Package 'MBEXCIS'

August 3, 2022

Type Package

Title Efficient Bayesian method for estimating the degree of skewness of X chromosome inactivation based on mixtures of pedigree and unrelated data

Version 0.1.0

Author Yi-Fan Kong and Ji-Yuan Zhou

Maintainer
Yi-Fan Kong <kongyifan_s@163.com> and Ji-Yuan Zhou <zhoujiyuan5460@hotmail.com>

Description This code contains the Bayesian method for estimating the degree of the skewness of X chromosome inactivation for either quantitative traits or qualitative traits, with or without covariates using pedigree data or mixtures of pedigree and unrelated data

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.2.1

R topics documented:

covariate_mixture1_2 *A dataset containing the covariates for quantitative traits based on mixture data*

Description

The dataset contains two covariates for 30 pedigrees, 260 pedigree-related individuals and 260 unrelated individuals.

Usage

```
covariate_mixture1_2
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

covariate1 A quantitative covariate.

covariate2 A qualitative covariate.

covariate_mixture3_4 *A dataset containing the covariates for qualitative traits based on mixture data*

Description

The dataset contains two covariates for 30 pedigrees, 260 pedigree-related individuals and 260 unrelated individuals

Usage

```
covariate_mixture3_4
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

covariate1 A quantitative covariate.

covariate2 A qualitative covariate.

covariate_pedigree1_2 *A dataset containing the covariates for quantitative traits based on pedigree data*

Description

The dataset contains two covariates for 30 pedigrees, 260 pedigree-related individuals.

Usage

```
covariate_pedigree1_2
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

covariate1 A quantitative covariate.

covariate2 A qualitative covariate.

covariate_pedigree3_4 *A dataset containing the covariates for qualitative traits based on pedigree data*

Description

The dataset contains two covariates for 30 pedigrees, 260 pedigree-related individuals.

Usage

```
covariate_pedigree3_4
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

covariate1 A quantitative covariate.

covariate2 A qualitative covariate.

HMC_HPDI *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

```
HMC_HPDI(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.

Author(s)

Yi-Fan Kong and Ji-Yuan Zhou

Examples

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

HMC_mode *A function to obtain the mode of the samples*

Description

A function to obtain the mode of the samples.

Usage

```
HMC_mode(sam_chain)
```

Arguments

sam_chain	A vector contains the samples.

Value

The HMC_mode() returns a value.

Author(s)

Yi-Fan Kong and Ji-Yuan Zhou

Examples

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

---

mixture_data1 *A dataset containing the quantitative trait and the genotype of a SNP with no missing values for mixture data*

Description

The dataset includes 30 pedigrees, 260 pedigree-related individuals and 260 unrelated individuals.

Usage

```
mixture_data1
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

type The type of data, type=1 for pedigree data; type=2 for unrelated data.

trait A numeric variable of the quantitative trait.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

---

mixture_data2 *A dataset containing the quantitative trait and the genotype of a SNP with missing values (denoted by NA) for mixture data*

Description

The dataset includes 30 pedigrees, 260 pedigree-related individuals and 260 unrelated individuals. The missing values are denoted by NA.

Usage

```
mixture_data2
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

type The type of data, type=1 for pedigree data; type=2 for unrelated data.

trait A numeric variable of the quantitative trait.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

`mixture_data3` *A dataset containing the qualitative trait and the genotype of a SNP with no missing values for mixture data*

Description

The dataset includes 30 pedigrees, 260 pedigree-related individuals and 260 unrelated individuals.

Usage

```
mixture_data3
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

type The type of data, type=1 for pedigree data; type=2 for unrelated data.

trait A numeric variable of the affection status: 0=unaffected, 1=affected.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

`mixture_data4` *A dataset containing the qualitative trait and the genotype of a SNP with missing values (denoted by NA) for mixture data*

Description

The dataset includes 30 pedigrees, 260 pedigree-related individuals and 260 unrelated individuals. The missing values are denoted by NA.

Usage

```
mixture_data4
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

type The type of data, type=1 for pedigree data; type=2 for unrelated data.

trait A numeric variable of the affection status: 0=unaffected, 1=affected.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

`M_Bayes_XCI` *The Bayesian method for estimating the degree of the skewness of X chromosome inactivation based on mixtures of pedigree and unrelated data*

Description

This code contains the Bayesian method for estimating the degree of the skewness of X chromosome inactivation for either quantitative traits or qualitative traits, with or without covariates using mixtures of pedigree and unrelated data

Usage

```
M_Bayes_XCI(mixture_data, covariate=NULL,trait_type, trait_missing=NA,
               genotype_missing=NA, covariate_missing=NA, gamma_prior="normal",
                  prior_customize=NULL, chains_num=4, parallel_chains=4,
                iter_num=2000, warmup_num=1000, acceptance_rate=0.9, decimal=4)
```

Arguments

`mixture_data`	A data frame containing the mixtures of both pedigree and unrelated data in the form of pedigree. The first five columns of the data frame must include: famid (pedigree Id), iid (individual ID), fid (father ID), mid (mother ID) and sex. The fid and the mid of founders or unrelated individuals are both set to be 0. The numerical codes for sex are 0=unknown, 1=male, 2=female. The sixth column is type, which indicate the data type (type=1 for pedigree data; type=2 for unrelated data). The seventh column is the trait value. For qualitative traits, the numerical codes are 0=unaffected, 1=affected. The eighth column is the genotype of target SNP, which is coded as 0, 1 and 2, indicating the number of minor alleles. The details can be referred to example data "mixture_data1".
`covariate`	A data frame containing covariates (optional). The first five columns should be consistent with that in the parameter "mixture_data". The sixth to last columns are the covariates you want to add. The details can be referred to example data "covariate_mixture1_2".
`trait_type`	A character string either being "quantitative" or "qualitative", indicating the type of the trait.
`trait_missing`	The format of missing value for the trait in the parameter "mixture_data" and "covariate" (optional), and the default value is NA. It may be 9 in some data files; or other numeric values.
`genotype_missing`	The format of the missing value for the genotype in the parameter "mixture_data" and "covariate" (optional), and the default value is NA. It may be 9 in some data files; or other numeric values.
`covariate_missing`	The format of the missing value for the covariate in the parameter "covariate" (optional), and the default value is NA. It may be 9 in some data files; or other numeric values.
`gamma_prior`	A character string either being "normal", "uniform" or "customize". The "normal" represents that the prior distribution of γ is a truncated normal distribution with both parameters being 1 and the values ranging from 0 to 2; The "uniform" represents that the prior distribution of γ is the uniform distribution γ~U(0, 2); The "customize" indicates that the users could specify the prior distributions of γ and other unknown parameters according to their own research background. When "normal" or "uniform", other unknown parameters of model are set to defaults, which are according to the first article in the references.
`prior_customize`	A Stan model in Stan language, activated only when the parameter "gamma_prior" is set to "customize". You can customize the prior distributions of γ and other unknown parameters here. Please see the example 3 and example 4 for the details.
`chains_num`	A positive integer specifying the number of Markov chains. The default is 4.
`parallel_chains`	A positive integer specifying the maximum number of MCMC chains to run in parallel. The default is 4. You can set it to the maximum number of CPU cores of your computer.
`iter_num`	A positive integer specifying the number of post-warmup iterations to run per chain. The default is 2000.
`warmup_num`	A positive integer specifying the number of warmup iterations to run per chain. The default is 1000.

`acceptance_rate`
: A value between 0 and 1 that represents the target acceptance rate, the default is 0.9.

`decimal`
: A positive integer specifying the number of decimal digits remained in results

Details

Please install the "cmdstanr" package and make sure it can work before using the M_Bayes_XCI. The "cmdstanr" package cannot be installed by CRAN, and please refer to the tutorial https://mc-stan.org/cmdstanr/index.html to install. Note that we measure the degree of the skewness of X chromosome inactivation in the presence of association. So the trait and the target SNP you include in the parameter "mixture_data" should be associated. The pedigree structure in the parameter "mixture_data" must be complete, for example, even the fid and mid of a offspring are missing, the individuals corresponding them also should be included in the data as NA or other values. The results may be different for different runs, because of the sampling randomness of the HMC algorithm. If the fixed results are wanted, seed number should be set before running the function. Because cmdstanr runs HMC sampling in C language, the stan file in R language needs to be compiled into C language before each run, which may take some extra time.

Value

`Point_Estimate`
: The point estimate of the degree of the skewness of X chromosome inactivation for the target SNP based on the mixtures of pedigree and unrelated data by Bayesian method.

`HPDI_Lower`
: The lower bound of the HPDI.

`HPDI_Upper`
: The upper bound of the HPDI.

`Rhat`
: A diagnostic factor assesses convergence of Markov chains. Rhat>1.05 indicates that the Markov chain does not converge, then you should increase "iter_num" and "acceptance_rate", or reconsider prior)

Note

The interval not containing 1 indicates the skewed X chromosome inactivation (XCI-S), otherwise it suggests the random X chromosome inactivation (XCI-R) or the escape from X chromosome inactivation (XCI-E).

Author(s)

Yi-Fan Kong and Ji-Yuan Zhou

References

Yi-Fan Kong, Shi-Zhu Li, Kai-Wen Wang, Bin Zhu, Yu-Xin Yuan, Meng-Kai Li and Ji-Yuan Zhou. An Efficient Bayesian Method for Estimating the Degree of Skewness of X Chromosome Inactivation Based on General Pedigrees and Unrelated Individuals. 2022

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

Examples

```
library(cmdstanr)
library(kinship2)

##example 1:
##mixture data for quantitative trait with covariates.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data1,covariate=covariate_mixture1_2,trait_type="quantitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.3694     0.0078     0.8148 1.0006

##example 2:
##mixture data for quantitative trait with covariates.
##the prior of gamma is set to uniform distribution and other parameters are set to defaults.
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data1,covariate=covariate_mixture1_2,trait_type="quantitative",
            gamma_prior="uniform")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.3596      4e-04     0.7724 1.0006

##example 3:
##mixture data for quantitative trait with covariates.
##you want to customize the prior of gamma and other unknown parameters.
##users are required to define the prior of gamma and other parameters by "prior_customize" according to
##their own research background. The format of "prior_customize" must be in stan language, and you can refer
##to https://mc-stan.org/rstan/articles/rstan.html. We give an example as follow:
prior_customize<-"
data {                        // Define the input data type (changes are not recommended)
 int<lower=0> N ;
 int<lower=0> covariate_number ;
 vector[N] y ;
 vector[N] X0 ;
 vector[N] GRM_values ;
 matrix[N,2] X ;
 matrix[N,covariate_number] covariate ;
 }
parameters {                // Define unknown parameter type (changes are not recommended)
 real<lower=0,upper=2> gamma;
 real beta0 ;
 real beta ;
 vector [covariate_number] beta_covariate;
 real<lower=0> tao ;
 real<lower=0> sigma;
 }
transformed parameters{    // Define the transformed parameters after EVD (changes are not recommended)
 vector[N] mean;
 vector[N] std;
 mean = beta0*X0 + beta*gamma*X[,1] + beta*(2-gamma)*X[,2] + covariate*beta_covariate;
 std = sqrt(tao*GRM_values + sigma^2);
 }
model {                          // Define the prior distribution and likelihood function
 y ~ normal(mean,std);           // likelihood function (changes are not recommended)
```

```
 beta0 ~ normal (0, 100);         // prior of intercept (you can customize)
 beta ~ normal (0, 10);          // prior of regression coefficient beta (you can customize)
 beta_covariate ~ normal(0, 10); // prior of regression coefficient of covariates (you can customize)
 gamma ~ normal (1, 2);           // prior of gamma (you can customize)
 tao ~ exponential(1);         // prior of the standard deviation of polygenic effects (you can customize)
 sigma ~ exponential(1);       // prior of the standard deviation of residuals (you can customize)
}"
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data1,covariate=covariate_mixture1_2,trait_type="quantitative",
            gamma_prior="customize",prior_customize=prior_customize)
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.3405      8e-04     0.7925 1.0006

##example 4:
##mixture data for qualitative trait with covariates.
##you want to customize the prior of gamma and other unknown parameters.
##the format of "prior_customize" between quantitative and qualitative traits are little different.
##We give the example for qualitative traits as follow:
prior_customize<-"
data {                          // Define the input data type (changes are not recommended)
 int<lower=0> N ;
 array[N] int y ;
 int<lower=0> covariate_number ;
 matrix[N,2] X;
 matrix[N,N] GRM_female;
 matrix[N,covariate_number] covariate ;
}
transformed data {       // Define the input data after Cholesky (changes are not recommended)
 matrix[N,N] C;
 C = cholesky_decompose(GRM_female);
}
parameters {                 // Define unknown parameter type (changes are not recommended)
 real<lower=0,upper=2> gamma;
 real beta0 ;
 real beta ;
 real<lower=0> tao ;
 vector [covariate_number] beta_covariate;
 vector[N] z;
}
model {                         // Define the prior distribution and likelihood function
 vector[N] p;
 vector[N] d;
 z ~ normal (0,1);
 d = tao *( C * z );
 p = beta0 + beta*gamma*X[,1] + beta*(2-gamma)*X[,2] + covariate*beta_covariate + d;
 y ~ bernoulli_logit(p);          // likelihood function (changes are not recommended)
 beta0 ~ normal(0, 100);          // prior of intercept (you can customize)
 beta ~ normal(0, 10);          // prior of regression coefficient beta (you can customize)
 beta_covariate ~ normal(0, 10); // prior of regression coefficient of covariates (you can customize)
 gamma ~ normal(1, 2);            // prior of gamma (you can customize)
 tao ~ exponential(1);         // prior of the standard deviation of polygenic effects (you can customize)
}"
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data3,covariate=covariate_mixture3_4,trait_type="qualitative",
            gamma_prior="customize",prior_customize=prior_customize)
#results:
```

```
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs234567         1.2507     0.7905     1.9196 1.0011

##example 5:
##mixture data for quantitative trait with covariates and missing values.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data2,covariate=covariate_mixture1_2,trait_type="quantitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.4527     0.0079     1.0871 1.0006

##example 6:
##mixture data for qualitative trait with covariates and missing values.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data4,covariate=covariate_mixture3_4,trait_type="qualitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs234567         1.3081     0.7731     1.9549 1.0005

##example 7:
##mixture data for quantitative trait without covariates.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data1,covariate=NULL,trait_type="quantitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.3941     0.0074     0.8238 1.0007

##example 8:
##mixture data for qualitative trait without covariates.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
M_Bayes_XCI(mixture_data=mixture_data3,covariate=NULL,trait_type="qualitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs234567         1.2463     0.7823     1.8924 1.0006
```

`pedigree_data1` *A dataset containing the quantitative trait and the genotype of a SNP with no missing values for only pedigree data*

Description

The dataset includes 30 pedigrees and 260 pedigree-related individuals.

Usage

```
pedigree_data1
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

trait A numeric variable of the quantitative trait.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

`pedigree_data2` *A dataset containing the quantitative trait and the genotype of a SNP with missing values (denoted by NA) for only pedigree data*

Description

The dataset includes 30 pedigrees and 260 pedigree-related individuals. The missing values are denoted by NA.

Usage

```
pedigree_data2
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

trait A numeric variable of the quantitative trait.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

pedigree_data3 *A dataset containing the qualitative trait and the genotype of a SNP with no missing values for only pedigree data*

Description

The dataset includes 30 pedigrees and 260 pedigree-related individuals.

Usage

```
pedigree_data3
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

trait A numeric variable of the affection status: 0=unaffected, 1=affected.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

pedigree_data4 *A dataset containing the qualitative trait and the genotype of a SNP with missing values (denoted by NA) for only pedigree data*

Description

The dataset includes 30 pedigrees and 260 pedigree-related individuals. The missing values are denoted by NA.

Usage

```
pedigree_data4
```

Format

famid Pedigree ID.

iid Individual ID.

fid 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.

trait A numeric variable of the affection status: 0=unaffected, 1=affected.

genotype The genotype of the target SNP, coded as 0, 1 or 2, indicating the number of the minor alleles.

P_Bayes_XCI *The Bayesian method for estimating the degree of the skewness of X chromosome inactivation based on pedigree data*

Description

This code contains the Bayesian method for estimating the degree of the skewness of X chromosome inactivation for either quantitative traits or qualitative traits, with or without covariates using only pedigree data

Usage

```
P_Bayes_XCI(pedigree_data, covariate=NULL, trait_type, trait_missing=NA,
               genotype_missing=NA, covariate_missing=NA, gamma_prior="normal",
                  prior_customize=NULL, chains_num=4, parallel_chains=4,
                iter_num=2000, warmup_num=1000, acceptance_rate=0.9, decimal=4)
```

Arguments

pedigree_data	A data frame containing the pedigree data. The first five columns of the data frame must include: famid (pedigree Id), iid (individual ID), fid (father ID), mid (mother ID) and sex. The numerical codes for sex are 0=unknown, 1=male, 2=female. The sixth column is the trait value. For qualitative traits, the numerical codes are 0=unaffected, 1=affected. The seventh column is the genotype of target SNP, which is coded as 0, 1 and 2, indicating the number of minor alleles. The details can be referred to example data "pedigree_data1".
covariate	A data frame containing covariates (optional). The first five columns should be consistent with that in the parameter "pedigree_data". The sixth to last columns are the covariates you want to add. The details can be referred to example data "covariate_pedigree1_2".
trait_type	A character string either being "quantitative" or "qualitative", indicating the type of the trait.
trait_missing	The format of missing value for the trait in the parameter "pedigree_data" and "covariate" (optional), and the default value is NA. It may be 9 in some data files; or other numeric values.
genotype_missing	The format of missing value for the genotype in the parameter "pedigree_data" and "covariate" (optional), and the default value is NA. It may be 9 in some data files; or other numeric values.
covariate_missing	The format of missing value for the covariate in the parameter "covariate" (optional), and the default value is NA. It may be 9 in some data files; or other numeric values.
gamma_prior	A character string either being "normal", "uniform" or "customize". The "normal" represents that the prior distribution of γ is a truncated normal distribution with both parameters being 1 and the values ranging from 0 to 2; The "uniform" represents that the prior distribution of γ is the uniform distribution γ~U(0, 2); The "customize" indicates that the users could specify the prior distributions of γ and other unknown parameters according to their own research background.

When "normal" or "uniform", other unknown parameters of model are set to defaults, which are according to the first article in the references.

`prior_customize`
: A Stan model in Stan language, activated only when the parameter "gamma_prior" is set to "customize". You can customize the prior distributions of γ and other unknown parameters here. Please see the example 3 and example 4 for the details.

`chains_num`
: A positive integer specifying the number of Markov chains. The default is 4.

`parallel_chains`
: A positive integer specifying the maximum number of MCMC chains to run in parallel. The default is 4. You can set it to the maximum number of CPU cores of your computer.

`iter_num`
: A positive integer specifying the number of post-warmup iterations to run per chain. The default is 2000.

`warmup_num`
: A positive integer specifying the number of warmup iterations to run per chain. The default is 1000.

`acceptance_rate`
: A value between 0 and 1 that represents the target acceptance rate, the default is 0.9.

`decimal`
: A positive integer specifying the number of decimal digits remained in results

Details

Please install the "cmdstanr" package and make sure it can work before using the P_Bayes_XCI. The "cmdstanr" package cannot be installed by CRAN, and please refer to the tutorial https://mc-stan.org/cmdstanr/index.html to install. Note that we measure the degree of the skewness of X chromosome inactivation in the presence of association. So the trait and the target SNP you include in the parameter "pedigree_data" must be associated. The pedigree structure in the parameter "pedigree_data" should be complete, for example, even the fid and mid of a offspring are missing, the individuals corresponding them also should be included in the data as NA or other values. The results may be different for different runs, because of the sampling randomness of the HMC algorithm. If the fixed results are wanted, seed number should be set before running the function. Because cmdstanr runs HMC sampling in C language, the stan file in R language needs to be compiled into C language before each run, which may take some extra time.

Value

`Point_Estimate`
: The point estimate of the degree of the skewness of X chromosome inactivation for the target SNP based on the pedigree data by Bayesian method.

`HPDI_Lower`
: The lower bound of the HPDI.

`HPDI_Upper`
: The upper bound of the HPDI.

`Rhat`
: A diagnostic factor assesses convergence of Markov chains. Rhat>1.05 indicates that the Markov chain does not converge, then you should increase "iter_num" and "acceptance_rate", or reconsider prior)

Note

The interval not containing 1 indicates the skewed X chromosome inactivation (XCI-S), otherwise it suggests the random X chromosome inactivation (XCI-R) or the escape from X chromosome inactivation (XCI-E).

Author(s)

Yi-Fan Kong and Ji-Yuan Zhou

References

Yi-Fan Kong, Shi-Zhu Li, Kai-Wen Wang, Bin Zhu, Yu-Xin Yuan, Meng-Kai Li and Ji-Yuan Zhou. An Efficient Bayesian Method for Estimating the Degree of Skewness of X Chromosome Inactivation Based on General Pedigrees and Unrelated Individuals. 2022

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

Examples

```
library(cmdstanr)
library(kinship2)

##example 1:
##pedigree data for quantitative trait with covariates.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
P_Bayes_XCI(pedigree_data=pedigree_data1,covariate=covariate_pedigree1_2,trait_type="quantitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.6351          0     1.4811 1.0004

##example 2:
##pedigree data for quantitative trait with covariates.
##the prior of gamma is set to uniform distribution and other parameters are set to defaults.
set.seed(123456)
M_Bayes_XCI(pedigree_data=pedigree_data1,covariate=covariate_pedigree1_2,trait_type="quantitative",
            gamma_prior="uniform")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.5972          0     1.5246 1.0016

##example 3:
##pedigree data for quantitative trait with covariates.
##you want to customize the prior of gamma and other unknown parameters.
##users are required to define the prior of gamma and other parameters by "prior_customize".
##The example of "prior_customize" is same as that in example 3 of function "M_Bayes_XCI".
set.seed(123456)
P_Bayes_XCI(pedigree_data=pedigree_data1,covariate=covariate_pedigree1_2,trait_type="quantitative",
            gamma_prior="customize",prior_customize=prior_customize)
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.5728     0.0018     1.4738 1.0006

##example 4:
##pedigree data for qualitative trait with covariates.
##you want to customize the prior of gamma and other unknown parameters.
##users are required to define the prior of gamma and other parameters by "prior_customize".
##The example of "prior_customize" is same as that in example 4 of function "M_Bayes_XCI".
set.seed(123456)
P_Bayes_XCI(pedigree_data=pedigree_data1,covariate=covariate_pedigree1_2,trait_type="qualitative",
            gamma_prior="customize",prior_customize=prior_customize)
```

```
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs234567         1.1429     0.5193     1.9999 1.0001

##example 5:
##pedigree data for quantitative trait with covariates and missing values.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
P_Bayes_XCI(pedigree_data=pedigree_data2,covariate=covariate_pedigree1_2,trait_type="quantitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.9163     0.1115     1.9141 1.0001

##example 6:
##pedigree data for qualitative trait with covariates and missing values.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
P_Bayes_XCI(pedigree_data=pedigree_data4,covariate=covariate_pedigree3_4,trait_type="qualitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs234567         1.1992     0.5012     1.9757 1.0004

##example 7:
##pedigree data for quantitative trait without covariates.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
P_Bayes_XCI(pedigree_data=pedigree_data1,covariate=NULL,trait_type="quantitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs123456         0.6802     0.0202     1.4178 1.0011

##example 8:
##pedigree data for qualitative trait without covariates.
##the prior of gamma is set to truncated normal distribution and other parameters are set to defaults.
set.seed(123456)
P_Bayes_XCI(pedigree_data=pedigree_data3,covariate=NULL,trait_type="qualitative",
            gamma_prior="normal")
#results:
#SNP_Name Point_Estimate HPDI_Lower HPDI_Upper   Rhat
#rs234567         1.1201     0.5049     1.9841 1.0021
```

Index