BGGM

Williams, Donald R.; Mulder, Joris

Published in:
The Journal of Open Source Software

DOI:
10.21105/joss.02111

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Williams, D. R., & Mulder, J. (2020). BGGM: Bayesian Gaussian Graphical Models in R. *The Journal of Open Source Software*, *5*(51), [2111]. https://doi.org/10.21105/joss.02111

The Journal of Open Source Software

BGGM: Bayesian Gaussian Graphical Models in R

Donald R. Williams[1] **and Joris Mulder**[2]

1 Department of Psychology, University of California, Davis **2** Department of Methodology and Statistics, Tilburg University

DOI: 10.21105/joss.02111

Software

- Review
- Repository
- Archive

Editor: Anisha Keshavan

Reviewers:

- @jayrobwilliams
- @paulgovan

Submitted: 15 January 2020
Published: 21 July 2020

BGGM: Bayesian Gaussian Graphical Models

The `R` package **BGGM** provides tools for making Bayesian inference in Gaussian graphical models (GGM). The methods are organized around two general approaches for Bayesian inference: (1) estimation and (2) hypothesis testing. The key distinction is that the former focuses on either the posterior or posterior predictive distribution (Gelman, Meng, & Stern, 1996; see section 5 in Rubin, 1984) , whereas the latter focuses on model comparison with the Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995).

What is a Gaussian Graphical Model ?

A Gaussian graphical model captures conditional (in)dependencies among a set of variables. These are pairwise relations (partial correlations) controlling for the effects of all other variables in the model.

Applications

The Gaussian graphical model is used across the sciences, including (but not limited to) economics (Millington & Niranjan, 2020), climate science (Zerenner, Friederichs, Lehnertz, & Hense, 2014), genetics (Chu, Weiss, Carey, & Raby, 2009), and psychology (Rodriguez, Williams, Rast, & Mulder, 2020).

Overview

The methods in **BGGM** build upon existing algorithms that are well-known in the literature. The central contribution of **BGGM** is to extend those approaches:

1. Bayesian estimation with the novel matrix-F prior distribution (Mulder & Pericchi, 2018)
   - Estimation (Williams, 2018)
2. Bayesian hypothesis testing with the matrix-F prior distribution (Williams & Mulder, 2019)
   - Exploratory hypothesis testing
   - Confirmatory hypothesis testing
3. Comparing Gaussian graphical models (Williams, 2018; Williams, Rast, Pericchi, & Mulder, 2020)

- Partial correlation differences
- Posterior predictive check
- Exploratory hypothesis testing
- Confirmatory hypothesis testing

4. Extending inference beyond the conditional (in)dependence structure (Williams, 2018)

- Predictability(e.g., Haslbeck & Waldorp, 2018)
- Posterior uncertaintyintervals for the partial correlations
- Custom Network Statistics

Supported Data Types

- **Continuous**: The continuous method was described in Williams (2018). Note that this is based on the customary Wishartdistribution.
- **Binary**: The binary method builds directly upon Talhouk, Doucet, & Murphy (2012) that, in turn, built upon the approaches of Lawrence, Bingham, Liu, & Nair (2008) and Webb & Forster (2008) (to name a few).
- **Ordinal**: The ordinal methods require sampling thresholds. There are two approach included in **BGGM**. The customary approach described in Albert & Chib (1993) (the default) and the 'Cowles' algorithm described in Cowles (1996).
- **Mixed**: The mixed data (a combination of discrete and continuous) method was introduced in Hoff (2007). This is a semi-parametric copula model (i.e., a copula GGM) based on the ranked likelihood. Note that this can be used for *only* ordinal data (not restricted to "mixed" data).

The computationally intensive tasks are written in c++ via the `R` package **Rcpp** (Eddelbuettel et al., 2011) and the c++ library **Armadillo** (Sanderson & Curtin, 2016). The Bayes factors are computed with the `R` package **BFpack** (Mulder et al., 2019). Furthermore, there are plotting functions for each method, control variables can be included in the model (e.g., `~ gender`), and there is support for missing values (see `bggm_missing`).

Comparison to Other Software

BGGM is the only `R` package to implement all of these algorithms and methods. The `mixed` data approach is also implemented in the package **sbgcop** (base `R`, Hoff, 2007). The `R` package **BDgraph** implements a Gaussian copula graphical model in c++ (Mohammadi & Wit, 2015), but not the binary or ordinal approaches. Furthermore, **BGGM** is the only package for confirmatory testing and comparing graphical models with the methods described in Williams et al. (2020).

Acknowledgements

DRW was supported by a National Science Foundation Graduate Research Fellowship under Grant No. 1650042 and JM was supported by a ERC Starting Grant (758791).

References

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, *88*(422), 669–679. doi:10.1080/01621459.1993.10476321

Chu, J.-h., Weiss, S. T., Carey, V. J., & Raby, B. A. (2009). A graphical model approach for inferring large-scale networks integrating gene expression and genetic polymorphism. *BMC systems biology*, *3*(1), 55. doi:10.1186/1752-0509-3-55

Cowles, M. K. (1996). Accelerating monte carlo markov chain convergence for cumulative-link generalized linear models. *Statistics and Computing*, *6*(2), 101–111. doi:10.1007/bf00162520

Eddelbuettel, D., François, R., Allaire, J., Ushey, K., Kou, Q., Russel, N., Chambers, J., et al. (2011). Rcpp: Seamless r and c++ integration. *Journal of Statistical Software*, *40*(8), 1–18.

Gelman, A., Meng, X.-L., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, *6*(4), 733–807.

Haslbeck, J. M., & Waldorp, L. J. (2018). How well do network models predict observations? On the importance of predictability in network models. *Behavior research methods*, *50*(2), 853–861. doi:10.3758/s13428-017-0910-x

Hoff, P. D. (2007). Extending the rank likelihood for semiparametric copula estimation. *The Annals of Applied Statistics*, *1*(1), 265–283. doi:10.1214/07-AOAS107

Jeffreys, H. (1961). *The theory of probability*. Oxford: Oxford University Press. ISBN: 0191589675

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Lawrence, E., Bingham, D., Liu, C., & Nair, V. N. (2008). Bayesian inference for multivariate ordinal data using parameter expansion. *Technometrics*, *50*(2), 182–191. doi:10.1198/004017008000000064

Millington, T., & Niranjan, M. (2020). Partial correlation financial networks. *Applied Network Science*, *5*(1), 11. doi:10.1007/s41109-020-0251-z

Mohammadi, R., & Wit, E. C. (2015). BDgraph: An r package for bayesian structure learning in graphical models. *Journal of Statistical Software*, *89*(3). doi:10.18637/jss.v089.i03

Mulder, J., Gu, X., Olsson-Collentine, A., Tomarken, A., Böing-Messing, F., Hoijtink, H., Meijerink, M., et al. (2019). BFpack: Flexible bayes factor testing of scientific theories in r. *arXiv preprint arXiv:1911.07728*.

Mulder, J., & Pericchi, L. (2018). The Matrix-F Prior for Estimating and Testing Covariance Matrices. *Bayesian Analysis*, (4), 1–22. doi:10.1214/17-BA1092

Rodriguez, J. E., Williams, D. R., Rast, P., & Mulder, J. (2020). On formalizing theoretical expectations: Bayesian testing of central structures in psychological networks. *PsyArXiv*. doi:10.31234/osf.io/zw7pf

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 1151–1172. doi:10.1214/aos/1176346785

Sanderson, C., & Curtin, R. (2016). Armadillo: A template-based c++ library for linear algebra. *Journal of Open Source Software*, *1*(2), 26. doi:10.21105/joss.00026

Talhouk, A., Doucet, A., & Murphy, K. (2012). Efficient bayesian inference for multivariate probit models with sparse inverse correlation matrices. *Journal of Computational and Graphical Statistics*, *21*(3), 739–757. doi:10.1080/10618600.2012.679239

Webb, E. L., & Forster, J. J. (2008). Bayesian model determination for multivariate ordinal and binary data. *Computational statistics & data analysis*, *52*(5), 2632–2649. doi:10.1016/j.csda.2007.09.008

Williams, D. R. (2018). Bayesian Estimation for Gaussian Graphical Models: Structure Learning, Predictability, and Network Comparisons. *arXiv*. doi:10.31234/OSF.IO/X8DPR

Williams, D. R., & Mulder, J. (2019). Bayesian Hypothesis Testing for Gaussian Graphical Models: Conditional Independence and Order Constraints. *PsyArXiv*. doi:10.31234/osf.io/ypxd8

Williams, D. R., Rast, P., Pericchi, L. R., & Mulder, J. (2020). Comparing gaussian graphical models with the posterior predictive distribution and bayesian model selection. *Psychological Methods*. doi:10.1037/met0000254

Zerenner, T., Friederichs, P., Lehnertz, K., & Hense, A. (2014). A gaussian graphical model approach to climate networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *24*(2), 023103. doi:10.1063/1.4870402