Package 'BEXCIS'

January 10, 2022

Type Package

Title Measuring the degree of the skewness of the X chromosome inactivation by the Bayesian method, the penalized Fieller's method and the Fieller's method.

Version 1.0

Date 2022-01-10

Author Wen-Yi Yu and Ji-Yuan Zhou

Maintainer Wen-Yi Yu <wenyiyu2019@163.com> and Ji-Yuan Zhou <zhoujiyuan5460@hotmail.com>

Description This code contains the Bayesian method, the penalized Fieller's method and the Fieller's method for measuring the degree of the skewness of the X chromosome inactivation for either quantitative trait or qualitative trait, with or without covariates using unrelated female individuals.

License GPL-2

R topics documented:

`BEXCIS-package` *Measuring the degree of the skewness of the X chromosome inactivation by the Bayesian method, the penalized Fieller's method and the Fieller's method.*

Description

This code contains the Bayesian method, the penalized Fieller's method and the Fieller's method for measuring the degree of the skewness of the X chromosome inactivation for either quantitative trait or qualitative trait, with or without covariates using unrelated female individuals.

Author(s)

Wen-Yi Yu and Ji-Yuan Zhou

Maintainer: Wen-Yi Yu <wenyiyu2019@163.com> and Ji-Yuan Zhou <zhoujiyuan5460@hotmail.com>

References

Wen-Yi Yu, Yu Zhang, Meng-Kai Li, Zi-Ying Yang, Wing Kam Fung, Pei-Zhen Zhao and Ji-Yuan Zhou. BEXCIS: Bayesian methods for estimating the degree of the skewness of X chromosome inactivation. 2022

Annis J, Miller BJ, Palmeri TJ. Bayesian inference with Stan: a tutorial on adding custom distributions. Behav Res Methods 2017;49:863-86.

Wang P, Xu SQ, Wang YX, et al. Penalized Fieller's confidence interval for the ratio of bivariate normal means. Biometrics 2020;1-14.

Wang P, Zhang Y, Wang BQ, et al. A statistical measure for the skewness of X chromosome inactivation based on case-control design. BMC Bioinformatics 2019;20(1):11.

Li BH, Yu WY, Zhou JY. A statistical measure for the skewness of X chromosome inactivation for quantitative traits and its application to the MCTFR data. BMC Genomic Data 2021;22(1):24.

`Bayes_XCI` *The Bayesian method for measuring the degree of the skewness of X chromosome inactivation.*

Description

This code contains the Bayesian method for measuring the degree of the skewness of X chromosome inactivation for either quantitative trait or qualitative trait, with or without covariates using unrelated female individuals.

Usage

```
Bayes_XCI(ped, trait_missing = NA, allele_missing = 0,
ped.header = F, trait_type = "quantitative",
covariate = NULL, covariate_prior = NULL,
covariate_missing = NA, covariate_prior_limit = NULL,
covariate.header = F, gamma_prior = "normal",
chains_num = 8, total_sample = 20000,
warmup_sample = 10000, adapt_delta_value = 0.99)
```

Arguments

`ped` The name of a standard linkage pedigree file or a matrix/dataframe containing the data. The pedigree file contains the following information: pedigree ID (famid), individual ID (iid), father's ID (fid), 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. There should be two columns for each SNP respectively representing two alleles. Numeric coding for genotype is 0=unknown, 1 = allele 1, 2 = allele 2. 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. The male individuals are hemizygote at all the SNPs on X chromosome and the second allele should be the same as the first allele at each SNP.

`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.

`allele_missing`
: The input variable "allele_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.

`ped.header`
: Logical scalar defaulting to False (or F) indicating whether the ped file contains variable names or not.

`trait_type`
: A character string either being "quantitative" or "qualitative", the default value is "quantitative".

`covariate`
: The covariates needed to be adjusted, can be a txt file or a dataframe/matrix, the first five columns should be pedigree ID (famid), individual ID (iid), father's ID (fid), mother's ID (mid) and sex.

`covariate_prior`
: A matrix of two columns with each row providing the prior distribution of the effect size for each covariate. A string describing the prior distribution should be given in the first column e.g., "normal", "uniform" and "exponential", and a string containing the parameters required to specify the distribution should be given in the second column e.g., "1,1" for normal distribution N(1,1). Note that the order of the prior should be the same as that of the covariate file or dataframe/matrix.

`covariate_missing`
: The input variable "covariate_missing" is the missing value for the covariates in the data file, and the default value is NA.

`covariate_prior_limit`
: A matrix of two columns with each row specifying the minimum and the maximum of the effect size of each covariate. The minimum and the maximum should be respectively given with two strings in the first and second columns, e.g., matrix(c('0', '1'), , 2). If there is no limitation of the effect size of the covariate or only the minimum or the maximum is available, the minimum and the maximum can be set to NA, e.g., rbind(c(NA, NA),c('0', NA)). Note that the order of the covariate_prior_limit should be the same as that of the covariate file or dataframe/matrix.

`covariate.header`
: Logical scalar defaulting to False (or F) indicating whether the covariate file contains variable names or not.

`gamma_prior`
: A character string either being 'normal' or 'uniform', which represents the prior distribution of the degree of the skewness of the X chromosome inactivation.

`chains_num`
: A positive integer specifying the number of Markov chains. The default is 8.

`total_sample`
: A positive integer specifying the number of iterations for each chain (including warmup). The default is 20000.

`warmup_sample`
: A positive integer specifying the number of warmup (also known as burnin) iterations per chain. If step-size adaptation is on (which is by default), this also controls the number of iterations for which adaptation is run (and hence these

warmup samples should not be used for inference). The number of warmup iterations should be smaller than the number of iterations and the default is 10000.

`adapt_delta_value`
: A value between 0 and 1 that represents the target acceptance rate, the default is 0.99.

Details

Please install the rstan package and make sure it can work before using the Bayes_XCI. Note that we measure the degree of the skewness of X chromosome inactivation in the presence of association. In case of multiple markers on X chromosome which are associated with the trait, each marker will be analyzed separately. 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. 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 4.1.0.

Value

`Point_Estimate`
: The point estimate of the degree of the skewness of X chromosome inactivation for the marker based on the Bayesian method.

`HPDI_Lower`
: The lower bound of the estimated interval.

`HPDI_Upper`
: The upper bound of the estimated interval.

Warning

The following warning is normal and can be ignored:

Warning message: In system(paste(CXX, ARGS), ignore.stdout = TRUE, ignore.stderr = TRUE) : 'C:/rtools40/usr/mingw_/bin/g++' not found

Errors

If the following error is returned:

[1] "Error in sampler$call_sampler(args_list[[i]]) : Initialization failed."

[1] "error occurred during calling the sampler; sampling not done".

Then, the value limits set for the coefficients of the covariates may be inappropriate.

If the following error is returned:

Error in h(simpleError(msg, call)) :

error in evaluating the argument 'object' in selecting a method for function 'extract': object 'Fit_H1' not found.

Then, the prior distributions set for the coefficients of the covariates may be inappropriate.

Note

The interval not containing 1 indicates skewed X chromosome inactivation (XCI), otherwise it suggests random XCI. When the skewed XCI is detected at a marker, the ratio (denoted by a) of the point estimate to 2 indicates the proportion of the cells in a heterozygous female which express allele 2 and hence (1-a) represents the the proportion of the cells in a heterozygous female which express allele 1.

Author(s)

Wen-Yi Yu and Ji-Yuan Zhou

References

Wen-Yi Yu, Yu Zhang, Meng-Kai Li, Zi-Ying Yang, Wing Kam Fung, Pei-Zhen Zhao and Ji-Yuan Zhou. BEXCIS: Bayesian methods for estimating the degree of the skewness of X chromosome inactivation. 2022

Annis J, Miller BJ, Palmeri TJ. Bayesian inference with Stan: a tutorial on adding custom distributions. Behav Res Methods 2017;49:863-86.

Examples

```
##example 1:
#the ped file without header
#set "ped='ped_qualitative.txt'" and "ped.header=F"

#the trait is qualitative
#set "trait_type='qualitative'"

#no covariate
#set "covariate=NULL"

#using uniform distribution as the prior distribution for gamma
#set "gamma_prior='uniform'"

#set seed to get the fixed result
  set.seed(123)
  example1<-Bayes_XCI(ped='ped_qualitative.txt', trait_missing=9,
allele_missing=0, ped.header=F, trait_type='qualitative',
covariate=NULL, covariate_prior=NULL,
covariate_missing=NA, covariate.header=F,
covariate_prior_limit=NULL, gamma_prior='uniform',
chains_num=2, total_sample=12000,
warmup_sample=2000, adapt_delta_value=0.85)

#Although the number of Markov chains, the number of iterations for each
#chain, the number of warmup iterations and the target acceptance rate
#are respectively set to 2, 12000, 2000 and 0.85 in this example, we recommend
#respectively using 8, 20000, 10000 and 0.99.

#Result
  write.table(example1,"example1.txt",row.names=F,quote=F)
  print(example1)

#      Point_Estimate HPDI_Lower HPDI_Upper
#snp_1       1.715434  0.1687426   1.999994



##example 2:
#the ped file with header
#set "ped='ped_qualitative_header.txt'" and "ped.header=T"

#the trait is qualitative
#set "trait_type='qualitative'"
```

```
#the covariate file without header
#set "covariate='covariate.txt'" and "covariate.header=F"

#using normal distribution as the prior distribution for gamma
#set "gamma_prior='normal'"

#define the prior distributions of the coefficients of the covariates
  covariate_prior<-rbind(c('uniform','-2,2'),c('normal','0,10'))

#define the value limits of the coefficients of the covariates
  covariate_prior_limit <- rbind(c(NA,'2'),c(NA,NA))

#set seed to get the fixed result
  set.seed(123)
  example2<-Bayes_XCI(ped='ped_qualitative_header.txt',
trait_missing=9, allele_missing=0,
ped.header=T, trait_type='qualitative',
covariate='covariate.txt', covariate_prior=covariate_prior,
covariate_missing=NA, covariate.header=F,
covariate_prior_limit=covariate_prior_limit,
gamma_prior='normal', chains_num=2,
total_sample=12000, warmup_sample=2000,
adapt_delta_value=0.85)

#Although the number of Markov chains, the number of iterations for each
#chain, the number of warmup iterations and the target acceptance rate
#are respectively set to 2, 12000, 2000 and 0.85 in this example, we recommend
#respectively using 8, 20000, 10000 and 0.99.

#Result
  write.table(example2,"example2.txt",row.names=F,quote=F)
  print(example2)

#        Point_Estimate HPDI_Lower HPDI_Upper
#rs001_1       1.601871  0.2058925   1.999482



##example 3:
#the ped file with header
#set "ped='ped_quantitative_header.txt'" and "ped.header=T"

#the trait is quantitative
#set "trait_type='quantitative'"

#the covariate file with header
#set "covariate='covariate_header.txt'" and "covariate.header=T"

#using uniform distribution as the prior distribution for gamma
#set "gamma_prior='uniform' "

#define the prior distributions of the coefficients of the covariates
  covariate_prior<-rbind(c('uniform','-2,2'),c('normal','0,10'))

#define the value limits of the coefficients of the covariates
```

```
  covariate_prior_limit <- rbind(c(NA,'2'),c(NA,NA))

#set seed to get the fixed result
  set.seed(123)
  example3<-Bayes_XCI(ped='ped_quantitative_header.txt',
trait_missing=NA, allele_missing=0, ped.header=T,
trait_type='quantitative', gamma_prior='uniform',
covariate='covariate_header.txt',
covariate_missing=NA,
covariate_prior=covariate_prior,
covariate.header=T,
covariate_prior_limit=covariate_prior_limit,
chains_num=2, total_sample=12000,
warmup_sample=2000, adapt_delta_value=0.85)

#Although the number of Markov chains, the number of iterations for each
#chain, the number of warmup iterations and the target acceptance rate
#are respectively set to 2, 12000, 2000 and 0.85 in this example, we recommend
#respectively using 8, 20000, 10000 and 0.99.

#Result
  write.table(example3,"example3.txt",row.names=F,quote=F)
  print(example3)

#        Point_Estimate HPDI_Lower HPDI_Upper
#rs001_1       1.417922  0.3589927   1.999366



##example 4:
#the ped dataframe with colnames
#set "ped=ped"
  ped<-read.table('ped_quantitative_header.txt',header=T)

#the trait is quantitative
#set "trait_type='quantitative'"

#the covariate dataframe with colnames
#set "covariate=covariate"
  covariate<-read.table('covariate_header.txt',header=T)

#using uniform distribution as the prior distribution for gamma
#set "gamma_prior='uniform'"

#define the prior distributions of the coefficients of the covariates
  covariate_prior<-rbind(c('uniform','-4,4'),c('normal','0,10'))

#if no constraint for the coefficients of the two covariates
  covariate_prior_limit <- rbind(c(NA,NA),c(NA,NA))

#set seed to get the fixed result
  set.seed(123)
  example4<-Bayes_XCI(ped=ped, trait_missing=NA,
allele_missing=0,
ped.header=F, trait_type='quantitative',
covariate=covariate,
```

```
covariate_prior=covariate_prior,
covariate_missing=NA, covariate.header=F,
covariate_prior_limit=covariate_prior_limit,
gamma_prior='uniform', chains_num=2,
total_sample=12000, warmup_sample=2000,
adapt_delta_value=0.85)

#Although the number of Markov chains, the number of iterations for each
#chain, the number of warmup iterations and the target acceptance rate
#are respectively set to 2, 12000, 2000 and 0.85 in this example, we recommend
#respectively using 8, 20000, 10000 and 0.99.

#Result
  write.table(example4,"example4.txt",row.names=F,quote=F)
  print(example4)

#setting different prior distributions and different limits
#for the coefficients of the covariates affects the results

#        Point_Estimate HPDI_Lower HPDI_Upper
#rs001_1       1.458581  0.3624141   1.999038
```

`Frequen_XCI` *The penalized Fieller's method and the Fieller's method for measuring the degree of the skewness of X chromosome inactivation.*

Description

This code contains the penalized Fieller's method and the Fieller's method for measuring the degree of the skewness of X chromosome inactivation for either quantitative trait or qualitative trait, with or without covariates using unrelated female individuals.

Usage

```
Frequen_XCI(ped, trait_missing = NA, allele_missing = 0,
ped.header = F, trait_type = "quantitative",
covariate = NULL, covariate.header = F,
covariate_missing = NA)
```

Arguments

`ped`	The name of a standard linkage pedigree file or a matrix/dataframe containing the data. The pedigree file contains the following information: pedigree ID (famid), individual ID (iid), father's ID (fid), 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. There should be two columns for each SNP respectively representing two alleles. Numeric coding for genotype is 0=unknown, 1 = allele 1, 2 = allele 2. 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. The male individuals are hemizygote at all the SNPs on X chromosome and the second allele should be the same as the first allele at each SNP.

`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.

`allele_missing`
: The input variable "allele_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.

`ped.header`
: Logical scalar defaulting to False (or F) indicating whether the ped file contains variable names or not.

`trait_type`
: A character string either being "quantitative" or "qualitative", the default value is "quantitative".

`covariate`
: The covariates needed to be adjusted, can be a txt file or a dataframe/matrix, the first five columns should be pedigree ID (famid), individual ID (iid), father's ID (fid), mother's ID (mid) and sex.

`covariate.header`
: Logical scalar defaulting to False (or F) indicating whether the covariate file contains variable names or not.

`covariate_missing`
: The input variable "covariate_missing" is the missing value for the covariates in the data file, and the default value is NA.

Details

Please install the rootSolve package and make sure it can work before using the Frequen_XCI. Note that we measure the degree of the skewness of X chromosome inactivation in the presence of association. In case of multiple markers on X chromosome which are associated with the trait, each marker will be analyzed separately.

Value

`F_Point_Estimate`
: The point estimate of the degree of the skewness of X chromosome inactivation for the marker obtained by the Fieller's method.

`F_Lower`
: The lower bound of the estimated interval by the Fieller's method.

`F_Upper`
: The upper bound of the estimated interval by the Fieller's method.

`F_discontinuous`
: A value indicates whether or not the interval obtained by the Fieller's method is continuous or discontinuous. F_discontinuous = 0 means the interval is continuous and F_discontinuous = 1 means the interval is discontinuous.

`PF_Point_Estimate`
: The penalized point estimate of the degree of the skewness of X chromosome inactivation for the marker obtained by the penalized Fieller's method.

`PF_Lower`
: The lower bound of the estimated interval of the penalized Fieller's method.

`PF_Upper`
: The upper bound of the estimated interval of the penalized Fieller's method.

Note

The interval not containing 1 indicates skewed X chromosome inactivation (XCI), otherwise it suggests random XCI. When the skewed XCI is detected at a marker, the ratio (denoted by a) of the point estimate to 2 indicates the proportion of the cells in a heterozygous female which express allele 2 and hence (1-a) represents the the proportion of the cells in a heterozygous female which express allele 1.

Author(s)

Wen-Yi Yu and Ji-Yuan Zhou

References

Wen-Yi Yu, Yu Zhang, Meng-Kai Li, Zi-Ying Yang, Wing Kam Fung, Pei-Zhen Zhao and Ji-Yuan Zhou. BEXCIS: Bayesian methods for estimating the degree of the skewness of X chromosome inactivation. 2022

Wang P, Xu SQ, Wang YX, et al. Penalized Fieller's confidence interval for the ratio of bivariate normal means. Biometrics 2020;1-14.

Wang P, Zhang Y, Wang BQ, et al. A statistical measure for the skewness of X chromosome inactivation based on case-control design. BMC Bioinformatics 2019;20(1):11.

Li BH, Yu WY, Zhou JY. A statistical measure for the skewness of X chromosome inactivation for quantitative traits and its application to the MCTFR data. BMC Genomic Data 2021;22(1):24.

Examples

```
##example 5:
#the ped file without header
#set "ped='ped_qualitative.txt'" and "ped.header=F"

#no covariate
#set "covariate=NULL"

#the trait is qualitative
#set "trait_type='qualitative'"

  example5<-Frequen_XCI(ped='ped_qualitative.txt',
trait_missing=9,
allele_missing=0, ped.header=F,
trait_type='qualitative', covariate=NULL,
covariate.header=F, covariate_missing=NA)


#Result
  write.table(example5,"example5.txt",row.names=F,quote=F)
  print(example5)

#       F_Point_Estimate    F_Lower    F_Upper    F_discontinuous
#snp_1       2               0          2            0
#     PF_Point_Estimate    PF_Lower    PF_Upper
#snp_1        1.756293       0          2



##example 6:
#the ped file without header
#set "ped='ped_qualitative.txt'" and "ped.header=F"

#the covariate file with header
#set "covariate='covariate_header.txt'" and "covariate.header=T"

#the trait is qualitative
#set "trait_type='qualitative'"
```

```
  example6<-Frequen_XCI(ped='ped_qualitative.txt',
trait_missing=9, allele_missing=0,
ped.header=F, trait_type='qualitative',
covariate='covariate_header.txt',
covariate.header=T, covariate_missing=NA)


#Result
  write.table(example6,"example6.txt",row.names=F,quote=F)
  print(example6)

#       F_Point_Estimate    F_Lower    F_Upper    F_discontinuous
#snp_1       2                  0           2            0
#          PF_Point_Estimate    PF_Lower    PF_Upper
#snp_1       2                  0           2




##example 7:
#the ped dataframe without colnames
  ped<-read.table('ped_quantitative.txt',header=F)

#the trait is quantitative
#set "trait_type='quantitative'"

#the covariate dataframe without colnames
  covariate<-read.table('covariate.txt',header=F)

  example7<-Frequen_XCI(ped=ped, trait_missing=NA,
allele_missing=0, ped.header=F,
trait_type='quantitative', covariate=covariate,
covariate.header=F, covariate_missing=NA)

#Result
  write.table(example7,"example7.txt",row.names=F,quote=F)
  print(example7)

#       F_Point_Estimate    F_Lower    F_Upper    F_discontinuous
#snp_1       1.271208          0          2            0
#       PF_Point_Estimate    PF_Lower    PF_Upper
#snp_1       1.239847          0          2
```

`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 that specifies the proportion of samples that should be included in an interval, the default is 0.95.

Value

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

Author(s)

Wen-Yi Yu and Ji-Yuan Zhou

Examples

```
##The highest posterior density interval
# including 95% of samples
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(samples)
```

Arguments

samples	A vector providing the samples.

Value

The modeofHMC() returns a value.

Author(s)

Wen-Yi Yu and Ji-Yuan Zhou

Examples

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

`PenFieller` *A function for obtaining the confidence interval of a ratio by the penalized Fieller's method.*

Description

This function is used to calculate the upper and lower limits of the penalized Fieller's confidence interval for the ratio estimate.

Usage

```
PenFieller(mu_n, mu_d, var_n, var_d,
rho = 0, df_n = NULL, df_d = NULL,
con_level = 0.95)
```

Arguments

`mu_n`	The estimated mean of the numerator.
`mu_d`	The estimated mean of the denominator.
`var_n`	A positive value gives the estimated variance of the numerator.
`var_d`	A positive value gives the estimated variance of the denominator.
`rho`	A value between -1 and 1 that represents the estimated correlation coefficient of the numerator and the denominator.
`df_n`	The degree of freedom of the numerator. The default value is NULL.
`df_d`	The degree of freedom of the denominator. The default value is NULL.
`con_level`	The confidence level. Should be between 0 and 1. The default is 0.95.

Value

`point_res`	The penalized point estimate of the ratio.
`PenFieller_Lower`	The lower bound of the estimated interval of the penalized Fieller's method.
`PenFieller_Upper`	The upper bound of the estimated interval of the penalized Fieller's method.

Author(s)

Peng Wang

References

Wang P, Xu SQ, Wang YX, et al. Penalized Fieller's confidence interval for the ratio of bivariate normal means. Biometrics 2020;1-14.