in R." *Journal of Statistical Software*, **100**(14), 1–27. doi:10.18637/jss.v100.i14.

Lindgren F, Rue H (2015). "Bayesian Spatial Modelling with R-**INLA**." *Journal of Statistical Software*, **63**(19), 1–25. doi:10.18637/jss.v063.i19.

Lunn D, Spiegelhalter D, Thomas A, Best N (2009). "The BUGS Project: Evolution, Critique and Future Directions." *Statistics in Medicine*, **28**(25), 3049–3067. doi:10.1002/sim.3680.

Martin GM, Frazier DT, Robert CP (2020). "Computing Bayes: Bayesian Computation from 1763 to the 21st Century." *arXiv 2004.06425*, arXiv.org E-Print Archive. URL https://arxiv.org/abs/2004.06425.

Mayrink VD, Duarte JDN, Demarqui FN (2021). "**pexm**: A **JAGS** Module for Applications Involving the Piecewise Exponential Distribution." *Journal of Statistical Software*, **100**(8), 1–28. doi:10.18637/jss.v100.i08.

Merkle EC, Fitzsimmons E, Uanhoro J, Goodrich B (2021). "Efficient Bayesian Structural Equation Modeling in Stan." *Journal of Statistical Software*, **100**(6), 1–22. doi:10.18637/jss.v100.i06.

Merkle EC, Rosseel Y (2018). "**blavaan**: Bayesian Structural Equation Models via Parameter Expansion." *Journal of Statistical Software*, **85**(4), 1–30. doi:10.18637/jss.v085.i04.

Michaud N, De Valpine P, Turek D, Paciorek CJ, Nguyen D (2021). "Sequential Monte Carlo Methods in the **nimble** and **nimbleSMC** R Packages." *Journal of Statistical Software*, **100**(3), 1–39. doi:10.18637/jss.v100.i03.

Mulder J, Williams DR, Gu X, Tomarken A, Böing-Messing F, Olsson-Collentine A, Meijerink M, Menke J, van Aert R, Fox JP, Hoijtink H, Rosseel Y, Wagenmakers EJ, Van Lissa C (2021). "**BFpack**: Flexible Bayes Factor Testing of Scientific Theories in R." *Journal of Statistical Software*, **100**(18), 1–63. doi:10.18637/jss.v100.i18.

Park JH (2021). *CRAN Task View: Bayesian Inference*. Version 2021-11-04, URL https://CRAN.R-project.org/view=Bayesian.

Plummer M (2003). "**JAGS**: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In K Hornik, F Leisch, A Zeileis (eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. Technische Universität Wien, Vienna, Austria. URL https://www.R-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.

Plummer M, Stukalov A, Denwood M (2021). ***rjags**: Bayesian Graphical Models Using MCMC*. R package version 4-12, URL https://CRAN.R-project.org/package=rjags.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Stan Development Team (2021). ***rstan**: R Interface to Stan*. R package version 2.21.2, URL https://CRAN.R-project.org/package=rstan.

Umlauf N, Klein N, Simon T, Zeileis A (2021). "**bamlss**: A Lego Toolbox for Flexible Bayesian Regression (and Beyond)." *Journal of Statistical Software*, **100**(4), 1–53. doi:10.18637/jss.v100.i04.

Van Niekerk J, Bakka H, Rue H, Schenk O (2021). "New Frontiers in Bayesian Modeling Using the **INLA** Package in R." *Journal of Statistical Software*, **100**(2), 1–28. doi:10.18637/jss.v100.i02.

Van Rossum G, *et al.* (2021). *Python Programming Language*. URL https://www.python.org/.

Venturini S, Piccarreta R (2021). "A Bayesian Approach for Model-Based Clustering of Several Binary Dissimilarity Matrices: The **dmbc** Package in R." *Journal of Statistical Software*, **100**(16), 1–35. doi:10.18637/jss.v100.i16.

Weber S, Li Y, Seaman JW, Kakizume T, Schmidli H (2021). "Applying Meta-Analytic-Predictive Priors with the R Bayesian Evidence Synthesis Tools." *Journal of Statistical Software*, **100**(19), 1–32. doi:10.18637/jss.v100.i19.

Wood SN (2017). *Generalized Additive Models: An Introduction with R*. 2nd edition. Chapman & Hall/CRC, Boca Raton.

Zhao Z, Banterle M, Bottolo L, Richardson S, Lewin A, Zucknick M (2021). "**BayesSUR**: An R Package for High-Dimensional Multivariate Bayesian Variable and Covariance Selection in Linear Regression." *Journal of Statistical Software*, **100**(11), 1–32. doi:10.18637/jss.v100.i11.

Affiliation:

Michela Cameletti
Department of Economics
Università degli Studi di Bergamo
Via dei Caniana 2
24127 Bergamo, Italy
E-mail: michela.cameletti@unibg.it
URL: https://www.unibg.it/ugov/person/2441

Virgilio Gómez-Rubio
Department of Mathematics
Escuela Técnica Superior de Ingenieros Industriales
Universidad de Castilla-La Mancha
02071 Albacete, Spain
E-mail: Virgilio.Gomez@uclm.es
URL: https://becarioprecario.github.io/

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
November 2021, Volume 100, Issue 1 *Submitted:* 2021-11-28
doi:10.18637/jss.v100.i01 *Accepted:* 2021-11-29