BGGM: Bayesian Gaussian Graphical Models in R

Donald R. Williams¹ and Joris Mulder²

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

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 uncertainty intervals 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 Wishart distribution.

- **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


