

# Package: CPBayes (via r-universe)

September 4, 2024

**Title** Bayesian Meta Analysis for Studying Cross-Phenotype Genetic Associations

**Version** 1.1.0

**Date** 2020-12-01

**Author** Arunabha Majumdar <statgen.arunabha@gmail.com> [aut, cre],  
Tanushree Haldar <tanushree.haldar@gmail.com> [aut], John Witte  
[ctb]

**Maintainer** Arunabha Majumdar <statgen.arunabha@gmail.com>

**Description** A Bayesian meta-analysis method for studying cross-phenotype genetic associations. It uses summary-level data across multiple phenotypes to simultaneously measure the evidence of aggregate-level pleiotropic association and estimate an optimal subset of traits associated with the risk locus. CPBayes is based on a spike and slab prior. The methodology is available from: A Majumdar, T Haldar, S Bhattacharya, JS Witte (2018) <doi:10.1371/journal.pgen.1007139>.

**Depends** R (>= 3.2.0)

**License** GPL-3

**LazyData** TRUE

**URL** <https://github.com/ArunabhaCodes/CPBayes>

**BugReports** <https://github.com/ArunabhaCodes/CPBayes/issues>

**RoxygenNote** 7.1.1

**Suggests** testthat, knitr, rmarkdown

**VignetteBuilder** knitr

**Imports** MASS, stats, forestplot, grDevices, purrr, mvtnorm

**Repository** <https://arunabhacodes.r-universe.dev>

**RemoteUrl** <https://github.com/arunabhacodes/cpbayes>

**RemoteRef** HEAD

**RemoteSha** 4df15cb9d3da002e578ab57afd6adc9a066c9ed1

## Contents

analytic_locFDR_BF_cor . . . . .	2
analytic_locFDR_BF_uncor . . . . .	3
CPBayes . . . . .	5
cpbayes_cor . . . . .	6
cpbayes_uncor . . . . .	8
estimate_corln . . . . .	11
ExampleDataCor . . . . .	13
ExampleDataUncor . . . . .	14
forest_cpbayes . . . . .	14
post_summaries . . . . .	16
SampleOverlapMatrix . . . . .	18

<b>Index</b>	<b>19</b>
--------------	-----------

---

analytic\_locFDR\_BF\_cor

*Analytic calculation of the local FDR & Bayes factor for correlated summary statistics.*

---

### Description

Run the [analytic\\_locFDR\\_BF\\_cor](#) function to analytically compute the local FDR & Bayes factor (BF) that quantifies the evidence of aggregate-level pleiotropic association for correlated summary statistics. Here a fixed value of slab variance is considered instead of a range of it in [cpbayes\\_cor](#).

### Usage

```
analytic_locFDR_BF_cor(BetaHat, SE, Corln, SpikeVar = 1e-04, SlabVar = 0.8)
```

### Arguments

BetaHat	A numeric vector of length K where K is the number of phenotypes. It contains the beta-hat values across studies/traits. No default.
SE	A numeric vector with the same dimension as BetaHat providing the standard errors corresponding to BetaHat. Every element of SE must be positive. No default.
Corln	A numeric square matrix of order K by K providing the correlation matrix of BetaHat. The number of rows & columns of Corln must be the same as the length of BetaHat. No default is specified. See <a href="#">estimate_corln</a> .
SpikeVar	Variance of spike (normal distribution with small variance) representing the null effect distribution. Default is $10^{-4}$ .
SlabVar	Variance of slab normal distribution representing the non-null effect distribution. Default is 0.8.

**Value**

The output produced by the function is a list which consists of the local FDR and log10(Bayes factor).

locFDR	It provides the analytically computed local false discovery rate (posterior probability of null association) under CPBayes model (a Bayesian analog of the p-value) which is a measure of the evidence of the aggregate-level pleiotropic association. Bayes factor is adjusted for prior odds, but locFDR is solely a function of the posterior odds.
log10_BF	It provides the analytically computed log10(Bayes factor) produced by CPBayes that measures the evidence of the overall pleiotropic association.

**References**

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. PLoS Genet 14(2): e1007139.

**See Also**

[cpbayes\\_cor](#), [estimate\\_corln](#), [analytic\\_locFDR\\_BF\\_uncor](#), [cpbayes\\_uncor](#), [post\\_summaries](#), [forest\\_cpbayes](#)

**Examples**

```
data(ExampleDataCor)
BetaHat <- ExampleDataCor$BetaHat
BetaHat
SE <- ExampleDataCor$SE
SE
cor <- ExampleDataCor$cor
cor
result <- cpbayes_cor(BetaHat, SE, cor)
str(result)
```

---

analytic\_locFDR\_BF\_uncor

*Analytic calculation of the local FDR & Bayes factor for uncorrelated summary statistics.*

---

**Description**

Run the [analytic\\_locFDR\\_BF\\_uncor](#) function to analytically compute the local FDR & Bayes factor (BF) that quantifies the evidence of aggregate-level pleiotropic association for uncorrelated summary statistics. Here a fixed value of slab variance is considered instead of a range of it in [cpbayes\\_uncor](#).

**Usage**

```
analytic_locFDR_BF_uncor(BetaHat, SE, SpikeVar = 1e-04, SlabVar = 0.8)
```

**Arguments**

BetaHat	A numeric vector of length K where K is the number of phenotypes. It contains the beta-hat values across studies/traits. No default.
SE	A numeric vector with the same dimension as BetaHat providing the standard errors corresponding to BetaHat. Every element of SE must be positive. No default.
SpikeVar	Variance of spike (normal distribution with small variance) representing the null effect distribution. Default is $10^{-4}$ .
SlabVar	Variance of slab normal distribution representing the non-null effect distribution. Default is 0.8.

**Value**

The output produced by the function is a list which consists of the local FDR and  $\log_{10}$ (Bayes factor).

locFDR	It provides the analytically computed local false discovery rate (posterior probability of null association) under CPBayes model (a Bayesian analog of the p-value) which is a measure of the evidence of the aggregate-level pleiotropic association. Bayes factor is adjusted for prior odds, but locFDR is solely a function of the posterior odds.
$\log_{10\_BF}$	It provides the analytically computed $\log_{10}$ (Bayes factor) produced by CPBayes that measures the evidence of the overall pleiotropic association.

**References**

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. PLoS Genet 14(2): e1007139.

**See Also**

[cpbayes\\_uncor](#), [analytic\\_locFDR\\_BF\\_cor](#), [cpbayes\\_cor](#), [estimate\\_corln](#), [post\\_summaries](#), [forest\\_cpbayes](#)

**Examples**

```
data(ExampleDataUncor)
BetaHat <- ExampleDataUncor$BetaHat
BetaHat
SE <- ExampleDataUncor$SE
SE
result <- analytic_locFDR_BF_uncor(BetaHat, SE)
str(result)
```

---

CPBayes

*CPBayes: An R-package implementing a Bayesian meta analysis method for studying cross-phenotype genetic associations.*

---

## Description

Simultaneous analysis of genetic associations with multiple phenotypes may reveal shared genetic susceptibility across traits (pleiotropy). CPBayes is a Bayesian meta analysis method for studying cross-phenotype genetic associations. It uses summary-level data across multiple phenotypes to simultaneously measure the evidence of aggregate-level pleiotropic association and estimate an optimal subset of traits associated with the risk locus. CPBayes is based on a spike and slab prior.

## Details

The package consists of following functions: [analytic\\_locFDR\\_BF\\_uncor](#), [cpbayes\\_uncor](#); [analytic\\_locFDR\\_BF\\_cor](#), [cpbayes\\_cor](#); [post\\_summaries](#), [forest\\_cpbayes](#), [estimate\\_corln](#).

## Functions

[analytic\\_locFDR\\_BF\\_uncor](#) It analytically computes the local FDR (locFDR) and Bayes factor (BF) quantifying the evidence of aggregate-level pleiotropic association for uncorrelated summary statistics.

[cpbayes\\_uncor](#) It implements CPBayes (based on MCMC) for uncorrelated summary statistics to figure out the optimal subset of non-null traits underlying a pleiotropic signal and other insights. The summary statistics across traits/studies are uncorrelated when the studies have no overlapping/genetically related subjects.

[analytic\\_locFDR\\_BF\\_cor](#) It analytically computes the local FDR (locFDR) and Bayes factor (BF) for correlated summary statistics.

[cpbayes\\_cor](#) It implements CPBayes (based on MCMC) for correlated summary statistics to figure out the optimal subset of non-null traits underlying a pleiotropic signal and other insights. The summary statistics across traits/studies are correlated when the studies have overlapping/genetically related subjects or the phenotypes were measured in a cohort study.

[post\\_summaries](#) It summarizes the MCMC data produced by [cpbayes\\_uncor](#) or [cpbayes\\_cor](#). It computes additional summaries to provide a better insight into a pleiotropic signal. It works in the same way for both [cpbayes\\_uncor](#) and [cpbayes\\_cor](#).

[forest\\_cpbayes](#) It creates a forest plot presenting the pleiotropy result obtained by [cpbayes\\_uncor](#) or [cpbayes\\_cor](#). It works in the same way for both [cpbayes\\_uncor](#) and [cpbayes\\_cor](#).

[estimate\\_corln](#) It computes an approximate correlation matrix of the beta-hat vector for multiple overlapping case-control studies using the sample-overlap count matrices.

## References

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. PLoS Genet 14(2): e1007139.

---

 cpbayes\_cor

*Run correlated version of CPBayes.*


---

### Description

Run correlated version of CPBayes when the main genetic effect (beta/log(odds ratio)) estimates across studies/traits are correlated.

### Usage

```
cpbayes_cor(
  BetaHat,
  SE,
  Corln,
  Phenotypes,
  Variant,
  UpdateSlabVar = TRUE,
  MinSlabVar = 0.6,
  MaxSlabVar = 1,
  MCMCiter = 7500,
  Burnin = 500
)
```

### Arguments

BetaHat	A numeric vector of length K where K is the number of phenotypes. It contains the beta-hat values across studies/traits. No default is specified.
SE	A numeric vector with the same dimension as BetaHat providing the standard errors corresponding to BetaHat. Every element of SE must be positive. No default is specified.
Corln	A numeric square matrix of order K by K providing the correlation matrix of BetaHat. The number of rows & columns of Corln must be the same as the length of BetaHat. No default is specified. See <a href="#">estimate_corln</a> .
Phenotypes	A character vector of the same length as BetaHat providing the name of the phenotypes. Default is specified as trait1, trait2, . . . , traitK. Note that BetaHat, SE, Corln, and Phenotypes must be in the same order.
Variant	A character vector of length one providing the name of the genetic variant. Default is 'Variant'.
UpdateSlabVar	A logical vector of length one. If TRUE, the variance of the slab distribution that presents the prior distribution of non-null effects is updated at each MCMC iteration in a range (MinSlabVar – MaxSlabVar) (see next). If FALSE, it is fixed at (MinSlabVar + MaxSlabVar)/2. Default is TRUE.
MinSlabVar	A numeric value greater than 0.01 providing the minimum value of the variance of the slab distribution. Default is 0.6.

MaxSlabVar	A numeric value smaller than 10.0 providing the maximum value of the variance of the slab distribution. Default is 1.0. <b>**Note that, a smaller value of the slab variance will increase the sensitivity of CPBayes while selecting the optimal subset of associated traits but at the expense of lower specificity. Hence the slab variance parameter in CPBayes is inversely related to the level of false discovery rate (FDR) in a frequentist FDR controlling procedure. For a specific dataset, an user can experiment different choices of these three arguments: UpdateSlabVar, MinSlabVar, and MaxSlabVar.</b>
MCMCiter	A positive integer greater than or equal to 2200 providing the total number of iterations in the MCMC. Default is 7500.
Burnin	A positive integer greater than or equal to 200 providing the burn in period in the MCMC. Default is 200. Note that the MCMC sample size (MCMCiter - Burnin) must be at least 2000, which is 7000 by default.

### Value

The output produced by [cpbayes\\_cor](#) is a list which consists of various components.

variantName	It is the name of the genetic variant provided by the user. If not specified by the user, default name is 'Variant'.
log10_BF	It provides the log10(Bayes factor) produced by CPBayes that measures the evidence of the overall pleiotropic association.
locFDR	It provides the local false discovery rate (posterior probability of null association) produced by CPBayes which is a measure of the evidence of aggregate-level pleiotropic association. Bayes factor is adjusted for prior odds, but locFDR is solely a function of the posterior odds. locFDR can sometimes be small indicating an association, but log10_BF may not indicate an association. Hence, always check both log10_BF and locFDR.
subset	It provides the optimal subset of associated/non-null traits selected by CPBayes. It is NULL if no phenotype is selected.
important_traits	It provides the traits which yield a trait-specific posterior probability of association (PPA <sub>j</sub> ) > 20%. Even if a phenotype is not selected in the optimal subset of non-null traits, it can produce a non-negligible value of PPA <sub>j</sub> . Note that, 'important_traits' is expected to include the traits already contained in 'subset'. It provides both the name of the important traits and their corresponding value of PPA <sub>j</sub> . Always check 'important_traits' even if 'subset' contains a single trait. It helps to better explain an observed pleiotropic signal.
auxi_data	It contains supplementary data including the MCMC data which is used later by <a href="#">post_summaries</a> and <a href="#">forest_cpbayes</a> : <ol style="list-style-type: none"> <li>1. traitNames: Name of all the phenotypes.</li> <li>2. K: Total number of phenotypes.</li> <li>3. mcmc.samplesize: MCMC sample size.</li> <li>4. PPA<sub>j</sub>: Trait-specific posterior probability of association for all the traits.</li> <li>5. Z.data: MCMC data on the latent association status of all the traits (Z).</li> </ol>

6. sim.beta: MCMC data on the unknown true genetic effect (beta) on each trait.
  7. betahat: The beta-hat vector provided by the user which will be used by [forest\\_cpbayes](#).
  8. se: The standard error vector provided by the user which will be used by [forest\\_cpbayes](#).
- uncor\_use 'Yes' or 'No'. Whether the combined strategy of CPBayes (implemented for correlated summary statistics) used the uncorrelated version or not.
- runtime It provides the runtime (in seconds) taken by [cpbayes\\_cor](#). It will help the user to plan the whole analysis.

## References

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. *PLoS Genet* 14(2): e1007139.

## See Also

[analytic\\_locFDR\\_BF\\_cor](#), [estimate\\_corln](#), [post\\_summaries](#), [forest\\_cpbayes](#), [analytic\\_locFDR\\_BF\\_uncor](#), [cpbayes\\_uncor](#)

## Examples

```
data(ExampleDataCor)
BetaHat <- ExampleDataCor$BetaHat
BetaHat
SE <- ExampleDataCor$SE
SE
cor <- ExampleDataCor$cor
cor
traitNames <- paste("Disease", 1:10, sep = "")
SNP1 <- "rs1234"
result <- cpbayes_cor(BetaHat, SE, cor, Phenotypes = traitNames, Variant = SNP1)
str(result)
```

---

cpbayes\_uncor

*Run uncorrelated version of CPBayes.*

---

## Description

Run uncorrelated version of CPBayes when the main genetic effect (beta/log(odds ratio)) estimates across studies/traits are uncorrelated.



**Usage**

```
cpbayes_uncor(
  BetaHat,
  SE,
  Phenotypes,
  Variant,
  UpdateSlabVar = TRUE,
  MinSlabVar = 0.6,
  MaxSlabVar = 1,
  MCMCiter = 7500,
  Burnin = 500
)
```

**Arguments**

BetaHat	A numeric vector of length K where K is the number of phenotypes. It contains the beta-hat values across studies/traits. No default is specified.
SE	A numeric vector with the same dimension as BetaHat providing the standard errors corresponding to BetaHat. Every element of SE must be positive. No default is specified.
Phenotypes	A character vector of the same length as BetaHat providing the name of the phenotypes. Default is specified as trait1, trait2, . . . , traitK. Note that BetaHat, SE, and Phenotypes must be in the same order.
Variant	A character vector of length one specifying the name of the genetic variant. Default is 'Variant'.
UpdateSlabVar	A logical vector of length one. If TRUE, the variance of the slab distribution that presents the prior distribution of non-null effects is updated at each MCMC iteration in a range (MinSlabVar – MaxSlabVar) (see next). If FALSE, it is fixed at (MinSlabVar + MaxSlabVar)/2. Default is TRUE.
MinSlabVar	A numeric value greater than 0.01 providing the minimum value of the variance of the slab distribution. Default is 0.6.
MaxSlabVar	A numeric value smaller than 10.0 providing the maximum value of the variance of the slab distribution. Default is 1.0. <b>**Note that, a smaller value of the slab variance will increase the sensitivity of CPBayes while selecting the optimal subset of associated traits but at the expense of lower specificity. Hence the slab variance parameter in CPBayes is inversely related to the level of false discovery rate (FDR) in a frequentist FDR controlling procedure. For a specific dataset, a user can experiment different choices of these three arguments: UpdateSlabVar, MinSlabVar, and MaxSlabVar.</b>
MCMCiter	A positive integer greater than or equal to 2200 providing the total number of iterations in the MCMC. Default is 7500.
Burnin	A positive integer greater than or equal to 200 providing the burn in period in the MCMC. Default is 500. Note that the MCMC sample size (MCMCiter - Burnin) must be at least 2000, which is 7000 by default.

**Value**

The output produced by the function is a list which consists of various components.

variantName	It is the name of the genetic variant provided by the user. If not specified by the user, default name is 'Variant'.
log10_BF	It provides the log10(Bayes factor) produced by CPBayes that measures the evidence of the overall pleiotropic association.
locFDR	It provides the local false discovery rate (posterior probability of null association) produced by CPBayes which is a measure of the evidence of the aggregate-level pleiotropic association. Bayes factor is adjusted for prior odds, but locFDR is solely a function of the posterior odds. locFDR can sometimes be small indicating an association, but log10_BF may not indicate an association. Hence, always check both log10_BF and locFDR.
subset	It provides the optimal subset of associated/non-null traits selected by CPBayes. It is NULL if no phenotype is selected.
important_traits	It provides the traits which yield a trait-specific posterior probability of association (PPAj) > 20%. Even if a phenotype is not selected in the optimal subset of non-null traits, it can produce a non-negligible value of trait-specific posterior probability of association (PPAj). Note that, 'important_traits' is expected to include the traits already contained in 'subset'. It provides both the name of the important traits and their corresponding values of PPAj. Always check 'important_traits' even if 'subset' contains a single trait. It helps to better explain an observed pleiotropic signal.
auxi_data	It contains supplementary data including the MCMC data which is used later by <a href="#">post_summaries</a> and <a href="#">forest_cpbayes</a> : <ol style="list-style-type: none"> <li>1. traitNames: Name of all the phenotypes.</li> <li>2. K: Total number of phenotypes.</li> <li>3. mcmc.samplesize: MCMC sample size.</li> <li>4. PPAj: Trait-specific posterior probability of association for all the traits.</li> <li>5. Z.data: MCMC data on the latent association status of all the traits (Z).</li> <li>6. sim.beta: MCMC data on the unknown true genetic effect (beta) on all the traits.</li> <li>7. betahat: The beta-hat vector provided by the user which will be used by <a href="#">forest_cpbayes</a>.</li> <li>8. se: The standard error vector provided by the user which will be used by <a href="#">forest_cpbayes</a>.</li> </ol>
runtime	It provides the runtime (in seconds) taken by <a href="#">cpbayes_uncor</a> . It will help the user to plan the whole analysis.

**References**

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. *PLoS Genet* 14(2): e1007139.

**See Also**

[analytic\\_locFDR\\_BF\\_uncor](#), [post\\_summaries](#), [forest\\_cpbayes](#), [analytic\\_locFDR\\_BF\\_cor](#), [cpbayes\\_cor](#), [estimate\\_corln](#)

**Examples**

```
data(ExampleDataUncor)
BetaHat <- ExampleDataUncor$BetaHat
BetaHat
SE <- ExampleDataUncor$SE
SE
traitNames <- paste("Disease", 1:10, sep = "")
SNP1 <- "rs1234"
result <- cpbayes_uncor(BetaHat, SE, Phenotypes = traitNames, Variant = SNP1)
str(result)
```

---

estimate_corln	<i>Estimate correlation structure of beta-hat vector for multiple overlapping case-control studies using sample-overlap matrices.</i>
----------------	---

---

**Description**

It computes an approximate correlation matrix of the estimated beta (log odds ratio) vector for multiple overlapping case-control studies using the sample-overlap matrices which describe the number of cases or controls shared between studies/traits, and the number of subjects who are case for one study/trait but control for another study/trait. For a cohort study, the phenotypic correlation matrix should be a reasonable substitute of this correlation matrix. These approximations are accurate when none of the diseases/traits is associated with the environmental covariates and genetic variant.

**Usage**

```
estimate_corln(n11, n00, n10)
```

**Arguments**

n11	An integer square matrix (number of rows must be the same as the number of studies/traits) providing the number of cases shared between all possible pairs of studies/traits. So (k,l)-th element of n11 is the number of subjects who are case for both k-th and l-th study/trait. Note that the diagonal elements of n11 are the number of cases in the studies/traits. If no case is shared between studies/traits, the off-diagonal elements of n11 will be zero. No default is specified.
n00	An integer square matrix (number of rows must be the same as the number of studies/traits) providing the number of controls shared between all possible pairs of studies/traits. So (k,l)-th element of n00 is the number subjects who are control for both k-th and l-th study/trait. Note that the diagonal elements of n00 are the number of controls in the studies/traits. If no control is shared between studies/traits, the off-diagonal elements will be zero. No default is specified.

`n10` An integer square matrix (number of rows must be the same as the number of studies/traits) providing the number of subjects who are case for one study/trait and control for another study/trait. Clearly, the diagonal elements will be zero. An off diagonal element, e.g., (k,l)-th element of `n10` is the number of subjects who are case for k-th study/trait and control for l-th study/trait. If there is no such overlap, all the elements of `n10` will be zero. No default is specified.

### Details

\*\*\*Important note on the estimation of correlation structure of correlated beta-hat vector:\*\*\* In general, environmental covariates are expected to be present in a study and associated with the phenotypes of interest. Also, a small proportion of genome-wide genetic variants are expected to be associated. Hence the above approximation of the correlation matrix may not be accurate. So in general, we recommend an alternative strategy to estimate the correlation matrix using the genome-wide summary statistics data across traits as follows. First, extract all the SNPs for each of which the trait-specific univariate association p-value across all the traits are  $> 0.1$ . The trait-specific univariate association p-values are obtained using the beta-hat and standard error for each trait. Each of the SNPs selected in this way is either weakly or not associated with any of the phenotypes (null SNP). Next, select a set of independent null SNPs from the initial set of null SNPs by using a threshold of  $r^2 < 0.01$  ( $r$ : the correlation between the genotypes at a pair of SNPs). In the absence of in-sample linkage disequilibrium (LD) information, one can use the reference panel LD information for this screening. Finally, compute the correlation matrix of the effect estimates (beta-hat vector) as the sample correlation matrix of the beta-hat vector across all the selected independent null SNPs. This strategy is more general and applicable to a cohort study or multiple overlapping studies for binary or quantitative traits with arbitrary distributions. It is also useful when the beta-hat vector for multiple non-overlapping studies become correlated due to genetically related individuals across studies. Misspecification of the correlation structure can affect the results produced by CPBayes to some extent. Hence, if genome-wide summary statistics data across traits is available, we highly recommend to use this alternative strategy to estimate the correlation matrix of the beta-hat vector. See our paper for more details.

### Value

This function returns an approximate correlation matrix of the beta-hat vector for multiple overlapping case-control studies. See the example below.

### References

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. PLoS Genet 14(2): e1007139.

### See Also

[cpbayes\\_cor](#)

### Examples

```
data(SampleOverlapMatrix)
n11 <- SampleOverlapMatrix$n11
n11
```

```
n00 <- SampleOverlapMatrix$n00
n00
n10 <- SampleOverlapMatrix$n10
n10
cor <- estimate_corln(n11, n00, n10)
cor
```

---

ExampleDataCor

*An example data for correlated summary statistics.*

---

## Description

ExampleDataCor is a list consisting of three components: BetaHat, SE, cor. ExampleDataCor\$BetaHat is a numeric vector that contains the main genetic effect (beta/log(odds ratio)) estimates for a SNP across 10 overlapping case-control studies for 10 different diseases. Each of the 10 studies has a distinct set of 7000 cases and a common set of 10000 controls shared across all the studies. In each case-control study, we fit a logistic regression of the case-control status on the genotype coded as the minor allele count for all the individuals in the sample. One can also include various covariates, such as, age, gender, principal components (PCs) of ancestries in the logistic regression. From each logistic regression for a disease, we obtain the estimate of the main genetic association parameter (beta/log(odds ratio)) along with the corresponding standard error. Since the studies have overlapping subjects, the beta-hat across traits are correlated. ExampleDataCor\$SE contains the standard error vector corresponding to the correlated beta-hat vector. ExampleDataCor\$cor is a numeric square matrix providing the correlation matrix of the correlated beta-hat vector.

## Usage

```
data(ExampleDataCor)
```

## Format

A list consisting of two numeric vectors (each of length 10) and a numeric square matrix of dimension 10 by 10:

**BetaHat** beta hat vector of length 10.

**SE** standard error vector corresponding to the beta-hat vector.

**cor** correlation matrix of the beta-hat vector.

## Examples

```
data(ExampleDataCor)
BetaHat <- ExampleDataCor$BetaHat
BetaHat
SE <- ExampleDataCor$SE
SE
cor <- ExampleDataCor$cor
cor
cpbayes_cor(BetaHat, SE, cor)
```

---

ExampleDataUncor	<i>An example data for uncorrelated summary statistics.</i>
------------------	---

---

### Description

ExampleDataUncor is a list which has two components: BetaHat, SE. The numeric vector ExampleDataUncor\$BetaHat contains the main genetic effect (beta/log(odds ratio)) estimates for a single nucleotide polymorphism (SNP) obtained from 10 separate case-control studies for 10 different diseases. In each case-control study comprising a distinct set of 7000 cases and 10000 controls, we fit a logistic regression of the case-control status on the genotype coded as the minor allele count for all the individuals in the sample. One can also include various covariates, such as, age, gender, principal components (PCs) of ancestries in the logistic regression. From each logistic regression for a disease, we obtain the estimate of the main genetic association parameter (beta/log(odds ratio)) along with the corresponding standard error. Since the studies do not have any overlapping subject, the beta-hat across the traits are uncorrelated. ExampleDataUncor\$SE is the second numeric vector that contains the standard errors corresponding to the uncorrelated beta-hat vector.

### Usage

```
data(ExampleDataUncor)
```

### Format

A list of two numeric vectors each of length 10 (for 10 studies):

**BetaHat** beta hat vector of length 10.

**SE** standard error vector corresponding to beta-hat vector.

### Examples

```
data(ExampleDataUncor)
BetaHat <- ExampleDataUncor$BetaHat
BetaHat
SE <- ExampleDataUncor$SE
SE
cpbayes_uncor(BetaHat, SE)
```

---

forest_cpbayes	<i>Forest plot presenting pleiotropy result obtained by CPBayes.</i>
----------------	--

---

### Description

Run the [forest\\_cpbayes](#) function to create a forest plot that presents the pleiotropy result obtained by [cpbayes\\_uncor](#) or [cpbayes\\_cor](#).

**Usage**

```
forest_cpbayes(mcmc_output, level = 0.05, PPAj_cutoff = 0.01)
```

**Arguments**

mcmc_output	A list returned by either <a href="#">cpbayes_uncor</a> or <a href="#">cpbayes_cor</a> . This list contains all the primary results and MCMC data produced by <a href="#">cpbayes_uncor</a> or <a href="#">cpbayes_cor</a> . No default is specified. See the example below.
level	A numeric value. (1-level)% confidence interval of the unknown true genetic effect (beta/log(odds ratio)) on each trait is plotted in the forest plot. Default choice is 0.05.
PPAj_cutoff	A numeric value. It's a user-specified threshold of PPAj (trait-specific posterior probability of association). Only those traits having PPAj values above this cut-off are included in the forest plot. So, the choice of this variable as '0.0' includes all traits in the forest plot. Default is 0.01.

**Value**

The output produced by this function is a diagram file in .pdf format. The details of the diagram are as follows:

file_name	The pdf file is named after the genetic variant. So, if the argument 'Variant' in <a href="#">cpbayes_uncor</a> or <a href="#">cpbayes_cor</a> is specified as 'rs1234', the figure file is named as rs1234.pdf.
Column1	First column in the figure specifies the name of the phenotypes.
Column2	Second column provides the trait-specific univariate association p-value for a trait.
Column3	Third column provides the trait-specific posterior probability of association (PPAj) produced by CPBayes.
Column4	Fourth column states whether a phenotype was selected in the optimal subset of associated/non-null traits detected by CPBayes. If a phenotype was not selected, selected and positively associated, selected and negatively associated, its association status is stated as null, positive and negative, respectively.
Column5	In the right section of the figure, the primary estimate and confidence interval of the beta/log odds ratio parameter for a trait is plotted.

**References**

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. PLoS Genet 14(2): e1007139.

**See Also**

[cpbayes\\_uncor](#), [cpbayes\\_cor](#)

## Examples

```
data(ExampleDataUncor)
BetaHat <- ExampleDataUncor$BetaHat
SE <- ExampleDataUncor$SE
traitNames <- paste("Disease", 1:10, sep = "")
SNP1 <- "rs1234"
result <- cpbayes_uncor(BetaHat, SE, Phenotypes = traitNames, Variant = SNP1)
## Not run: forest_cpbayes(result, level = 0.05)
```

---

post_summaries	<i>Post summary of the MCMC data generated by the uncorrelated or correlated version of CPBayes.</i>
----------------	--

---

## Description

Run the `post_summaries` function to summarize the MCMC data produced by `cpbayes_uncor` or `cpbayes_cor` and obtain meaningful insights into an observed pleiotropic signal.

## Usage

```
post_summaries(mcmc_output, level = 0.05)
```

## Arguments

mcmc_output	A list returned by either <code>cpbayes_uncor</code> or <code>cpbayes_cor</code> . This list contains the primary results and MCMC data produced by <code>cpbayes_uncor</code> or <code>cpbayes_cor</code> . No default is specified. See the example below.
level	A numeric value. (1-level)% credible interval (Bayesian analog of the confidence interval) of the unknown true genetic effect (beta/odds ratio) on each trait is computed. Default choice is 0.05.

## Value

The output produced by this function is a list that consists of various components.

variantName	It is the name of the genetic variant provided by the user. If not specified by the user, default name is 'Variant'.
log10_BF	It provides the log10(Bayes factor) produced by CPBayes that measures the evidence of the overall pleiotropic association.
locFDR	It provides the local false discovery rate (posterior probability of null association) produced by CPBayes (a Bayesian analog of the p-value) which is a measure of the evidence of aggregate-level pleiotropic association. Bayes factor is adjusted for prior odds, but locFDR is solely a function of posterior odds. locFDR can sometimes be significantly small indicating an association, but log10_BF may not. Hence, always check both log10_BF and locFDR.



subset	A data frame providing the optimal subset of associated/non-null traits along with their trait-specific posterior probability of association (PPAj) and direction of associations. It is NULL if no phenotype is selected by CPBayes.
important_traits	It provides the traits which yield a trait-specific posterior probability of association (PPAj) > 20%. Even if a phenotype is not selected in the optimal subset of non-null traits, it can produce a non-negligible value of trait-specific posterior probability of association. We note that 'important_traits' is expected to include the traits already contained in 'subset'. It provides the name of the important traits and their trait-specific posterior probability of association (PPAj) and the direction of associations. Always check 'important_traits' even if 'subset' contains a single trait. It helps to better explain an observed pleiotropic signal.
traitNames	It returns the name of all the phenotypes specified by the user. Default is trait1, trait2, ... , traitK.
PPAj	Data frame providing the trait-specific posterior probability of association for all the phenotypes.
poste_summary_beta	Data frame providing the posterior summary of the unknown true genetic effect (beta) on each trait. It gives posterior mean, median, standard error, credible interval (lower and upper limits) of the true beta corresponding to each trait.
poste_summary_OR	Data frame providing the posterior summary of the unknown true genetic effect (odds ratio) on each trait. It gives posterior mean, median, standard error, credible interval (lower and upper limits) of the true odds ratio corresponding to each trait.

## References

Majumdar A, Haldar T, Bhattacharya S, Witte JS (2018) An efficient Bayesian meta analysis approach for studying cross-phenotype genetic associations. *PLoS Genet* 14(2): e1007139.

## See Also

[cpbayes\\_uncor](#), [cpbayes\\_cor](#)

## Examples

```
data(ExampleDataUncor)
BetaHat <- ExampleDataUncor$BetaHat
BetaHat
SE <- ExampleDataUncor$SE
SE
traitNames <- paste("Disease", 1:10, sep = "")
SNP1 <- "rs1234"
result <- cpbayes_uncor(BetaHat, SE, Phenotypes = traitNames, Variant = SNP1)
PleioSumm <- post_summaries(result, level = 0.05)
str(PleioSumm)
```

---

SampleOverlapMatrix *An example data of sample-overlap matrices.*

---

### Description

An example data of sample-overlap matrices for five different diseases in the Kaiser GERA cohort (a real data). SampleOverlapMatrix is a list that contains an example of the sample overlap matrices for five different diseases in the Kaiser GERA cohort. SampleOverlapMatrix\$n11 provides the number of cases shared between all possible pairs of diseases. SampleOverlapMatrix\$n00 provides the number of controls shared between all possible pairs of diseases. SampleOverlapMatrix\$n10 provides the number of subjects who are case for one disease and control for another disease.

### Usage

```
data(SampleOverlapMatrix)
```

### Format

A list consisting of three integer square matrices (each of dimension 5 by 5):

**n11** number of cases shared between all possible pairs of diseases.

**n00** number of controls shared between all possible pairs of diseases.

**n10** number of subjects who are case for one disease and control for another disease.

### Examples

```
data(SampleOverlapMatrix)
n11 <- SampleOverlapMatrix$n11
n11
n00 <- SampleOverlapMatrix$n00
n00
n10 <- SampleOverlapMatrix$n10
n10
estimate_corln(n11,n00,n10)
```

# Index

## \* datasets

ExampleDataCor, [13](#)

ExampleDataUncor, [14](#)

SampleOverlapMatrix, [18](#)

analytic\_locFDR\_BF\_cor, [2](#), [2](#), [4](#), [5](#), [8](#), [11](#)

analytic\_locFDR\_BF\_uncor, [3](#), [3](#), [5](#), [8](#), [11](#)

CPBayes, [5](#)

cpbayes\_cor, [2–5](#), [6](#), [7](#), [8](#), [11](#), [12](#), [14–17](#)

cpbayes\_uncor, [3–5](#), [8](#), [8](#), [10](#), [14–17](#)

estimate\_corln, [2–6](#), [8](#), [11](#), [11](#)

ExampleDataCor, [13](#)

ExampleDataUncor, [14](#)

forest\_cpbayes, [3–5](#), [7](#), [8](#), [10](#), [11](#), [14](#), [14](#)

post\_summaries, [3–5](#), [7](#), [8](#), [10](#), [11](#), [16](#), [16](#)

SampleOverlapMatrix, [18](#)