

Efficient Bayesian Structural Equation Modeling in Stan

Merkle, E. C. , Fitzsimmons, E. , Uanhoro, J. , & Goodrich, B. . (2020). Efficient bayesian structural equation modeling in stan. arXiv: [2008.07733](https://arxiv.org/abs/2008.07733)

Efficient Bayesian Structural Equation Modeling in Stan

Purpose

Backgroud

the original Stan approach

New Stan approach

Stan in blavaan

Priors for the covariance parameters

The proposed method

Comparison between JAGS and Stan

missing data

Latent variable scaling

Performance

Real data analysis

Simulation-based Calibration

Conclusion

My Thoughts

Purpose

describe and illustrate a new approach to Stan SEM estimation, which greatly improves the speed and efficiency of model estimation.

Compare: the original JAGS approach described in Merkle and Rosseel (2018), the original Stan approach that is being formally described in this paper for the first time, and the new Stan approach that is the primary focus of this paper.

Backgroud

Packages for Bayesian Modeling: **brms** (Bürkner 2017) for mixed and multivariate models, **ctsem** (Driver et al. 2017) for time series models, **edstan** (Furr 2017) for item response models, and **pcFactorStan** (Pritikin and Furr 2019) for pairwise comparison factor models.

Traditional Bayesian SEM: rely on sampling the η in tandem with other model parameters. This is advantageous because observed variables are often independent conditioned on the η , so that the conditional distribution of each observed variable is a univariate normal. However, the sampling of latent variables greatly increases the dimension of the parameters space, which can reduce sampling speed and efficiency.

the original JAGS approach described in Merkle and Rosseel (2018): [parameter expansion](#)

the original Stan approach

The phantom latent variable approach is not essential in Stan and, in testing, we found that the approach did not lead to gains in sampling speed or efficiency.

capitalizing on the structure of the SEM latent variable covariance matrix Ψ . This capitalization was inspired by related work on estimating multivariate autoregressive models in Stan (Joseph 2016).

We provide an overview of this approach here. For traditional SEMs, the distribution of latent variables can typically be expressed as (see Equation (2))

$$\eta \sim N((I - B)^{-1}\alpha, (I - B)^{-1}\Psi(I - B^{\top})^{-1}).$$

Evaluation of this multivariate normal log-likelihood is time-consuming in Stan because we need to compute the inverse and determinant of the covariance matrix. However, for many models, the structure of the covariance matrix leads to simplifications. For example, the matrix B is often triangular with zeros along its diagonal (leading to a so-called “recursive” model), and the matrix Ψ is often diagonal. When both of these properties are fulfilled, we can use standard matrix properties (e.g., Petersen and Pedersen 2012) to write the determinant as a product of scalar values:

$$\begin{aligned} \det(I - B)^{-1}\Psi(I - B^{\top})^{-1} &= (\det(I - B))^{-1}\det(\Psi)(\det(I - B))^{-1} \\ &= 1 \cdot \prod_{i=1}^m \psi_{ii} \cdot 1. \end{aligned}$$

Relatedly, the inverse of Σ is simplified as

$$((I - B)^{-1}\Psi(I - B^{\top})^{-1})^{-1} = (I - B^{\top})\Psi^{-1}(I - B),$$

which completely removes the need to compute matrix inversions when Ψ is diagonal. When either B is triangular or Ψ is diagonal (but not both), we can use a subset of the above

New Stan approach

Stan in blavaan

embed a stan program: we have written a single Stan program that can estimate the majority of multivariate normal SEMs that a user could specify.

limitations:

- there is some inflexibility in choice of prior distributions.
- equality constraints: allows for equality constraints within a class of parameters (e.g., loadings can be constrained equal to one another or intercepts can be constrained equal to one another), it does not allow for constraints across classes of parameters.

Priors for the covariance parameters

matrices without fixed values: inverse Wishart prior / LKJ prior (Lewandowski, Kurowicka, and Joe 2009).

other prior distributions proposed for covariance matrices (e.g., Chung, Gelman, Rabe-Hesketh, Liu, and Dorie 2015; Consonni and Veronese 2003; Mulder and Pericchi 2018; Spezia 2018).

The proposed method

In the Stan approaches, we consequently decompose the covariance matrices into standard deviations and correlations. For example, Θ is written as

$$\Theta = D_{\Theta} R_{\Theta} D_{\Theta}, \quad (7)$$

where D_{Θ} is a diagonal matrix of standard deviations and R_{Θ} is a correlation matrix. Prior distributions are then placed individually on the free standard deviation parameters and on the free correlation parameters within the two matrices. This approach is similar to that of Barnard, McCulloch, and Meng (2000), and Liu, Zhang, and Grimm (2016) provide a comparison of this approach to the use of inverse Wisharts in the context of growth curve models.

The use of an independent prior on each free parameter can sometimes lead to a non-positive definite covariance matrix during MCMC sampling. Stan is able to reject such a covariance matrix and continue sampling, whereas JAGS will terminate. This is why we developed

Prior (default):

	ν	λ	β	θ	ψ	ρ
Set 1 (default)	N(0,32)	N(0,10)	N(0,10)	Gamma(1,.5)	Gamma(1,.5)	Beta(1,1)
Set 2 (informative)	N(0,32)	N(1.25,.25)	N(1.5,.25)	Gamma(10,10)	Gamma(10,10)	Beta(5,5)

The strategy is worthwhile because it is relatively easy to specify informative prior distributions for individual standard deviation and correlation parameters in the model.

My example JAGS code (cannot work because the non-positive definite covariance matrix problem described above, just for illustration)

```
#####
# Specify the (prior) distribution for the latent variables
#####

for (i in 1:n){
  for (z in 1:Z){
    kappa[i,z] <- mu.f[z] + inprod(dat[i,(J+1):(J+P)], beta[1:P,z])
  }
  error[i,1:Z] ~ dmnorm(rep(0,Z), inv.psi[1:Z,1:Z]) # distribution for the latent variables
  ksi[i,1:Z] <- kappa[i,1:Z] + error[i,1:Z]
}

inv.psi[1:Z,1:Z] <- inverse(psi[1:Z,1:Z])
psi[1:Z,1:Z] <- psi.sd[1:Z,1:Z]%*%psi.cor[1:Z,1:Z]%*%psi.sd[1:Z,1:Z]

psi.sd[1,2] <- 0
psi.sd[2,1] <- 0
psi.sd[1,3] <- 0
psi.sd[3,1] <- 0
psi.sd[2,3] <- 0
psi.sd[3,2] <- 0

psi.cor[1,2] <- -1 + 2*psi.cor.new[1,2]
psi.cor[2,1] <- -1 + 2*psi.cor.new[2,1]
psi.cor[1,3] <- -1 + 2*psi.cor.new[1,3]
psi.cor[3,1] <- -1 + 2*psi.cor.new[3,1]
psi.cor[2,3] <- -1 + 2*psi.cor.new[2,3]
psi.cor[3,2] <- -1 + 2*psi.cor.new[3,2]

psi.cor.new[1,2] ~ dbeta(1,1)
psi.cor.new[2,1] ~ dbeta(1,1)
psi.cor.new[1,3] ~ dbeta(1,1)
psi.cor.new[3,1] ~ dbeta(1,1)
psi.cor.new[2,3] ~ dbeta(1,1)
psi.cor.new[3,2] ~ dbeta(1,1)

for (z in 1:Z){

  psi.sd[z,z] ~ dgamma(1, 0.5)
  psi.cor[z,z] <- 1

  mu.f[z] <- 0
}
```

Comparison between JAGS and Stan

missing data

- In JAGS, one can include NA values in the data, and JAGS will sample these missing values as if they were extra model parameters.
- Stan does not allow NA values in the data, so that one must handle the missing data manually.

blavaan: utilize a “full information” likelihood (e.g., Woithke 2000) in our Stan models, which is the same likelihood that is used to handle missing data in lavaan and other software that performs maximum likelihood SEM estimation. This requires some additional overhead in preparing the data to be sent to Stan

Latent variable scaling

(i) fixing each latent variable’s variance to 1, or (ii) fixing a single loading to 1 for each latent variable.

(ii) is most straightforward to implement in a Bayesian setting because (i) may cause non-convergence (Two sets of exact but opposite sign solutions)

when use (i):

- JAGS: a truncated normal prior (truncated from below at 0)
- Stan: the sign flipping is allowed to occur during MCMC sampling. Because after sampling, in the “generated quantities” block. In this block, one loading per latent variable is transformed to always be positive, and the signs of associated parameters (loadings, regressions, and covariance parameters) are flipped every time the sampled value of the focal loading is negative.

Performance

Real data analysis

conduct three real data analysis to demonstrate the efficiency of the proposed method

results: e.g., the marginal Stan (the proposed) method at 23.4 seconds per 100 iterations, the JAGS method at 28.4 seconds per 100 iterations, and the old Stan method at 577.27 seconds per 100 iterations.

```
R> fit1 <- bsem(model, data = PoliticalDemocracy, target = "jags")
R> fit2 <- bsem(model, data = PoliticalDemocracy, target = "stanclassic")
R> fit3 <- bcfa(model, data = PoliticalDemocracy, target = "stan")
```

Simulation-based Calibration

we use the simulation-based calibration method proposed by Talts, Betancourt, Simpson, Vehtari, and Gelman (2018) to study the calibration of blavaan’s new (marginal) Stan implementation. This involved repeatedly generating data from the model’s prior distribution, fitting the model to the generated data, and examining the ranks of the posterior MCMC samples relative to samples from the prior distribution. If the MCMC algorithm is calibrated, then these ranks should be approximately uniformly

distributed. Deviations from uniformity are then taken as miscalibration.

two sets of priors (listed above)

surprisingly, the distributions with default prior are far from uniform, with peaks generally occurring near zero and one.

The non-uniformity occurs because our model has a large number of parameters with independent prior distributions, with the parameters all contributing to the model-implied covariance matrix

the results of the informative priors are now much closer to uniform, because the information in the prior distributions now generally leads to positive definite model covariance matrices.

Conclusion

The results in this paper show that we can improve sampling efficiency by integrating the latent variables out of the model likelihood, which is the opposite of most popular approaches to Bayesian SEM estimation (where the popular approaches are largely based on results summarized by, e.g., Lee 2007; Song and Lee 2012).

We can expect the marginal sampling efficiency to be even more advantageous as sample sizes increase, because the sample size has no impact on the dimension of the parameter space here. In contrast, the dimension of the parameter space increases with sample size under conditional approaches, where latent variables count as parameters.

These problems include the fact that the marginal likelihood does not have a closed form when we have non-normal observed variables (e.g., ordinal variables) or when we have latent variable interactions. We think that some progress can be made here by employing other Bayesian methods, including data augmentation (e.g., Chib and Greenberg 1998) in the ordinal case.

Further, even if those methods did exhibit reasonable efficiency, the marginal likelihood is generally necessary for obtaining suitable information criteria such as DIC (Spiegelhalter, Best, Carlin, and Linde 2002) or WAIC (Watanabe 2010). . Merkle, Furr, and Rabe-Hesketh (2019) discuss why the marginal likelihood is preferable here, and Zhang, Tao, Wang, and Shi (2019) discuss related applications of DIC to multilevel item response models.

My Thoughts

The limitation about dealing with missing values using Stan may greatly affect its application.....

Compared with blavaan, Mplus is so efficient...

■

[Return](#)