Package ‘MBSP’

January 20, 2025

Type Package

Title Multivariate Bayesian Model with Shrinkage Priors

Version 4.0

Date 2023-05-17

Author Ray Bai, Malay Ghosh

Maintainer Ray Bai <raybaistat@gmail.com>

Description Gibbs sampler for fitting multivariate Bayesian linear regression with shrinkage priors (MBSP), using the three parameter beta normal family. The method is described in Bai and Ghosh (2018) <doi:10.1016/j.jmva.2018.04.010>.

Depends R (>= 3.6.0)

Imports stats, MCMCpack, GIGrvg, utils, mvtnorm

NeedsCompilation yes

Repository CRAN

Date/Publication 2023-05-17 21:00:03 UTC

Contents

`matrix_normal` *Matrix-Normal Distribution*

Description

This function provides a way to draw a sample from the matrix-normal distribution, given the mean matrix, the covariance structure of the rows, and the covariance structure of the columns.

Usage

```
matrix_normal(M, U, V)
```

Arguments

M	mean $a \times b$ matrix
U	$a \times a$ covariance matrix (covariance of rows).
V	$b \times b$ covariance matrix (covariance of columns).

Details

This function provides a way to draw a random $a \times b$ matrix from the matrix-normal distribution,

$$MN(M, U, V),$$

where M is the $a \times b$ mean matrix, U is an $a \times a$ covariance matrix, and V is a $b \times b$ covariance matrix.

Value

A randomly drawn $a \times b$ matrix from $MN(M, U, V)$.

Author(s)

Ray Bai and Malay Ghosh

Examples

```
# Draw a random 50x20 matrix from MN(O,U,V),
# where:
#    O = zero matrix of dimension 50x20
#    U has AR(1) structure,
#    V has sigma^2*I structure

# Specify Mean.mat
p <- 50
q <- 20
Mean_mat <- matrix(0, nrow=p, ncol=q)

# Construct U
rho <- 0.5
times <- 1:p
H <- abs(outer(times, times, "-"))
U <- rho^H

# Construct V
sigma_sq <- 2
V <- sigma_sq*diag(q)

# Draw from MN(Mean_mat, U, V)
mn_draw <- matrix_normal(Mean_mat, U, V)
```

MBSP *MBSP Model with Three Parameter Beta Normal (TPBN) Family*

Description

This function provides a fully Bayesian approach for obtaining a (nearly) sparse estimate of the $p \times q$ regression coefficients matrix B in the multivariate linear regression model,

$$Y = XB + E,$$

using the three parameter beta normal (TPBN) family. Here Y is the $n \times q$ matrix with n samples of q response variables, X is the $n \times p$ design matrix with n samples of p covariates, and E is the $n \times q$ noise matrix with independent rows. The complete model is described in Bai and Ghosh (2018).

If there are r confounding variables which *must* remain in the model and should *not* be regularized, then these can be included in the model by putting them in a separate $n \times r$ confounding matrix Z. Then the model that is fit is

$$Y = XB + ZC + E,$$

where C is the $r \times q$ regression coefficients matrix corresponding to the confounders. In this case, we put a flat prior on C. By default, confounders are not included.

If the user desires, two information criteria can be computed: the Deviance Information Criterion (DIC) of Spiegelhalter et al. (2002) and the widely applicable information criterion (WAIC) of Watanabe (2010).

Usage

```
MBSP(Y, X, confounders=NULL, u=0.5, a=0.5, tau=NA,
     max_steps=6000, burnin=1000, save_samples=TRUE,
     model_criteria=FALSE)
```

Arguments

`Y`	Response matrix of n samples and q response variables.
`X`	Design matrix of n samples and p covariates. The MBSP model regularizes the regression coefficients B corresponding to X.
`confounders`	Optional design matrix Z of n samples of r confounding variables. By default, confounders are not included in the model (`confounders=NULL`). However, if there are some confounders that *must* remain in the model and should *not* be regularized, then the user can include them here.
`u`	The first parameter in the TPBN family. Defaults to $u = 0.5$ for the horseshoe prior.
`a`	The second parameter in the TPBN family. Defaults to $a = 0.5$ for the horseshoe prior.

tau	The global parameter. If the user does not specify this (tau=NA), the Gibbs sampler will use $\tau = 1/(p * n * log(n))$. The user may also specify any value for τ strictly greater than 0; otherwise it defaults to $1/(p * n * log(n))$.
max_steps	The total number of iterations to run in the Gibbs sampler. Defaults to 6000.
burnin	The number of burn-in iterations for the Gibbs sampler. Defaults to 1000.
save_samples	A Boolean variable for whether to save all of the posterior samples of the regression coefficients matrix B and the covariance matrix Sigma. Defaults to "TRUE".
model_criteria	A Boolean variable for whether to compute the following information criteria: DIC (Deviance Information Criterion) and WAIC (widely applicable information criterion). Can be used to compare models with (for example) different choices of u, a, or tau. Defauls to "FALSE".

Details

The function performs (nearly) sparse estimation of the regression coefficients matrix B and variable selection from the p covariates. The lower and upper endpoints of the 95 percent posterior credible intervals for each of the pq elements of B are also returned so that the user may assess uncertainty quantification.

In the three parameter beta normal (TPBN) family, $(u, a) = (0.5, 0.5)$ corresponds to the horseshoe prior, $(u, a) = (1, 0.5)$ corresponds to the Strawderman-Berger prior, and $(u, a) = (1, a), a > 0$ corresponds to the normal-exponential-gamma (NEG) prior. This function uses the horseshoe prior as the default shrinkage prior.

The user also has the option of including an $n \times r$ matrix with r confounding variables. These confounders are variables which are included in the model but should *not* be regularized.

Finally, if the user specifies model_criteria=TRUE, then the MBSP function will compute two model selection criteria: the Deviance Information Criterion (DIC) of Spiegelhalter et al. (2002) and the widely applicable information criterion (WAIC) of Watanabe (2010). This permits model comparisons between (for example) different choices of u, a, and tau. The default horseshoe prior and choice of tau performs well, but the user may wish to experiment with u, a, and tau. In general, models with *lower* DIC or WAIC are preferred.

Value

The function returns a list containing the following components:

B_est	The point estimate of the $p \times q$ matrix B (taken as the componentwise posterior median for all pq entries).
B_CI_lower	The 2.5th percentile of the posterior density (or the lower endpoint of the 95 percent credible interval) for all pq entries of B.
B_CI_upper	The 97.5th percentile of the posterior density (or the upper endpoint of the 95 percent credible interval) for all pq entries of B.
active_predictors	The row indices of the active (nonzero) covariates chosen by our model from the p total predictors.
B_samples	All max_steps-burnin samples of B.

C_est	The point estimate of the $r \times q$ matrix C corresponding to the confounders (taken as the componentwise posterior median for all rq entries). This matrix is not returned if there are no confounders (i.e. confounders=NULL).
C_CI_lower	The 2.5th percentile of the posterior density (or the lower endpoint of the 95 percent credible interval) for all rq entries of C. This is not returned if there are no confounders (i.e. confounders=NULL).
C_CI_upper	The 97.5th percentile of the posterior density (or the upper endpoint of the 95 percent credible interval) for all rq entries of C. This is not returned if there are no confounders (i.e. confounders=NULL)
C_samples	All max_steps-burnin samples of C. This is not returned if there are no confounders (i.e. confounders=NULL)
Sigma_est	The point estimate of the $q \times q$ covariance matrix Σ (taken as the componentwise posterior median for all q^2 entries).
Sigma_CI_lower	The 2.5th percentile of the posterior density (or the lower endpoint of the 95 percent credible interval) for all q^2 entries of Σ.
Sigma_CI_upper	The 97.5th percentile of the posterior density (or the upper endpoint of the 95 percent credible interval) for all q^2 entries of Σ.
Sigma_samples	All max_steps-burnin samples of C.
DIC	The Deviance Information Criterion (DIC), which can be used for model comparison. Models with smaller DIC are preferred. This only returns a numeric value if model_criteria=TRUE is specified.
WAIC	The widely applicable information criterion (WAIC), which can be used for model comparison. Models with smaller WAIC are preferred. This only returns a numeric value if model_criteria=TRUE is specified. The WAIC tends to be more stable than DIC.

Author(s)

Ray Bai and Malay Ghosh

References

Armagan, A., Clyde, M., and Dunson, D.B. (2011) Generalized beta mixtures of Gaussians. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger (Eds.) *Advances in Neural Information Processing Systems 24*, 523-531.

Bai, R. and Ghosh, M. (2018). High-dimensional multivariate posterior consistency under global-local shrinkage priors. *Journal of Multivariate Analysis*, **167**: 157-170.

Berger, J. (1980). A robust generalized Bayes estimator and confidence region for a multivariate normal mean. *Annals of Statistics*, **8**(4): 716-761.

Carvalho, C.M., Polson, N.G., and Scott., J.G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, **97**(2): 465-480.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4): 583-639.

Strawderman, W.E. (1971). Proper Bayes Minimax Estimators of the Multivariate Normal Mean. *Annals of Mathematical Statistics*, **42**(1): 385-388.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, **11**: 3571-3594.

Examples

```
###################################
# Set n, p, q, and sparsity level #
###################################

n <- 100
p <- 40
q <- 3 # number of response variables is 3
p_act <- 5 # number of active (nonzero) predictors is 5

#############################
# Generate design matrix X. #
#############################
set.seed(1234)
times <- 1:p
rho <- 0.5
H <- abs(outer(times, times, "-"))
V <- rho^H
mu <- rep(0, p)
# Rows of X are simulated from MVN(0,V)
X <- mvtnorm::rmvnorm(n, mu, V)
# Center X
X <- scale(X, center=TRUE, scale=FALSE)

############################################
# Generate true coefficient matrix B_true. #
############################################
# Entries in nonzero rows are drawn from Unif[(-5,-0.5)U(0.5,5)]
B_act <- runif(p_act*q,-5,4)
disjoint <- function(x){
if(x <= -0.5)
return(x)
else
return(x+1)
}
B_act <- matrix(sapply(B_act, disjoint),p_act,q)
# Set rest of the rows equal to 0
B_true <- rbind(B_act,matrix(0,p-p_act,q))
B_true <- B_true[sample(1:p),] # permute the rows

#########################################
# Generate true error covariance Sigma. #
#########################################
sigma_sq=2
times <- 1:q
```

```
H <- abs(outer(times, times, "-"))
Sigma <- sigma_sq * rho^H

############################
# Generate noise matrix E. #
############################
mu <- rep(0,q)
E <- mvtnorm::rmvnorm(n, mu, Sigma)

##############################
# Generate response matrix Y #
##############################
Y <- crossprod(t(X),B_true) + E

# Note that there are no confounding variables in this synthetic example

#########################################
# Fit the MBSP model on synthetic data. #
#########################################

# Should use default of max_steps=6000, burnin=1000 in practice.
mbsp_model = MBSP(Y=Y, X=X, max_steps=1000, burnin=500, model_criteria=FALSE)

# Recommended to use the default, i.e. can simply use: mbsp_model = MBSP(Y, X)
# If you want to return the DIC and WAIC, have to set model_criteria=TRUE.


# indices of the true nonzero rows
true_active_predictors <- which(rowSums(B_true)!=0)
true_active_predictors

# variables selected by the MBSP model
mbsp_model$active_predictors

# true regression coeficients in the true nonzero rows
B_true[true_active_predictors, ]

# the MBSP model's estimates of the nonzero rows
mbsp_model$B_est[true_active_predictors, ]
```

Index