

# Package ‘GCPBayes’

October 21, 2021

**Type** Package

**Title** Bayesian Meta-Analysis of Pleiotropic Effects Using Group Structure

**Version** 3.1.0

**Author** Taban Baghfalaki

**Maintainer** Taban Baghfalaki <t.baghfalaki@gmail.com>

## Description

Run a Gibbs sampler for a multivariate Bayesian sparse group selection model with Dirac, continuous and hierarchical spike prior for detecting pleiotropy on the traits. This package is designed for summary statistics containing estimated regression coefficients and its estimated covariance matrix. The methodology is available from: Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Liqueet, B. (2021) <[doi:10.1002/sim.8855](https://doi.org/10.1002/sim.8855)>.

**License** GPL (>= 2.0)

**Encoding** UTF-8

**Imports** MASS, arm, mvtnorm, invgamma, gdata, truncnorm, postpack, wiqid, usethis, Rcpp

**LazyData** true

**RoxygenNote** 7.1.1

**Depends** R (>= 3.5.0)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-10-21 17:30:02 UTC

## R topics documented:

CS . . . . .	2
DNAJC1 . . . . .	8
DS . . . . .	9
e2_Monte_Carlo_EM . . . . .	13
GCPBayes . . . . .	14
HS . . . . .	15

MCMCplot . . . . .	20
PARP2 . . . . .	22
PARP2_summary . . . . .	23
Simulated_individual . . . . .	24
Simulated_individual_survival . . . . .	25
Simulated_summary . . . . .	25

<b>Index</b>	<b>27</b>
--------------	-----------

---

CS	<i>Continuous Spike</i>
----	-------------------------

---

### Description

Run a Gibbs sampler for a multivariate Bayesian sparse group selection model with Continuous spike prior for detecting pleiotropic effects on the traits. This function is designed for summary statistics containing estimated regression coefficients and their estimated covariance matrices.

### Usage

```
CS(
  Betah,
  Sigmah,
  kappa0,
  tau20,
  zeta0,
  m,
  K,
  niter = 1000,
  burnin = 500,
  nthin = 2,
  nchains = 2,
  a1 = a1,
  a2 = a2,
  c1 = c1,
  c2 = c2,
  sigma2 = 10^-3,
  snpnames = snpnames,
  genename = genename
)
```

### Arguments

Betah	A list containing m-dimensional vectors of the regression coefficients for K studies.
Sigmah	A list containing the positive definite covariance matrices (m*m-dimensional) which is the estimated covariance matrices of K studies.
kappa0	Initial value for kappa (its dimension is equal to nchains).

tau20	Initial value for tau2 (its dimension is equal to nchains).
zeta0	Initial value for zeta.
m	Number of variables in the group.
K	Number of traits.
niter	Number of iterations for the Gibbs sampler.
burnin	Number of burn-in iterations.
nthin	The lag of the iterations used for the posterior analysis is defined (or thinning rate).
nchains	Number of Markov chains, when nchains>1, the function calculates the Gelman-Rubin convergence statistic, as modified by Brooks and Gelman (1998).
a1, a2	Hyperparameters of kappa. Default is a1=0.1 and a2=0.1.
c1, c2	Hyperparameters of tau2. Default is c1=0.1 and c2=0.1.
sigma2	Variance of spike (multivariate normal distribution with a diagonal covariance matrix with small variance) representing the null effect distribution. Default is 10 <sup>-3</sup> .
snpnames	Names of variables for the group.
genename	Name of group.

## Details

Run a Gibbs sampler using a Continuous spike

## Value

- mcmcchain: The list of simulation output for all parameters.
- Summary: Summary statistics for regression coefficients in each study.
- Criteria: genename, snpnames, PPA, log10BF, IBFDR, theta.
- Indicator: A table containing m rows of binary indicators for each study, the number of studies with nonzero signal and having pleiotropic effect by credible interval (CI). The first K columns show nonzero signals, K+1 th column includes the number of studies with nonzero signal and the last column shows an indicator for having pleiotropic effect of each SNP.

## Author(s)

Taban Baghfalaki.

## References

Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Liquet, B. (2021). Bayesian meta analysis models for cross cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, 40(6), 1498-1518.

## Examples

```
#####Gene DNAJC1 #####
data(DNAJC1)
Breast <- DNAJC1$Breast
Thyroid <- DNAJC1$Thyroid
genename <- "DNAJC1"
snpnames <- Breast$snp
Betah <- list(Breast$beta, Thyroid$beta)
Sigmah <- list(diag(Breast$se^2), diag(Thyroid$se^2))
K <- 2
m <- 14

pvalue <- matrix(0, K, m)
for (k in 1:K) {
  pvalue[k, ] <- 2 * pnorm(-abs(Betah[[k]] / sqrt(diag(Sigmah[[k]]))))
}

zinit <- rep(0, K)
for (j in 1:K) {
  index <- 1:m
  PVALUE <- p.adjust(pvalue[j, ])
  SIGNALS <- index[PVALUE < 0.05]
  modelf1 <- rep(0, m)
  modelf1[SIGNALS] <- 1
  if (max(modelf1) == 1) (zinit[j] <- 1)
}

RES <- CS(Betah, Sigmah,
  kappa0 = 0.5, tau20 = 1, zeta0 = zinit,
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1, a1 = 0.1, a2 = 0.1,
  c1 = 0.1, c2 = 0.1, sigma2 = 10^-3, snpnames = snpnames, genename = genename
)
## Not run:
print(RES)
RES1 <- CS(Betah, Sigmah,
  kappa0 = c(0.2, 0.5), tau20 = c(1, 2), zeta0 = zinit,
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
  a1 = 0.1, a2 = 0.1, c1 = 0.1, c2 = 0.1, sigma2 = 10^-3, snpnames, genename)
print(RES1)
#####Simulated summary level data with K=5 #####
data(Simulated_summary)
genename <- Simulated_summary$genename
snpnames <- Simulated_summary$snpnames
Betah <- Simulated_summary$simBeta
Sigmah <- Simulated_summary$simSIGMA
K <- 5
m <- 10

pvalue <- matrix(0, K, m)
for (k in 1:K) {
  pvalue[k, ] <- 2 * pnorm(-abs(Betah[[k]] / sqrt(diag(Sigmah[[k]]))))
}
```

```

zinit <- rep(0, K)
for (j in 1:K) {
  index <- 1:m
  PVALUE <- p.adjust(pvalue[j, ])
  SIGNALS <- index[PVALUE < 0.05]
  modelf1 <- rep(0, m)
  modelf1[SIGNALS] <- 1
  if (max(modelf1) == 1) (zinit[j] <- 1)
}

RES <- CS(Betah, Sigmah,
  kappa0 = 0.5, tau20 = 1, zeta0 = zinit,
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1, a1 = 0.1, a2 = 0.1,
  c1 = 0.1, c2 = 0.1, sigma2 = 10^-3, snpnames = snpnames, genename = genename)
print(RES)
RES1 <- CS(Betah, Sigmah,
  kappa0 = c(0.2, 0.5), tau20 = c(1, 2), zeta0 = zinit,
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
  a1 = 0.1, a2 = 0.1, c1 = 0.1, c2 = 0.1, sigma2 = 10^-3, snpnames, genename)
print(RES1)
#####Gene PARP2 #####
library(BhGLM)
data(PARP2)
Breast <- PARP2$Breast
Thyroid <- PARP2$Thyroid
genename <- "PARP2"
snpnames <- c("rs3093872", "rs3093921", "rs1713411", "rs3093926", "rs3093930", "rs878156")

Fit1 <- BhGLM::bglm(y1 ~ ., family = binomial(link = "logit"), data = Breast)
Betah1 <- Fit1$coefficients[-1]
Sigmah1 <- cov(coef(arm::sim(Fit1)))[-1, -1]

Fit2 <- BhGLM::bglm(y2 ~ ., family = binomial(link = "logit"), data = Thyroid)
Betah2 <- Fit2$coefficients[-1]
Sigmah2 <- cov(coef(arm::sim(Fit2)))[-1, -1]

Betah <- list(Betah1, Betah2)
Sigmah <- list(Sigmah1, Sigmah2)
K <- 2
m <- 6

pvalue <- matrix(0, K, m)
for (k in 1:K) {
  pvalue[k, ] <- 2 * pnorm(-abs(Betah[[k]] / sqrt(diag(Sigmah[[k]]))))
}

zinit <- rep(0, K)
for (j in 1:K) {
  index <- 1:m
  PVALUE <- p.adjust(pvalue[j, ])
  SIGNALS <- index[PVALUE < 0.05]

```

```

modelf1 <- rep(0, m)
modelf1[SIGNALS] <- 1
if (max(modelf1) == 1) (zinit[j] <- 1)
}

RES <- CS(Betah, Sigmah,
kappa0 = 0.5, tau20 = 1, zeta0 = zinit,
m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1, a1 = 0.1, a2 = 0.1,
c1 = 0.1, c2 = 0.1, sigma2 = 10^-3, snpnames = snpnames, genename = genename)
print(RES)
RES1 <- CS(Betah, Sigmah,
kappa0 = c(0.2, 0.5), tau20 = c(1, 2), zeta0 = zinit,
m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
a1 = 0.1, a2 = 0.1, c1 = 0.1, c2 = 0.1, sigma2 = 10^-3, snpnames, genename)
print(RES1)
#####Simulated individual level data with K=3 and continuous phynotype #####
library(BhGLM)
data(Simulated_individual)
Study1 <- Simulated_individual$Study1
Study2 <- Simulated_individual$Study2
Study3 <- Simulated_individual$Study3
K <- 3
m <- 30
genename <- "Simulated"
snpnames <- sprintf("SNP%s",seq(1:m))

Fit1 <- BhGLM::bglm(Y1~ ., family=gaussian,data=data.frame(Study1))
Betah1 <- Fit1$coefficients[-1]
Sigmah1 <- cov(coef(arm::sim(Fit1)))[-1,-1]

Fit2 <- BhGLM::bglm(Y2~ ., family=gaussian,data=data.frame(Study2))
Betah2 <- Fit2$coefficients[-1]
Sigmah2 <- cov(coef(arm::sim(Fit2)))[-1,-1]

Fit3 <- BhGLM::bglm(Y3~ ., family=gaussian,data=data.frame(Study3))
Betah3 <- Fit3$coefficients[-1]
Sigmah3 <- cov(coef(arm::sim(Fit3)))[-1,-1]

Betah <- list(Betah1,Betah2,Betah3)
Sigmah <- list(Sigmah1,Sigmah2,Sigmah3)

pvalue <- matrix(0,K,m)
for(k in 1:K){
  pvalue[k,] <- 2*pnorm(-abs(Betah[[k]]/sqrt(diag(Sigmah[[k]]))))
}

zinit=rep(0,K)
for(j in 1:K){
  index <- 1:m
  PVALUE <- p.adjust(pvalue[j,])
  SIGNALS <- index[PVALUE<0.05]
}

```

```

  modelf1 <- rep(0,m)
  modelf1[SIGNALS] <- 1
  if(max(modelf1)==1)(zinit[j] <- 1)
}

RES <- CS(Betah, Sigmah, kappa0=0.5, tau20=1, zeta0=zinit,
         m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=1, a1=0.1, a2=0.1,
         c1=0.1, c2=0.1, sigma2=10^-3, snpnames=snpnames, genename=genename)
print(RES)
RES1 <- CS(Betah, Sigmah, kappa0=c(0.2,0.5), tau20=c(1,2), zeta0=zinit,
         m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
         a1=0.1, a2=0.1, c1=0.1, c2=0.1, sigma2=10^-3, snpnames, genename)
print(RES1)

#####Simulated individual level data with K=2 and gene expression data #####
library(BhGLM)
data(Simulated_individual_survival)
Study1 <- Simulated_individual_survival$Study1
Study2 <- Simulated_individual_survival$Study2
K <- 2
m <- 10
genename <- "Simulated"
snpnames <- sprintf("G%s",seq(1:m))

Fit1=BhGLM::bcoxph(Study1$T ~ Study1$X)
Betah1 <- Fit1$coefficients
Sigmah1 <- Fit1$var

Fit2=BhGLM::bcoxph(Study2$T ~ Study2$X)
Betah2 <- Fit2$coefficients
Sigmah2 <- Fit2$var

Betah <- list(Betah1,Betah2)
Sigmah <- list(Sigmah1,Sigmah2)

pvalue <- matrix(0,K,m)
for(k in 1:K){
  pvalue[k,] <- 2*pnorm(-abs(Betah[[k]]/sqrt(diag(Sigmah[[k]]))))
}

zinit=rep(0,K)
for(j in 1:K){
  index <- 1:m
  PVALUE <- p.adjust(pvalue[j,])
  SIGNALS <- index[PVALUE<0.05]
  modelf1 <- rep(0,m)
  modelf1[SIGNALS] <- 1
  if(max(modelf1)==1)(zinit[j] <- 1)
}

```

```
RES1 <- CS(Betah, Sigmah, kappa0=c(0.2,0.5), tau20=c(1,2), zeta0=zinit,  
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,  
          a1=0.1, a2=0.1, c1=0.1, c2=0.1, sigma2=10^-3, snpnames, genename)  
print(RES1)  
  
## End(Not run)
```

---

DNAJC1

*Gene DNAJC1 from BCAC and Epithyr studies*

---

## Description

The summary statistics data for DNAJC1 protein coding gene including beta and standard error for pleiotropy investigation of breast and thyroid cancers. The summary statistics of breast and thyroid cancer are extracted from BCAC and Epithyr (Baghgalaki et al., 2021b) studies, respectively.

## Usage

DNAJC1

## Format

A list which contains two matrices for the summary statistics of each study.

**Breast** Summary statistics of breast cancer including the name of SNPs, beta and se

**Thyroid** Summary statistics of thyroid cancer including the name of SNPs, beta and se

## References

1. Baghgalaki, T., Sugier, Y. Asgari, P. E., Truong, & Liquet, B. (2021). GCPBayes: An R Package for Studying Cross-Phenotype Genetic Associations with Group-Level Bayesian Meta-Analysis. *Submitted*.

## See Also

[GCPBayes](#)



DS

*Dirac Spike***Description**

Run a Gibbs sampler for a multivariate Bayesian sparse group selection model with Dirac spike prior for detecting pleiotropic effects on the traits. This function is designed for summary statistics containing estimated regression coefficients and their estimated covariance matrices.

**Usage**

```
DS(
  Betah,
  Sigmah,
  kappa0,
  sigma20,
  m,
  K,
  niter = 1000,
  burnin = 500,
  nthin = 2,
  nchains = 2,
  a1 = 0.1,
  a2 = 0.1,
  d1 = 0.1,
  d2 = 0.1,
  snpnames,
  genename
)
```

**Arguments**

Betah	A list containing m-dimensional vectors of the regression coefficients for K studies.
Sigmah	A list containing the positive definite covariance matrices (m*m-dimensional) which is the estimated covariance matrices of K studies.
kappa0	Initial value for kappa (its dimension is equal to nchains).
sigma20	Initial value for sigma2 (its dimension is equal to nchains).
m	Number of variables in the group.
K	Number of traits.
niter	Number of iterations for the Gibbs sampler.
burnin	Number of burn-in iterations.
nthin	The lag of the iterations used for the posterior analysis is defined (or thinning rate).

nchains	Number of Markov chains, when nchains>1, the function calculates the Gelman-Rubin convergence statistic, as modified by Brooks and Gelman (1998).
a1, a2	Hyperparameters of kappa. Default is a1=0.1 and a2=0.1.
d1, d2	Hyperparameters of sigma2. Default is d1=0.1 and d2=0.1.
snpnames	Names of variables for the group.
genename	Name of group.

### Details

Run a Gibbs sampler using a Dirac spike

### Value

- mcmcchain: The list of simulation output for all parameters.
- Summary: Summary statistics for regression coefficients in each study.
- Criteria: genename, snpnames, PPA, log10BF, IBFDR, theta.
- Indicator: A table containing m rows of binary indicators for each study, the number of studies with nonzero signal and having pleiotropic effect by credible interval (CI). The first K columns show nonzero signals, K+1 th column includes the number of studies with nonzero signal and the last column shows an indicator for having pleiotropic effect of each SNP.

### Author(s)

Taban Baghfalaki.

### References

1. Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Liquet, B. (2021). Bayesian meta analysis models for cross cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, **40**(6), 1498-1518.

### Examples

```
#####Gene DNAJC1 #####
data(DNAJC1)
Breast <- DNAJC1$Breast
Thyroid <- DNAJC1$Thyroid
genename <- "DNAJC1"
snpnames <- Breast$snp
Betah <- list(Breast$beta, Thyroid$beta)
Sigmah <- list(diag(Breast$se^2), diag(Thyroid$se^2))
K <- 2
m <- 14

RES <- DS(Betah, Sigmah,
  kappa0 = 0.5, sigma20 = 1,
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1,
  a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, snpnames, genename
)
```

```

## Not run:
print(RES)
RES1 <- DS(Betah, Sigmah,
           kappa0 = c(0.2, 0.5), sigma20 = c(1, 2),
           m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
           a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, snpnames, genename)
print(RES1)
#####Simulated summary level data with K=5 #####
data(Simulated_summary)
genename <- Simulated_summary$genename
snpnames <- Simulated_summary$snpnames
Betah <- Simulated_summary$simBeta
Sigmah <- Simulated_summary$simSIGMA
K <- 5
m <- 10

RES <- DS(Betah, Sigmah,
          kappa0 = 0.5, sigma20 = 1,
          m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1,
          a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, snpnames, genename)
print(RES)

RES1 <- DS(Betah, Sigmah,
           kappa0 = c(0.2, 0.5), sigma20 = c(1, 2),
           m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
           a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, snpnames, genename)
print(RES1)
#####Gene PARP2 #####
library(BhGLM)
data(PARP2)
Breast <- PARP2$Breast
Thyroid <- PARP2$Thyroid
genename <- "PARP2"
snpnames <- c("rs3093872", "rs3093921", "rs1713411", "rs3093926", "rs3093930", "rs878156")

Fit1 <- BhGLM::bglm(y1~ ., family=binomial(link="logit"),data=Breast)
Betah1 <- Fit1$coefficients[-1]
Sigmah1 <- cov(coef(arm::sim(Fit1)))[-1,-1]

Fit2 <- BhGLM::bglm(y2~ ., family=binomial(link="logit"),data=Thyroid)
Betah2 <- Fit2$coefficients[-1]
Sigmah2 <- cov(coef(arm::sim(Fit2)))[-1,-1]

Betah <- list(Betah1,Betah2)
Sigmah <- list(Sigmah1,Sigmah2)
K <- 2
m <- 6

RES <- DS(Betah, Sigmah, kappa0=0.5, sigma20=1,
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=1,
          a1=0.1, a2=0.1, d1=0.1, d2=0.1, snpnames, genename)
print(RES)

```

```

RES1 <- DS(Betah, Sigmah, kappa0=c(0.2,0.5), sigma20=c(1,2),
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
          a1=0.1, a2=0.1, d1=0.1, d2=0.1, snpnames, genename)
print(RES1)
#####Simulated individual level data with K=3 and continuous phynotype #####
library(BhGLM)
data(Simulated_individual)
Study1 <- Simulated_individual$Study1
Study2 <- Simulated_individual$Study2
Study3 <- Simulated_individual$Study3
K <- 3
m <- 30
genename <- "Simulated"
snpnames <- sprintf("SNP%s",seq(1:m))

Fit1 <- BhGLM::bglm(Y1~ ., family=gaussian,data=data.frame(Study1))
Betah1 <- Fit1$coefficients[-1]
Sigmah1 <- cov(coef(arm::sim(Fit1)))[-1,-1]

Fit2 <- BhGLM::bglm(Y2~ ., family=gaussian,data=data.frame(Study2))
Betah2 <- Fit2$coefficients[-1]
Sigmah2 <- cov(coef(arm::sim(Fit2)))[-1,-1]

Fit3 <- BhGLM::bglm(Y3~ ., family=gaussian,data=data.frame(Study3))
Betah3 <- Fit3$coefficients[-1]
Sigmah3 <- cov(coef(arm::sim(Fit3)))[-1,-1]

Betah <- list(Betah1,Betah2,Betah3)
Sigmah <- list(Sigmah1,Sigmah2,Sigmah3)

RES <- DS(Betah, Sigmah, kappa0=0.5, sigma20=1,
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=1,
          a1=0.1, a2=0.1, d1=0.1, d2=0.1, snpnames, genename)
print(RES)

RES1 <- DS(Betah, Sigmah, kappa0=c(0.2,0.5), sigma20=c(1,2),
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
          a1=0.1, a2=0.1, d1=0.1, d2=0.1, snpnames, genename)
print(RES1)

#####Simulated individual level data with K=2 and gene expression data #####
library(BhGLM)
data(Simulated_individual_survival)
Study1 <- Simulated_individual_survival$Study1
Study2 <- Simulated_individual_survival$Study2
K <- 2
m <- 10
genename <- "Simulated"
snpnames <- sprintf("G%s",seq(1:m))

```

```

Fit1=BhGLM::bcoxph(Study1$T ~ Study1$X)
Betah1 <- Fit1$coefficients
Sigmah1 <- Fit1$var

Fit2=BhGLM::bcoxph(Study2$T ~ Study2$X)
Betah2 <- Fit2$coefficients
Sigmah2 <- Fit2$var

Betah <- list(Betah1,Betah2)
Sigmah <- list(Sigmah1,Sigmah2)

RES1 <- DS(Betah, Sigmah, kappa0=c(0.2,0.5), sigma20=c(1,2),
           m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
           a1=0.1, a2=0.1, d1=0.1, d2=0.1, snpnames, genename)
print(RES1)

## End(Not run)

```

---

e2\_Monte\_Carlo\_EM

*Internal: e2\_Monte\_Carlo\_EM*


---

## Description

Internal: e2\_Monte\_Carlo\_EM

## Usage

```

e2_Monte_Carlo_EM(
  Betah,
  Sigmah,
  kappa0 = kappa0,
  kappastar0 = kappastar0,
  sigma20 = sigma20,
  s20 = s20,
  m,
  K,
  a1 = a1,
  a2 = a2,
  d1 = d1,
  d2 = d2,
  c1 = c1,
  c2 = c2,
  e2 = e2,
  snpnames,
  genename
)

```

**Arguments**

Beta $\theta$	A matrix of dimension $K \times m$ represents the regression coefficients. Each row of this matrix includes the regression coefficients for each trait.
Sigma $\theta$	A symmetric block-diagonal matrix of dimension $K \times m$ is used. Each block of this matrix shows a positive definite covariance matrix which is an estimated covariance matrix of each trait.
kappa $\theta$	Initial value for kappa.
kappastar $\theta$	Initial value for kappastar.
sigma2 $\theta$	Initial value for sigma2.
s2 $\theta$	Initial value for s2.
m	Number of variables in the group.
K	Number of traits.
a1, a2	Hyperparameters of kappa. Default is a1=0.1 and a2=0.1.
d1, d2	Hyperparameters of sigma2. Default is d1=0.1 and d2=0.1.
c1, c2	Hyperparameters of kappastar. Default is c1=0.1 and c2=0.1.
e2	Initial value for doing Monte Carlo EM algorithm to estimate hyperparameter of s2.
snpname	Names of variables for the group.
genename	Name of group.

GCPBayes

*GCPBayes Package***Description**

Run a Gibbs sampler for a multivariate Bayesian sparse group selection model with Dirac, continuous and hierarchical spike prior for detecting pleiotropic effects on multiple traits. This package is designed for summary statistics containing estimated regression coefficients and their estimated covariance matrices.

**Author(s)**

Taban Baghfalaki <t.baghfalaki@gmail.com>, <t.baghfalaki@modares.ac.ir>

**References**

1. Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Lique, B. (2021). Bayesian meta analysis models for cross cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, **40**(6), 1498-1518.
2. Baghfalaki, T., Sugier, Y. Asgari, P. E., Truong, & Lique, B. (2021). GCPBayes: An R Package for Studying Cross-Phenotype Genetic Associations with Group-Level Bayesian Meta-Analysis. *Submitted*.

---

 HS *Hierarchical Spike*


---

**Description**

Run a Gibbs sampler for a multivariate Bayesian sparse group selection model with hierarchical spike prior for detecting pleiotropic effects on the traits. This function is designed for summary statistics containing estimated regression coefficients and their estimated covariance matrices.

**Usage**

```
HS(
  Betah,
  Sigmah,
  kappa0 = kappa0,
  kappastar0 = kappastar0,
  sigma20 = sigma20,
  s20 = s20,
  m,
  K,
  niter = 1000,
  burnin = 500,
  nthin = 2,
  nchains = 2,
  a1 = 0.1,
  a2 = 0.1,
  d1 = 0.1,
  d2 = 0.1,
  c1 = 1,
  c2 = 1,
  e2 = 1,
  snpnames,
  genename
)
```

**Arguments**

Betah	A list containing m-dimensional vectors of the regression coefficients for K studies.
Sigmah	A list containing the positive definite covariance matrices (m*m-dimensional) which is the estimated covariance matrices of K studies.
kappa0	Initial value for kappa (its dimension is equal to nchains).
kappastar0	Initial value for kappastar (its dimension is equal to nchains).
sigma20	Initial value for sigma2 (its dimension is equal to nchains).
s20	Initial value for s2 (its dimension is equal to nchains).

m	Number of variables in the group.
K	Number of traits.
niter	Number of iterations for the Gibbs sampler.
burnin	Number of burn-in iterations.
nthin	The lag of the iterations used for the posterior analysis is defined (or thinning rate).
nchains	Number of Markov chains, when nchains>1, the function calculates the Gelman-Rubin convergence statistic, as modified by Brooks and Gelman (1998).
a1, a2	Hyperparameters of kappa. Default is a1=0.1 and a2=0.1.
d1, d2	Hyperparameters of sigma2. Default is d1=0.1 and d2=0.1.
c1, c2	Hyperparameters of kappastar. Default is c1=0.1 and c2=0.1.
e2	Initial value for doing Monte Carlo EM algorithm to estimate hyperparameter of s2.
snpname	Names of variables for the group.
genename	Name of group.

### Details

Run a Gibbs sampler using a hierarchical spike.

### Value

- mcmchain: The list of simulation output for all parameters.
- Summary: Summary statistics for regression coefficients in each study.
- Indicator 1: A table containing m rows of binary indicators for each study, the number of studies with nonzero signal and having pleiotropic effect by credible interval (CI). The first K columns show nonzero signals, K+1 th column includes the number of studies with nonzero signal and the last column shows an indicator for having pleiotropic effect of each SNP.
- Indicator 2: A table containing m rows of binary indicators for each study, the number of studies with nonzero signal and having pleiotropic effect by median. The first K columns show nonzero signals, K+1 th column includes the number of studies with nonzero signal and the last column shows an indicator for having pleiotropic effect of each SNP.

### Author(s)

Taban Baghfalaki.

### References

Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Liquet, B. (2021). Bayesian meta analysis models for cross cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, 40(6), 1498-1518.



## Examples

```

data(PARP2_summary)
Breast <- PARP2_summary$Breast
Thyroid <- PARP2_summary$Thyroid
Betah <- list(Breast$beta, Thyroid$beta)
Sigmah <- list(diag(Breast$se), diag(Thyroid$se))
genename <- "PARP2"
snppnames <- Breast$snp
K <- 2
m <- 6

RES <- HS(Betah, Sigmah,
  kappa0 = 0.5, kappastar0 = 0.5, sigma20 = 1, s20 = 1,
  m = m, K = K, niter = 1000, burnin = 500, nthin = 1, nchains = 1,
  a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, c1 = 1, c2 = 1, e2 = 1, snppnames, genename
)
## Not run:
print(RES)
#####Gene DNAJC1 #####
data(DNAJC1)
Breast <- DNAJC1$Breast
Thyroid <- DNAJC1$Thyroid
genename <- "DNAJC1"
snppnames <- Breast$snp
Betah <- list(Breast$beta, Thyroid$beta)
Sigmah <- list(diag(Breast$se^2), diag(Thyroid$se^2))
K <- 2
m <- 14

RES <- HS(Betah, Sigmah,
  kappa0 = 0.5, kappastar0 = 0.5, sigma20 = 1, s20 = 1,
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 1, nchains = 1,
  a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, c1 = 1, c2 = 1, e2 = 1, snppnames, genename)
print(RES)

RES1 <- HS(Betah, Sigmah,
  kappa0 = c(0.5, 0.3), kappastar0 = c(0.5, 0.3), sigma20 = c(2, 1), s20 = c(1, 2),
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
  a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, c1 = 1, c2 = 1, e2 = 1, snppnames, genename)
print(RES1)
#####Simulated summary level data with K=5 #####
data(Simulated_summary)
genename <- Simulated_summary$genename
snppnames <- Simulated_summary$snppnames
Betah <- Simulated_summary$simBeta
Sigmah <- Simulated_summary$simSIGMA
K <- 5
m <- 10

RES <- HS(Betah, Sigmah,
  kappa0 = 0.5, kappastar0 = 0.5, sigma20 = 1, s20 = 1,
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1,

```

```

a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, c1 = 1, c2 = 1, e2 = 1, snpnames, geneName)
print(RES)

RES1 <- HS(Betah, Sigmah,
  kappa0 = c(0.5, 0.3), kappastar0 = c(0.5, 0.3), sigma20 = c(2, 1), s20 = c(1, 2),
  m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
  a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, c1 = 1, c2 = 1, e2 = 1, snpnames, geneName)
print(RES1)
#####Gene PARP2 #####
library(BhGLM)
data(PARP2)
Breast <- PARP2$Breast
Thyroid <- PARP2$Thyroid
geneName <- "PARP2"
snpnames <- c("rs3093872", "rs3093921", "rs1713411", "rs3093926", "rs3093930", "rs878156")

Fit1 <- BhGLM::bglm(y1~ ., family=binomial(link="logit"),data=Breast)
Betah1 <- Fit1$coefficients[-1]
Sigmah1 <- cov(coef(arm::sim(Fit1)))[-1,-1]

Fit2 <- BhGLM::bglm(y2~ ., family=binomial(link="logit"),data=Thyroid)
Betah2 <- Fit2$coefficients[-1]
Sigmah2 <- cov(coef(arm::sim(Fit2)))[-1,-1]

Betah <- list(Betah1,Betah2)
Sigmah <- list(Sigmah1,Sigmah2)
K <- 2
m <- 6

RES <- HS(Betah, Sigmah, kappa0=0.5, kappastar0=0.5, sigma20=1, s20=1,
  m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=1,
  a1=0.1, a2=0.1, d1=0.1, d2=0.1, c1=1, c2=1, e2=1, snpnames, geneName)
print(RES)

RES1 <- HS(Betah, Sigmah, kappa0=c(0.5,0.3), kappastar0=c(0.5,0.3), sigma20=c(2,1), s20=c(1,2),
  m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
  a1=0.1, a2=0.1, d1=0.1, d2=0.1, c1=1, c2=1, e2=1, snpnames, geneName)
print(RES1)
#####Simulated individual level data with K=3 and continuous phynotype #####
library(BhGLM)
data(Simulated_individual)
Study1 <- Simulated_individual$Study1
Study2 <- Simulated_individual$Study2
Study3 <- Simulated_individual$Study3
K <- 3
m <- 30
geneName <- "Simulated"
snpnames <- sprintf("SNP%s",seq(1:m))

Fit1 <- BhGLM::bglm(Y1~ ., family=gaussian,data=data.frame(Study1))
Betah1 <- Fit1$coefficients[-1]

```

```

Sigmah1 <- cov(coef(arm::sim(Fit1)))[-1,-1]

Fit2 <- BhGLM::bglm(Y2~ ., family=gaussian,data=data.frame(Study2))
Betah2 <- Fit2$coefficients[-1]
Sigmah2 <- cov(coef(arm::sim(Fit2)))[-1,-1]

Fit3 <- BhGLM::bglm(Y3~ ., family=gaussian,data=data.frame(Study3))
Betah3 <- Fit3$coefficients[-1]
Sigmah3 <- cov(coef(arm::sim(Fit3)))[-1,-1]

Betah <- list(Betah1,Betah2,Betah3)
Sigmah <- list(Sigmah1,Sigmah2,Sigmah3)

RES <- HS(Betah, Sigmah, kappa0=0.5, kappastar0=0.5, sigma20=1, s20=1,
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=1,
          a1=0.1, a2=0.1, d1=0.1, d2=0.1, c1=1, c2=1, e2=1, snpnames, geneName)
print(RES)

RES1 <- HS(Betah, Sigmah, kappa0=c(0.5,0.3), kappastar0=c(0.5,0.3), sigma20=c(2,1), s20=c(1,2),
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
          a1=0.1, a2=0.1, d1=0.1, d2=0.1, c1=1, c2=1, e2=1, snpnames, geneName)
print(RES1)

#####Simulated individual level data with K=2 and gene expression data #####
library(BhGLM)
data(Simulated_individual_survival)
Study1 <- Simulated_individual_survival$Study1
Study2 <- Simulated_individual_survival$Study2
K <- 2
m <- 10
geneName <- "Simulated"
snpnames <- sprintf("G%s",seq(1:m))

Fit1=BhGLM::bcoxph(Study1$T ~ Study1$X)
Betah1 <- Fit1$coefficients
Sigmah1 <- Fit1$var

Fit2=BhGLM::bcoxph(Study2$T ~ Study2$X)
Betah2 <- Fit2$coefficients
Sigmah2 <- Fit2$var

Betah <- list(Betah1,Betah2)
Sigmah <- list(Sigmah1,Sigmah2)

RES1 <- HS(Betah, Sigmah, kappa0=c(0.5,0.3), kappastar0=c(0.5,0.3), sigma20=c(2,1), s20=c(1,2),
          m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
          a1=0.1, a2=0.1, d1=0.1, d2=0.1, c1=1, c2=1, e2=1, snpnames, geneName)
print(RES1)

## End(Not run)

```

MCMCplot

*MCMC plot***Description**

Trace plot, density plot and ACF plot for the output of CS/DS/HS. The plot is able to draw at most ten SNPs.

**Usage**

```
MCMCplot(Result = Result, k = k, nchains = nchains, whichsnps = whichsnps)
```

**Arguments**

Result	All the generated results by CS/DS/HS function.
k	The number of study for drawing plots, k=1,2,...,K.
nchains	Number of Markov chains run in Result.
whichsnps	The name of SNPs.

**Details**

Trace plot, density plot and ACF plot for the output of CS/DS/HS for checking convergence of MCMC chains.

**Author(s)**

Taban Baghfalaki.

**References**

Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Lique, B. (2021). Bayesian meta analysis models for cross cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, 40(6), 1498-1518.

**Examples**

```
#####Gene DNAJC1 #####
data(DNAJC1)
Breast <- DNAJC1$Breast
Thyroid <- DNAJC1$Thyroid
genename <- "DNAJC1"
snpnames <- Breast$snp
Betah <- list(Breast$beta, Thyroid$beta)
Sigmah <- list(diag(Breast$se^2), diag(Thyroid$se^2))
K <- 2
m <- 14
```

```

RES1 <- DS(Betah, Sigmah,
           kappa0 = 0.5, sigma20 = 1,
           m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1,
           a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, snpnames, genename)
)

MCMCplot(Result = RES1, k = 2, nchains = 1, whichsnps = sample(snpnames, 7))
#####Simulated summary level data with K=5 #####
## Not run:
data(Simulated_summary)
genename <- Simulated_summary$genename
snpnames <- Simulated_summary$snpnames
Betah <- Simulated_summary$simBeta
Sigmah <- Simulated_summary$simSIGMA
K <- 5
m <- 10

RES1 <- DS(Betah, Sigmah,
           kappa0 = c(0.2, 0.5), sigma20 = c(1, 2),
           m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 2,
           a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, snpnames, genename)

MCMCplot(Result = RES1, k = 3, nchains = 2, whichsnps = sample(snpnames, 3))

RES1 <- DS(Betah, Sigmah,
           kappa0 = c(0.2, 0.5, 0.6), sigma20 = c(1, 2, 1.5),
           m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 3,
           a1 = 0.1, a2 = 0.1, d1 = 0.1, d2 = 0.1, snpnames, genename)

MCMCplot(Result = RES1, k = 3, nchains = 3, whichsnps = sample(snpnames, 5))
#####Gene DNAJC1 #####
pvalue <- matrix(0, K, m)
for (k in 1:K) {
  pvalue[k, ] <- 2 * pnorm(-abs(Betah[[k]]) / sqrt(diag(Sigmah[[k]])))
}

zinit <- rep(0, K)
for (j in 1:K) {
  index <- 1:m
  PVALUE <- p.adjust(pvalue[j, ])
  SIGNALS <- index[PVALUE < 0.05]
  modelf1 <- rep(0, m)
  modelf1[SIGNALS] <- 1
  if (max(modelf1) == 1) (zinit[j] <- 1)
}

RES <- CS(Betah, Sigmah,
          kappa0 = 0.5, tau20 = 1, zeta0 = zinit,
          m = m, K = K, niter = 2000, burnin = 1000, nthin = 2, nchains = 1, a1 = 0.1, a2 = 0.1,
          c1 = 0.1, c2 = 0.1, sigma2 = 10^-3, snpnames = snpnames, genename = genename)

MCMCplot(Result = RES1, k = 1, nchains = 1, whichsnps = sample(snpnames, 7))

```

```
#####Gene PARP2 #####
library(BhGLM)
data(PARP2)
Breast <- PARP2$Breast
Thyroid <- PARP2$Thyroid
genename <- "PARP2"
snpnames <- c("rs3093872", "rs3093921", "rs1713411", "rs3093926", "rs3093930", "rs878156")

Fit1 <- BhGLM::bglm(y1~ ., family=binomial(link="logit"),data=Breast)
Betah1 <- Fit1$coefficients[-1]
Sigmah1 <- cov(coef(arm::sim(Fit1)))[-1,-1]

Fit2 <- BhGLM::bglm(y2~ ., family=binomial(link="logit"),data=Thyroid)
Betah2 <- Fit2$coefficients[-1]
Sigmah2 <- cov(coef(arm::sim(Fit2)))[-1,-1]

Betah <- list(Betah1,Betah2)
Sigmah <- list(Sigmah1,Sigmah2)
K <- 2
m <- 6

RES1 <- DS(Betah, Sigmah, kappa0=c(0.2,0.5), sigma20=c(1,2),
           m=m, K=K, niter=1000, burnin=500, nthin=1, nchains=2,
           a1=0.1, a2=0.1, d1=0.1, d2=0.1, snpnames, genename)

MCMCplot(Result=RES1, k=1, nchains=2, whichsnps=snpnames)

RES1 <- DS(Betah, Sigmah, kappa0=c(0.2,0.5), sigma20=c(1,2),
           m=m, K=K, niter=2000, burnin=1000, nthin=2, nchains=2,
           a1=0.1, a2=0.1, d1=0.1, d2=0.1, snpnames, genename)

MCMCplot(Result=RES1, k=1, nchains=2, whichsnps=snpnames)

## End(Not run)
```

---

PARP2

*Gene PARP2 from CECILE study*


---

### Description

A list containing the individual level data for gene PARP2 including genotypes and phenotypes for pleiotropy investigation of breast and thyroid cancers. It is from CECILE study, a French population-based case-control study on breast cancer and from the French studies included in the EPITHYR consortium on thyroid cancer.

**Usage**

PARP2

**Format**

A list which contains two matrices.

**Breast** Summary statistics of breast cancer including the name of SNPs, beta and se

**Thyroid** Summary statistics of thyroid cancer including the name of SNPs, beta and se

**References**

1. Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Liquet, B. (2021). Bayesian meta analysis models for cross cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, **40**(6), 1498-1518.
2. Baghfalaki, T., Sugier, Y. Asgari, P. E., Truong, & Liquet, B. (2021). GCPBayes: An R Package for Studying Cross-Phenotype Genetic Associations with Group-Level Bayesian Meta-Analysis. *Submitted*.

**See Also**

[GCPBayes](#)

---

PARP2\_summary

*Summary statistics of gene PARP2 from CECILE study*

---

**Description**

The summary statistics data for PARP2 protein coding gene including beta and standard error for pleiotropy investigation of breast and thyroid cancers.

**Usage**

PARP2\_summary

**Format**

A list which contains two matrices for the summary statistics of each study.

**Breast** Summary statistics of breast cancer including the name of SNPs, beta and se

**Thyroid** Summary statistics of thyroid cancer including the name of SNPs, beta and se

## References

1. Baghfalaki, T., Sugier, P. E., Truong, T., Pettitt, A. N., Mengersen, K., & Lique, B. (2021). Bayesian meta analysis models for cross cancer genomic investigation of pleiotropic effects using group structure. *Statistics in Medicine*, **40**(6), 1498-1518.
2. Baghfalaki, T., Sugier, Y. Asgari, P. E., Truong, & Lique, B. (2021). GCPBayes: An R Package for Studying Cross-Phenotype Genetic Associations with Group-Level Bayesian Meta-Analysis. *Submitted*.

## See Also

[GCPBayes](#)

---

Simulated\_individual    *Simulated individual level data*

---

## Description

A list containing the individual level data including genotypes and phenotypes for pleiotropy investigation of three studies.

## Usage

```
Simulated_individual
```

## Format

A list which contains the name of gene, the name of the SNPs, the vectors and the matrices for the summary statistics of each study.

**Study1** The individual level data including genotypes and phenotypes of Study 1

**Study2** The individual level data including genotypes and phenotypes of Study 2

**Study3** The individual level data including genotypes and phenotypes of Study 3

## References

,

Baghfalaki, T., Sugier, Y. Asgari, P. E., Truong, & Lique, B. (2021). GCPBayes: An R Package for Studying Cross-Phenotype Genetic Associations with Group-Level Bayesian Meta-Analysis. *Submitted*.

## See Also

1. [GCPBayes](#)



---

Simulated\_individual\_survival  
*Simulated individual level survival data*

---

**Description**

A list containing the individual level data including gene expression data for survival outcomes for pleiotropy investigation of two studies. #' @name Simulated\_individual\_survival

**Usage**

```
Simulated_individual_survival
```

**Format**

A list which contains the name of gene, the name of the SNPs, the vectors and the matrices for the summary statistics of each study.

**Study1** The individual level data including survival outcomes and gene expression data of Study 1

**Study2** The individual level data including survival outcomes and gene expression data of Study 2

**References**

Baghfalaki, T., Sugier, Y. Asgari, P. E., Truong, & Liquet, B. (2021). GCPBayes: An R Package for Studying Cross-Phenotype Genetic Associations with Group-Level Bayesian Meta-Analysis. *Submitted*.

**See Also**

1. [GCPBayes](#)

---

Simulated\_summary      *Simulated summary statistics for K=5 traits*

---

**Description**

A list containing the summary statistics including regression coefficients and covariance matrices for K=5 studies.

**Usage**

```
Simulated_summary
```

**Format**

A list which contains the name of gene, the name of the SNPs, the vectors and the matrices for the summary statistics of each study.

**genename** The name of gene

**snpnames** The name of the SNPs

**simBeta** The regression coefficients of the studies

**simSIGMA** The covariance matrices for the studies

**References**

Baghfalaki, T., Sugier, Y. Asgari, P. E., Truong, & Liqueet, B. (2021). GCPBayes: An R Package for Studying Cross-Phenotype Genetic Associations with Group-Level Bayesian Meta-Analysis. *Submitted*.

**See Also**

1. [GCPBayes](#)

# Index

- \* **CS**
  - GCPBayes, [14](#)
- \* **DS**
  - GCPBayes, [14](#)
- \* **HS**
  - GCPBayes, [14](#)
- \* **MCMCplot**
  - GCPBayes, [14](#)
- \* **datasets**
  - DNAJC1, [8](#)
  - PARP2, [22](#)
  - PARP2\_summary, [23](#)
  - Simulated\_individual, [24](#)
  - Simulated\_individual\_survival, [25](#)
  - Simulated\_summary, [25](#)

[CS](#), [2](#)

[DNAJC1](#), [8](#)

[DS](#), [9](#)

[e2\\_Monte\\_Carlo\\_EM](#), [13](#)

[GCPBayes](#), [8](#), [14](#), [23–26](#)

[HS](#), [15](#)

[MCMCplot](#), [20](#)

[PARP2](#), [22](#)

[PARP2\\_summary](#), [23](#)

[Simulated\\_individual](#), [24](#)

[Simulated\\_individual\\_survival](#), [25](#)

[Simulated\\_summary](#), [25](#)