Package ‘BMRV’
March 24, 2015

Type Package

Title Bayesian models for rare variant association detection

Version 1.1

Date 2015-03-22

Author Liang He

Maintainer Liang He <liang.he@helsinki.fi>

Description This package provides two methods for detecting the association between rare variants and continuous traits. One of them detects interaction effect and can be applied to twin studies. The other incorporates genotype uncertainty information.

License None

Archs i386, x64

**R topics documented:**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMRV-package</td>
<td>1</td>
</tr>
<tr>
<td>blvcm</td>
<td>2</td>
</tr>
<tr>
<td>blvcm_bin</td>
<td>4</td>
</tr>
<tr>
<td>blvcm_bin_data</td>
<td>5</td>
</tr>
<tr>
<td>blvcm_data</td>
<td>6</td>
</tr>
<tr>
<td>hbmr</td>
<td>6</td>
</tr>
<tr>
<td>hbmr_data</td>
<td>7</td>
</tr>
</tbody>
</table>

Index

Description

This package provides two methods for detecting the association between rare variants and continuous binary traits. One of them detects interaction effect and can be applied to twin studies. The other incorporates genotype uncertainty information.

Details
Package: BMRV
Type: Package
Version: 1.1
Date: 2015-03-22
License: None

Author(s)
Liang He
Maintainer: Liang He <liang.he@helsinki.fi>

References

Examples
data(blvcm_data)
temp <- blvcm(blvcm_data$pheno_data, blvcm_data$geno_data, iter=20000, model = 3)

blvcm Bayesian latent variable collapsing model (BLVCM)

Description
The function implements BLVCM using Gibbs sampling method.

Usage
blvcm(pheno, geno, model = 3, iter = 30000, burnin = 500, var = -1, lambda = -1, cov = 0, init = c(0.5,0.5))

Arguments
pheno N x 3 Phenotypic data matrix (trait, family number, zyg), where N is the number of subjects. Please see the example data for more details. For faster convergence, it is recommended that phenotype should be standardized.
geno N x K Genotypic data matrix, where N is the number of subjects and K is the number of rare variants. The value can be 0 or 1. A missing genotype is represented by -9, which will be imputated by BLVCM based on HWE.
model Twin model: 3 for ACE model, 2 for AE model, 1 for independent subjects
iter The number of MCMC iterations
**blvcm**

- **burnin**: The number of burn-in.
- **var**: variance hyperparameter in the priors for beta and gamma. The default value is the variance of the phenotype.
- **lambda**: threshold lambda for hypothesis test. The default value is 0.2.
- **cov**: covariate matrix
- **init**: Initial values for beta and gamma. The default values are 0.5 and 0.

**Value**

- **BF_main**: The Bayes factor of the main effect
- **BF_int**: The Bayes factor of the interaction effect
- **post_odds_beta**: The posterior odds of beta
- **post_odds_gamma**: The posterior odds of gamma
- **com_a**: The inverse of the posterior mean of the precision for additive genetic component
- **com_c**: The inverse of the posterior mean of the precision for shared environmental component
- **mean_mu**: The posterior mean of intercept
- **mean_beta**: The posterior mean of beta
- **mean_gamma**: The posterior mean of gamma
- **sd_mu**: The posterior standard deviation of intercept
- **sd_beta**: The posterior standard deviation of beta
- **sd_gamma**: The posterior standard deviation of gamma
- **mean_rv**: The posterior mean of alphas
- **prior_var**: The variance hyperparameter in the priors for beta and gamma

**Author(s)**

Liang He

**References**


**Examples**

data(blvcm_data)
blvcm(blvcm_data$pheno_data, blvcm_data$geno_data, iter=20000, burnin=1000, model=3)
blvcm_bin  Bayesian latent variable collapsing model (BLVCM) for binary data with probit link

Description

The function implements BLVCM for binary traits using Gibbs sampling method with probit link function.

Usage

blvcm_bin(pheno, geno, model = 3, iter = 30000, burnin = 500, var = -1, lambda = -1, cov = 0, init = c(0.5, 0))

Arguments

pheno  N x 3 Phenotypic data matrix (trait, family number, zyg), where N is the number of subjects. The trait must be 0 or 1.
geno  N x K Genotypic data matrix, where N is the number of subjects and K is the number of rare variants. The value can be 0 or 1. A missing genotype is represented by -9, which will be imputed by BLVCM based on HWE.
model  Twin model: 3 for ACE model, 2 for AE model, 1 for independent subjects
iter  The number of MCMC iterations. The default value is 30000.
burnin  The number of burn-in. The default value is 500.
var  variance hyperparameter in the priors for beta and gamma. The default value is 1.
lambda  threshold lambda for hypothesis test. The default value is 0.2.
cov  covariate matrix
init  Initial values for beta and gamma. The default values are 0.5 and 0.

details

The Gibbs sampler uses the variable augmentation method for probit link described in Albert, J. H., & Chib, S. (1993). Since the variance of a binary variable is determined by its mean compared to quantitative traits, theta(s) are eliminated to avoid overfitting.

Value

BF_main  The Bayes factor of the main effect
BF_int  The Bayes factor of the interaction effect
post_odds_beta  The posterior odds of beta
post_odds_gamma  The posterior odds of gamma
com_a  The inverse of the posterior mean of the precision for additive genetic component
com_c  The inverse of the posterior mean of the precision for shared environmental component
blvcm_bin_data

mean_mu  The posterior mean of intercept
mean_beta The posterior mean of beta
mean_gamma The posterior mean of gamma
sd_mu    The posterior standard deviation of intercept
sd_beta  The posterior standard deviation of beta
sd_gamma The posterior standard deviation of gamma
mean_rv  The posterior mean of alphas

Author(s)
Liang He

References

Examples

data(blvcm_bin_data)
blvcm_bin(blvcm_bin_data$pheno_data, blvcm_bin_data$geno_data, iter=20000, burnin=1000, model=2)

---

Usage
data(blvcm_bin_data)

Format
The format is: List of 2 $ pheno_data: num [1:2000, 1:3] 0 1 1 0 ... attr(*, "dimnames")=List of 2 ...
$ geno_data: int [1:2000, 1:40] 0 0 0 0 ...

Examples
data(blvcm_bin_data)
Example data for BLVCM

Usage

data(blvcm_data)

Format

The format is: List of 2 $ pheno_data: num [1:600, 1:3] -0.0813 -1.0135 0.4363 0.7927 0.9597 ...
..- attr(*, "dimnames")=List of 2 .. $ : NULL .. $ : chr [1:3] "pheno" "fam" "zyg" $ geno_data:
int [1:600, 1:40] 0 0 0 0 0 0 0 0 0 0 ...

Examples

data(blvcm_data)
## maybe str(blvcm_data) ; plot(blvcm_data) ...

Hierarchical Bayesian multiple regression model incorporating genotype uncertainty (HBMR)

Description

The function implements HBMR using Gibbs sampling method.

Usage

hbmr(pheno, geno, qi, iter = 10000, burnin = 500, gq = -1, imp = -1, cov = 0, maf = -1,
rvinfo = FALSE, pa = -1, pb = -1)

Arguments

pheno Phenotypic vector (N x 1). For faster convergence, it is recommended that pheno-
type should be standardized.

geno N x K Genotypic data matrix, where N is the number of subjects and K is the
number of rare variants. Genotypic value is only for dominant coding, i.e. 0 or
1. Plug in 0 for imputed genotypes.

qi N x K Genotypic quality matrix, where N is the number of subjects and K is the
number of rare variants. If the genotype is sequenced, this must be an integer
>=1 and is its GQ score in VCF file. If the genotype is imputed, this must be a
value <1, and is its expected genotypic value based on the dominant coding.

iter Number of MCMC iterations
burnin Number of burn-in
gq cutoff for GQ score (lambda_Q). If not specified, default value is 20. See the
reference for more details.

imp cutoff for imputed genotype (lambda_I). If not specified, default value is 0.1. See the
reference for more details.
hbmr_data

N x M covariate data matrix, where N is the number of subjects and K is the number of covariates.

Minor allele frequency information vector (K by 1).

0 or 1. Default is 0. Indicator of showing estimated RV effect size and standard deviation.

The hyper-parameter a in the gamma distribution of Bayesian shrinkage prior. The default value is 1.3.

The hyper-parameter b in the gamma distribution of Bayesian shrinkage prior. The default value is 0.04.

The Bayes factor of delta=1 vs. delta=0

The BF estimated by using Rao-Blackwellization theorem

The mean of the posterior of beta_0

The inverse of the mean of posterior of precision 1/\sigma

The number of genotypes whose uncertainty are considered in estimation

The means of the posterior of gamma for the K RVs

The standard deviations of the posterior of gamma for the K RVs

The means of the posterior of for the M covariates

Liang He


data(hbmr_data)

hbmr(hbmr_data$pheno_data, hbmr_data$geno_data, hbmr_data$qual_data, iter=10000, burnin=1000)

Usage

data(hbmr_data)

Format

The format is: List of 3 $ pheno_data: num [1:600] -0.255 0.398 2.982 1.361 -0.165 ...

$ geno_data: num [1:600, 1:50] 1 0 0 0 0 0 0 0 0 0 ...

$ qual_data: num [1:600, 1:50] 5 5 5 99 99 99 99 99 99 99 ...

Example data for HBMR
Examples

data(hbmr_data)

## maybe str(hbmr_data) ; plot(hbmr_data) ...
Index

*Topic \textasciitilde kwd1
  blvcm, 2
  blvcm_bin, 4
  hbmr, 6

*Topic \textasciitilde kwd2
  blvcm, 2
  blvcm_bin, 4
  hbmr, 6

*Topic datasets
  blvcm_bin_data, 5
  blvcm_data, 6
  hbmr_data, 7

*Topic package
  BMRV-package, 1

blvcm, 2
blvcm_bin, 4
blvcm_bin_data, 5
blvcm_data, 6
BMRV (BMRV-package), 1
BMRV-package, 1

hbmr, 6
hbmr_data, 7