Package 'BayesDC'

April 3, 2025

Type Package

Title Bayesian methods for estimating the degree of dosage compensation on X chromosome.

Version 0.1.0

Author Yi Zhang and Ji-Yuan Zhou

Maintainer Yi Zhang <joyswhale@163.com>

Description This code contains the Bayesian method for estimating the degree of the dosage compensation on X chromosome (denoted as d) for either quantitative traits or qualitative traits, with or without covariates using unrelated subjects.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

R topics documented:

`BayesDC` *The Bayesian method for estimating the degree of the dosage compensation on X chromosome*

Description

This code contains the Bayesian method for estimating the degree of the dosage compensation on X chromosome (denoted as d) for either quantitative traits or qualitative traits, with or without covariates using unrelated subjects.

Usage

```
BayesDC(ped,covariate=NULL,trait_type,
         trait_missing=NA,genotype_missing=NA,covariate_missing=NA,
         prior,chains_num=8,iter_num=3000,warmup_num=1000,acceptance_rate=0.99)
```

Arguments

ped	The data frame "ped" contains the following information: pedigree ID (FID), individual ID (IID), father's ID (PID), mother's ID (MID), sex, phenotype, genotypes. Either father's or mother's ID is set to 0 for founders, i.e. individuals with no parents. Numeric coding for sex is 0 = unknown, 1 = male, 2 = female. Each genotype is coded as 0, 1 or 2 for female and 0, 2 for male, indicating the number of the minor alleles. For quantitative traits, phenotypes are individuals' trait values. Numeric coding for a qualitative trait is 1 = unaffected, 2 = affected. The ped provided to this function should only contain SNPs on X chromosome.
covariate	The covariates needed to be adjusted, can be a txt file or a dataframe/matrix, the first five columns should be pedigree ID (FID), individual ID (IID), father's ID (PID), mother's ID (MID), sex.
trait_type	A character string either being "quantitative" or "qualitative", indicating the type of the trait.
trait_missing	The input variable "trait_missing" is the missing value for the trait in the data file, and the default value is NA. It may be 9 in some data files; or other numeric value.
genotype_missing	The input variable "genotype_missing" represents that the allele at the locus is missing, and the default value is 0. It may be 9 in some data files; or other numeric value.
covariate_missing	The input variable "covariate_missing" is the missing value for the covariates in the data file, and the default value is NA.
prior	A character string either being "normal" or "uniform". "prior"="normal" represents that the prior distribution of d is a truncated normal distribution with mean=2 and sd=1, and the values ranging from 1/4 to 4, and the prior distributions of other unknown parameters are consistent with those proposed in the paper; "prior"="uniform" represents that the prior distribution of d is the uniform distribution specified in our paper, that is, d~U(1/4, 4), and the prior distributions of other unknown parameters are consistent with those proposed in the paper.
chains_num	A positive integer specifying the number of the Markov chains. The default number is 8.
iter_num	A positive integer specifying the number of the iterations for each chain (including warmup). The default number is 3,000.
warmup_num	A positive integer specifying the number of the warmup (also known as burnin) iterations per chain. The number of the warmup iterations should be smaller than the number of the iterations and the default is 1,000.
acceptance_rate	A value between 0 and 1 which represents the target acceptance rate, and the default is 0.99.

Details

Please install the "rstan" package and make sure that it can work before using this function. Note that we estimate the degree of the dosage compensation on X chromosome in the presence of association. The results may be different for different runs, because of the sampling randomness of the HMC algorithm. If the fixed results are wanted, the seed number should be set before running the function. Note that different version of R may lead to different results under the same seed number. The results of the examples given in this file are obtained under the R with version 4.1.2. Meanwhile, we recommend using "chains_num"=8, "iter_num"=3,000, "warmup_num"=1,000 and "acceptance_rate"=0.99 in practical applications.

Value

`Point_Estimate`	The point estimate of the degree of the dosage compensation on X chromosome for the SNP based on the Bayesian method.
`HPDI_Lower`	The lower bound of the HPDI.
`HPDI_Upper`	The upper bound of the HPDI.

Author(s)

Yi Zhang and Ji-Yuan Zhou

References

Yi Zhang. Bayesian methods for estimating the degree of dosage compensation on X chromosome. 2025

Annis, J.; Miller, B. J.; Palmeri, T. J. Bayesian inference with Stan: A tutorial on adding custom distributions. Behav. Res. Methods 2017, 49, 863-886.

Examples

```
#install.package("rstan")
library("rstan")
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)

##example 1:
##quantitative trait with covariate
##the prior distribution of d is a truncated normal distribution specified in our paper
set.seed(123)
BayesDC(ped1,covariate=covar,trait_type = "quantitative",
       trait_missing=NA,genotype_missing=NA,covariate_missing=NA,
       prior = "normal",chains_num=8,iter_num=3000,warmup_num=1000,acceptance_rate=0.99)

##example 2:
##quantitative trait with covariate
##the prior distribution of d is a uniform distribution specified in our paper
##the prior distributions of other unknown parameters are consistent with those in our paper
set.seed(123)
BayesDC(ped1,covariate=covar,trait_type = "quantitative",
       trait_missing=NA,genotype_missing=NA,covariate_missing=NA,
      prior = "uniform",chains_num=8,iter_num=3000,warmup_num=1000,acceptance_rate=0.99)

##example 3:
##quantitative trait without covariate
```

```
##the prior distribution of d is a truncated normal distribution specified in our paper
set.seed(123)
BayesDC(ped1,covariate=NULL,trait_type = "quantitative",
       trait_missing=NA,genotype_missing=NA,covariate_missing=NA,
       prior = "normal",chains_num=8,iter_num=3000,warmup_num=1000,acceptance_rate=0.99)

##example 4:
##qualitative trait with covariate
##the prior distribution of d is a truncated normal distribution specified in our paper
set.seed(123)
BayesDC(ped2,covariate=covar,trait_type = "qualitative",
       trait_missing=NA,genotype_missing=NA,covariate_missing=NA,
       prior = "normal",chains_num=8,iter_num=3000,warmup_num=1000,acceptance_rate=0.99)

##example 5:
##qualitative trait with covariate
##the prior distribution of d is a uniform distribution specified in our paper
set.seed(123)
BayesDC(ped2,covariate=covar,trait_type = "qualitative",
       trait_missing=NA,genotype_missing=NA,covariate_missing=NA,
      prior = "uniform",chains_num=8,iter_num=3000,warmup_num=1000,acceptance_rate=0.99)

##example 6:
##qualitative trait without covariate
##the prior distribution of d is a uniform distribution specified in our paper
set.seed(123)
BayesDC(ped2,covariate=NULL,trait_type = "qualitative",
       trait_missing=NA,genotype_missing=NA,covariate_missing=NA,
      prior = "uniform",chains_num=8,iter_num=3000,warmup_num=1000,acceptance_rate=0.99)
```

`covar` *A dataset of covariates.*

Description

The pedigree information in the first five columns includes: pedigree ID (pid), individual ID (iid), father ID (fid), mother ID (mid) and sex. The six and seven is tow covariates.

Usage

```
covar
```

Format

A dataset for 1,000 unrelated subjects and seven variables contain two covariates.

FID Pedigree ID.

IID Individual ID.

PID Father ID.

MID Mother ID.

sex The genetic sex of the individual, coded as 1 for males and 2 for females, following the PLINK default coding.

cov_1 A numeric variable of continuous covariate.

cov_2 A numeric variable of dichotomy covariate.

`HPDIofHMC` *A function to obtain the highest posterior density interval of the samples*

Description

A function to obtain the highest posterior density interval of the samples.

Usage

```
HPDIofHMC(sampleVec,credMass=0.95)
```

Arguments

`sampleVec`	A vector contains the samples.
`credMass`	A value between 0 and 1 used to specify the probability of the samples which should be included in an interval, and the default is 0.95.

Value

A vector contains the lower bound and the upper bound of the highest posterior density interval.

Examples

```
HPDIofHMC(rnorm(100,1,1),credMass = 0.95)
```

`modeofHMC` *A function to obtain the mode of the samples*

Description

A function to obtain the mode of the samples.

Usage

```
modeofHMC(sam_chain)
```

Arguments

`sam_chain`	A vector contains the samples.

Value

The modeofHMC() returns a value.

Examples

```
modeofHMC(runif(100,5,50))
```

---

ped1 *A dataset of simulated quantitative trait and two SNPs.*

Description

The pedigree information in the first five columns includes: pedigree ID (pid), individual ID (iid), father ID (fid), mother ID (mid) and sex. The six five column is quantitative trait, and seven and eight is tow SNPs.

Usage

```
ped1
```

Format

A dataset for 1,000 unrelated subjects and eight variables.

FID Pedigree ID.

IID Individual ID.

PID Father ID.

MID Mother ID.

sex The genetic sex of the individual, coded as 1 for males and 2 for females, following the PLINK default coding.

phen A numeric variable of the quantitative trait.

geno1 A numeric variable of code of SNP. Each genotype is coded as 0, 1 or 2 for female and 0, 2 for male, indicating the number of the minor alleles.

geno2 A numeric variable of code of SNP. Each genotype is coded as 0, 1 or 2 for female and 0, 2 for male, indicating the number of the minor alleles.

ped2 *A dataset of simulated qualitative trait and two SNPs.*

Description

The pedigree information in the first five columns includes: pedigree ID (pid), individual ID (iid), father ID (fid), mother ID (mid) and sex. The six five column is qualitative trait, and seven and eight is tow SNPs.

Usage

```
ped2
```

Format

A dataset for 1,000 unrelated subjects and eight variables.

FID Pedigree ID.

IID Individual ID.

PID Father ID.

MID Mother ID.

sex The genetic sex of the individual, coded as 1 for males and 2 for females, following the PLINK default coding.

phen A numeric variable of the qualitative trait.

geno1 A numeric variable of code of SNP. Each genotype is coded as 0, 1 or 2 for female and 0, 2 for male, indicating the number of the minor alleles.

geno2 A numeric variable of code of SNP. Each genotype is coded as 0, 1 or 2 for female and 0, 2 for male, indicating the number of the minor alleles.

Index