Package ‘BMRV’

October 12, 2022

Type Package

Title Bayesian Models for Rare Variant Association Analysis

Version 1.32

Date 2016-10-29

Author Liang He

Maintainer Liang He <liang.he@duke.edu>

Description Provides two Bayesian models for detecting the association between rare genetic variants and a trait that can be continuous, ordinal or binary. Bayesian latent variable collapsing model (BLVCM) detects interaction effect and is dedicated to twin design while it can also be applied to independent samples. Hierarchical Bayesian multiple regression model (HBMR) incorporates genotype uncertainty information and can be applied to either independent or family samples. Furthermore, it deals with continuous, binary and ordinal traits.

License GPL (>= 2)

Depends R (>= 3.1.1), stats

Encoding latin1

LinkingTo BH

Repository CRAN

Repository/R-Forge/Project bmrv

Repository/R-Forge/Revision 10

Repository/R-Forge/DateTimeStamp 2016-10-31 22:36:31

Date/Publication 2016-11-01 11:07:18

NeedsCompilation yes

R topics documented:

`BMRV-package` *Bayesian models for rare variant association analysis*

Description

This package provides two Bayesian models for detecting the association between a set of rare genetic variants and a trait that can be continuous, ordinal or binary. BLVCM detects interaction effect and is dedicated to twin design while it can also be applied to independent samples. HBMR incorporates genotype uncertainty information and can be applied to either independent or family samples. Furhtermore, it deals with continuous, binary and ordinal traits.

Details

Package:	BMRV
Type:	Package
Version:	1.32
Date:	2016-10-29
License:	GPL (>=2)

blvcm hbmr

Author(s)

Liang He

Maintainer: Liang He <liang.he@duke.edu>

References

He, L., Sillanpää, M. J., Ripatti, S., & Pitkäniemi, J. (2014). Bayesian Latent Variable Collapsing Model for Detecting Rare Variant Interaction Effect in Twin Study. Genetic epidemiology, 38(4), 310-324.

He, L., Pitkäniemi, J., Sarin, A. P., Salomaa, V., Sillanpää, M. J., & Ripatti, S. (2015). Hierarchical Bayesian Model for Rare Variant Association Analysis Integrating Genotype Uncertainty in Human Sequence Data. Genetic epidemiology, 39(2), 89-100.

Examples

```
data(blvcm_data)
temp<- blvcm(blvcm_data$pheno_data, blvcm_data$geno_data[,1:3], iter=10000, model = 3)
```

---

blvcm *Bayesian latent variable collapsing model (BLVCM)*

Description

The function implements BLVCM using a Gibbs sampler.

Usage

```
blvcm(pheno, geno, model = 3, iter = 30000, burnin = 500, var = -1, lambda = 0.2,
cov = 0, init = c(0,0))
```

Arguments

pheno	An N x 3 phenotypic data matrix (trait, family number, zyg=1 for MZ, 2 for DZ), where N is the number of subjects. Please see the example data for more details. For faster convergence, it is recommanded that the phenotype should be standardized.
geno	An 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, which must be positive.
burnin	The number of burn-ins, which must be positive.
var	The variance hyperparameter (must be positive) in the priors for β and γ. If not specified (var=-1), the default value is the variance of the phenotype.
lambda	The threshold λ (must be positive) for hypothesis test. The default value is 0.2.
cov	A matrix of other covariates.
init	Initial values for β and γ (must be non-negative). The default values are 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 β
post_odds_gamma	The posterior odds of γ
com_a	The inverse of the posterior mean of the precision for additive genetic component. NA for independent samples

com_c	The inverse of the posterior mean of the precision for shared environmental component. NA for independent samples or AE model
mean_mu	The posterior mean of the intercept μ
mean_beta	The posterior mean of β
mean_gamma	The posterior mean of γ
sd_mu	The posterior standard deviation of the intercept μ
sd_beta	The posterior standard deviation of β
sd_gamma	The posterior standard deviation of γ
mean_rv	The posterior mean of α. The number of α equals the number of RVs
mean_cov	The posterior mean of the effects of covariates
prior_var	The variance hyperparameters in the priors for β and γ

Author(s)

Liang He

References

He, L., Sillanpää, M. J., Ripatti, S., & Pitkäniemi, J. (2014). Bayesian Latent Variable Collapsing Model for Detecting Rare Variant Interaction Effect in Twin Study. Genetic epidemiology, 38(4), 310-324.

Examples

```
data(blvcm_data)
blvcm(blvcm_data$pheno, blvcm_data$geno[,1:3], iter=10000, 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 a Gibbs sampler with probit link function.

Usage

```
blvcm_bin(pheno, geno, model = 3, iter = 30000, burnin = 500, var = -1, lambda = 0.2,
cov = 0, init = c(0, 0))
```

Arguments

- `pheno` — An N x 3 phenotypic data matrix (trait, family number, zyg=1 for MZ, 2 for DZ), where N is the number of subjects. The trait must be 0 or 1.
- `geno` — An 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 (must be positive). The default value is 30000.
- `burnin` — The number of burn-ins (must be positive). The default value is 500.
- `var` — The variance hyperparameters (must be positive) in the priors for β and γ. The default value is 1.
- `lambda` — The threshold λ (must be positive) for hypothesis test. The default value is 0.2.
- `cov` — A matrix of other covariates to be adjusted.
- `init` — Initial values for β and γ. The default values are 0. The initial value for β must be non-negative.

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, θ(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 β
- `post_odds_gamma` — The posterior odds of γ
- `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 the intercept μ
- `mean_beta` — The posterior mean of β
- `mean_gamma` — The posterior mean of γ
- `sd_mu` — The posterior standard deviation of the intercept μ
- `sd_beta` — The posterior standard deviation of β
- `sd_gamma` — The posterior standard deviation of γ
- `mean_rv` — The posterior mean of α
- `mean_cov` — The posterior mean of the effects of covariates

Author(s)

Liang He

References

He, L., Sillanpää, M. J., Ripatti, S., & Pitkäniemi, J. (2014). Bayesian Latent Variable Collapsing Model for Detecting Rare Variant Interaction Effect in Twin Study. Genetic epidemiology, 38(4), 310-324.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American statistical Association, 88(422), 669-679.

Examples

```
data(blvcm_bin_data)
blvcm_bin(blvcm_bin_data$pheno, blvcm_bin_data$geno[,1:3], iter=5000, burnin=500, model=2)
```

`blvcm_bin_data` *Example data for BLVCM_bin*

Description

This is an example dataset consisting of binary traits for the blvcm_bin function.

Usage

```
data(blvcm_bin_data)
```

Format

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

Examples

```
data(blvcm_bin_data)
```

`blvcm_data` *Example data for BLVCM*

Description

This is an example dataset consisting of continuous traits for the blvcm function.

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)
```

`hbmr` *Hierarchical Bayesian multiple regression model incorporating genotype uncertainty (HBMR)*

Description

The function implements HBMR using Gibbs sampling method for quantitative traits.

Usage

```
hbmr(pheno, geno, qi = matrix(), fam = 0, kin = matrix(), iter = 10000, burnin = 500,
gq = 20, imp = 0.1, cov = matrix(), maf = c(), rvinfo = FALSE, pa = 1.3, pb = 0.04)
```

Arguments

pheno	Phenotypic vector (N x 1). For better inference, it is recommanded 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. Genotypic value is only for dominant coding, i.e. 0 or 1. Plug in 0 for imputed genotypes.

`qi`	An optional 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.
`fam`	fam=1 for family samples. In this case, a relatedness matrix should be given. See kin.
`kin`	In the case of fam=1, kin is an N x N relatedness matrix.
`iter`	The number of MCMC iterations. The default value is 10000.
`burnin`	The number of burn-ins. The default value is 500.
`gq`	A cutoff for GQ score (λ_Q). It should be an positive integer. If not specified, default value is 20. See the reference for more details.
`imp`	A cutoff for imputed genotype (λ_I). It should be a real number in (0,1). If not specified, default value is 0.1. See the reference for more details.
`cov`	An optional N x M covariate data matrix, where N is the number of subjects and M is the number of covariates.
`maf`	An optional minor allele frequency information vector (K x 1). If not specified, MAF will be estimated based on the genotype data.
`rvinfo`	TRUE or FALSE. Default is FALSE. Indicator of showing estimatd RV effect size and standard deviation.
`pa`	The positive hyper-parameter a in the gamma distribution of Bayesian shrinkage prior. The default value is 1.3.
`pb`	The positive hyper-parameter b in the gamma distribution of Bayesian shrinkage prior. The default value is 0.04.

Value

`BF`	The Bayes factor of $\delta = 1$ vs. $\delta = 0$
`BF_RB`	The BF estimated by using Rao-Blackwellization theorem
`p_upper`	For a BF larger than 2, we calculate p_upper that is the upper bound of the p value corresponding to the BF based on the connection $BF < (-1)/(e * p * log(p))$. The exact p value, which is smaller than p_upper, can be obtained through permutations.
`mean`	The mean of the posterior of β_0
`var`	The inverse of the mean of posterior of precision 1/σ
`est_geno`	The number of genotypes whose uncertainty are considered in estimation
`var_ran`	The estimated variance of the random effect for family design
`rv_mean_es`	The means of the posterior of γ for the K RVs
`rv_sd_es`	The standard deviations of the posterior of γ for the K RVs
`mean_cov`	The means of the posterior of for the M covariates

Author(s)

Liang He

References

He, L., Pitkäniemi, J., Sarin, A. P., Salomaa, V., Sillanpää, M. J., & Ripatti, S. (2015). Hierarchical Bayesian Model for Rare Variant Association Analysis Integrating Genotype Uncertainty in Human Sequence Data. Genetic epidemiology, 39(2), 89-100.

Examples

```
data(hbmr_data)
hbmr(hbmr_data$pheno_data, hbmr_data$geno_data[,1:3], hbmr_data$qual_data[,1:3],
iter=10000, burnin=1000)
```

---

hbmr_bin — *Hierarchical Bayesian multiple regression model incorporating genotype uncertainty (HBMR) for binary traits*

Description

The function implements HBMR using a Gibbs sampler with probit link function for binary traits.

Usage

```
hbmr_bin(pheno, geno, qi = matrix(), fam = 0, kin = matrix(), iter = 10000, burnin = 500,
gq = 20, imp = 0.1, cov = matrix(), maf = c(), pa = 1.3, pb = 0.04)
```

Arguments

pheno	A phenotypic vector (N x 1). The trait must be 0 or 1.
geno	An 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	An optional 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.
fam	fam=1 for family samples. In this case, a relatedness matrix should be given. See kin.
kin	In the case of fam=1, kin is an N x N relatedness matrix. The scale of its entries are twice the kinship coefs, i.e. the same as that in coxme.
iter	The number of MCMC iterations. The default value is 10000.
burnin	The number of burn-ins. The default value is 500.
gq	A cutoff for GQ score (λ_Q). It should be an positive integer. If not specified, default value is 20. See the reference for more details.
imp	A cutoff for imputed genotype (λ_I). It should be a real number in (0,1). If not specified, default value is 0.1. See the reference for more details.

- `cov` An optional N x M covariate data matrix, where N is the number of subjects and M is the number of covariates.
- `maf` An optional minor allele frequency information vector (K x 1). If not specified, MAF will be estimated based on the genotype data.
- `pa` The positive hyper-parameter a in the gamma distribution of Bayesian shrinkage prior. The default value is 1.3.
- `pb` The positive hyper-parameter b in the gamma distribution of Bayesian shrinkage prior. The default value is 0.04.

Value

- `BF` The Bayes factor of $\delta = 1$ vs. $\delta = 0$
- `BF_RB` The BF estimated by using Rao-Blackwellization theorem
- `p_upper` For a BF larger than 2, we calculate p_upper that is the upper bound of the p value corresponding to the BF based on the connection $BF < (-1)/(e * p * log(p))$. The exact p value, which is smaller than p_upper, can be obtained through permutations.
- `mean` The mean of the posterior of β_0
- `var` The inverse of the mean of posterior of precision 1/σ
- `est_geno` The number of genotypes whose uncertainty are considered in estimation
- `var_ran` The estimated variance of the random effect for family design
- `rv_mean_es` The means of the posterior of γ for the K RVs
- `rv_sd_es` The standard deviations of the posterior of γ for the K RVs
- `mean_cov` The means of the posterior of for the M covariates

Author(s)

Liang He

References

He, L., Pitkäniemi, J., Sarin, A. P., Salomaa, V., Sillanpää, M. J., & Ripatti, S. (2015). Hierarchical Bayesian Model for Rare Variant Association Analysis Integrating Genotype Uncertainty in Human Sequence Data. Genetic epidemiology, 39(2), 89-100.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American statistical Association, 88(422), 669-679.

Examples

```
data(hbmr_bin_data)
hbmr_bin(hbmr_bin_data$pheno[1:500], hbmr_bin_data$geno[1:500,1:3], fam=1,
kin= hbmr_bin_data$kin[1:500,1:500], iter=800, burnin=200)
```

hbmr_bin_data — *Example data for hbmr_bin*

Description

This is an example dataset consisting of binary traits for the hbmr_bin function.

Usage

```
data(hbmr_bin_data)
```

Format

The format is: List of 2 $ pheno: num [1:1720] 0 1 1 1 0 ... $ geno : int [1:1720, 1:40] 0 0 0 0 0 0 0 0 0 0 ... $ qi : int [1:1720, 1:40] 0 0 0 0 0 0 0 0 0 0 ... $ kin : int [1:1720, 1:1720] 0 0 0 0 0 0 0 0 0 0 ...

Examples

```
data(hbmr_bin_data)
```

hbmr_data — *Example data for HBMR*

Description

This is an example dataset consisting of continuous traits for the hbmr function.

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 ...

Examples

```
data(hbmr_data)
```

hbmr_ord *Hierarchical Bayesian multiple regression model incorporating genotype uncertainty (HBMR) for ordinal traits*

Description

The function implements HBMR using a Gibbs sampler with probit link function for ordinal traits.

Usage

```
hbmr_ord(pheno, geno, qi = matrix(), fam = 0, kin = matrix(), iter = 10000, burnin = 500,
gq = 20, imp = 0.1, cov = matrix(), maf = c(), pa = 1.3, pb = 0.04)
```

Arguments

pheno	A phenotypic vector (N x 1). The trait must be a natural number (1, 2, 3, 4, ...).
geno	An 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	An optional 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.
fam	fam=1 for family samples. In this case, a relatedness matrix should be given. See kin.
kin	In the case of fam=1, kin is an N x N relatedness matrix. The scale of its entries are twice the kinship coefs, i.e. the same as that in coxme.
iter	The number of MCMC iterations. The default value is 10000.
burnin	The number of burn-ins. The default value is 500.
gq	A cutoff for GQ score (λ_Q). It should be an positive integer. If not specified, default value is 20. See the reference for more details.
imp	A cutoff for imputed genotype (λ_I). It should be a real number in (0,1). If not specified, default value is 0.1. See the reference for more details.
cov	An optional N x M covariate data matrix, where N is the number of subjects and M is the number of covariates.
maf	An optional minor allele frequency information vector (K by 1). If not specified, MAF will be estimated based on the genotype data.
pa	The positive hyper-parameter a in the gamma distribution of Bayesian shrinkage prior. The default value is 1.3.
pb	The positive hyper-parameter b in the gamma distribution of Bayesian shrinkage prior. The default value is 0.04.

Value

BF	The Bayes factor of $\delta = 1$ vs. $\delta = 0$
BF_RB	The BF estimated by using Rao-Blackwellization theorem
p_upper	For a BF larger than 2, we calculate p_upper that is the upper bound of the p value corresponding to the BF based on the connection $BF < (-1)/(e * p * log(p))$. The exact p value, which is smaller than p_upper, can be obtained through permutations.
mean	The mean of the posterior of β_0
var	The inverse of the mean of posterior of precision $1/\sigma$
est_geno	The number of genotypes whose uncertainty are considered in estimation
var_ran	The estimated variance of the random effect for family design
rv_mean_es	The means of the posterior of γ for the K RVs
rv_sd_es	The standard deviations of the posterior of γ for the K RVs
mean_cov	The means of the posterior of for the M covariates

Author(s)

Liang He

References

He, L., Pitkäniemi, J., Sarin, A. P., Salomaa, V., Sillanpää, M. J., & Ripatti, S. (2015). Hierarchical Bayesian Model for Rare Variant Association Analysis Integrating Genotype Uncertainty in Human Sequence Data. Genetic epidemiology, 39(2), 89-100.

Kärkkäinen, H. P., & Sillanpää, M. J. (2013). Fast Genomic Predictions via Bayesian G-BLUP and Multilocus Models of Threshold Traits Including Censored Gaussian Data. G3: Genes| Genomes| Genetics, 3(9), 1511-1523.

Examples

```
data(hbmr_bin_data)
hbmr_ord(hbmr_bin_data$pheno[1:500], hbmr_bin_data$geno[1:500,1:3], fam=1,
kin= hbmr_bin_data$kin[1:500,1:500], iter=800, burnin=200)
```

Index