Package 'mixedBayes'

May 1, 2025

Type Package

Title Bayesian Longitudinal Regularized Quantile Mixed Model

Version 0.1.7

Date 2025-04-30

Description With high-dimensional omics features, repeated measure ANOVA leads to longitudinal gene-environment interaction studies that have intra-cluster correlations, outlying observations and structured sparsity arising from the ANOVA design. In this package, we have developed robust sparse Bayesian mixed effect models tailored for the above studies (Fan et al. (2025) <doi:10.1093/jrsssc/qlaf027>). An efficient Gibbs sampler has been developed to facilitate fast computation. The Markov chain Monte Carlo algorithms of the proposed and alternative methods are efficiently implemented in 'C++'. The development of this software package and the associated statistical methods have been partially supported by an Innovative Research Award from Johnson Cancer Research Center, Kansas State University.

Depends R (>= 4.2.0)

License GPL-2

Encoding UTF-8

URL `https://github.com/kunfa/mixedBayes`

Imports Rcpp

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.3.2

NeedsCompilation yes

Author Kun Fan [aut, cre],
Cen Wu [aut]

Maintainer Kun Fan <`kfan@ksu.edu`>

Repository CRAN

Date/Publication 2025-05-01 13:30:02 UTC

Contents

`mixedBayes-package` *Bayesian Longitudinal Regularized Quantile Mixed Model*

Description

In this package, we provide implementations of a set of high-dimensional robust Bayesian mixed-effect models to dissect longitudinal gene-environment interactions. The proposed method conducts robust Bayesian variable selection on both the main and interaction effects corresponding to individual and group levels (i.e. bi-level), respectively. Alternatively, selections only on individual levels by ignoring the grouping structure can also be performed. In addition, intra-cluster correlations among repeated measures are modeled via random intercept-and-slope and/or random intercept models. Imposing exact sparsity through spike-and-slab priors can be conducted on fixed effects with bi-level and/or individual level. In total, package mixedBayes provides implementations on 2 (robust and non-robust) × 2 (types of fixed effects) × 2 (types of random effects) × 2 (spike-and-slab or Laplacian priors) = 16 methods. Please read the details below for how to configure the method used.

Details

The user friendly, integrated interface **mixedBayes()** allows users to flexibly choose the fitting methods by specifying the following parameter:

slope:	whether to use random intercept-and-slope model or random intercept model.
robust:	whether to use robust or non-robust methods.
quant:	to specify different quantiles when using robust methods.
structure:	whether to specify bi-level or individual level.
sparse:	whether to use the spike-and-slab priors to impose sparsity.

The function mixedBayes() returns a mixedBayes object that contains the posterior estimates of each coefficients. S3 generic functions selection()and print() are implemented for mixedBayes objects. selection() takes a mixedBayes object and returns the variable selection results.

References

Fan, K., Jiang, Y., Ma, S., Wang, W. and Wu, C. (2025). Robust Sparse Bayesian Regression for Longitudinal Gene-Environment Interactions. *Journal of the Royal Statistical Society Series C: Applied Statistics*, qlaf027 doi:10.1093/jrsssc/qlaf027

Zhou, F., Ren, J., Li, G., Jiang, Y., Li, X., Wang, W. and Wu, C. (2019). Penalized Variable Selection for Lipid-Environment Interactions in a Longitudinal Lipidomics Study. *Genes*, 10(12), 1002 doi:10.3390/genes10121002

Zhou, F., Ren, J., Liu, Y., Li, X., Wang, W., and Wu, C. (2022). Interep: An r package for high-dimensional interaction analysis of the repeated measurement data. *Genes*, 13(3), 544 doi:10.3390/genes13030544

Zhou, F., Lu, X., Ren, J., Fan, K., Ma, S., and Wu, C. (2022). Sparse group variable selection for gene–environment interactions in the longitudinal study. *Genetic epidemiology*, 46(5-6), 317-340 doi:10.1002/gepi.22461

Ren, J., Zhou, F., Li, X., Ma, S., Jiang, Y. and Wu, C. (2023). Robust Bayesian variable selection for gene-environment interactions. *Biometrics*,79(2),684-694 doi:10.1111/biom.13670

Wu, C., and Ma, S. (2015). A selective review of robust variable selection with applications in bioinformatics. *Briefings in Bioinformatics*, 16(5), 873–883 doi:10.1093/bib/bbu046

Zhou, F., Ren, J., Lu, X., Ma, S. and Wu, C. (2021). Gene–Environment Interaction: a Variable Selection Perspective. *Epistasis. Methods in Molecular Biology.* 2212:191–223 doi:10.1007/9781-071609477_13

Ren, J., Zhou, F., Li, X., Chen, Q., Zhang, H., Ma, S., Jiang, Y. and Wu, C. (2020) Semi-parametric Bayesian variable selection for gene-environment interactions. *Statistics in Medicine*, 39: 617– 638 doi:10.1002/sim.8434

Wu, C., Jiang, Y., Ren, J., Cui, Y. and Ma, S. (2018). Dissecting gene-environment interactions: A penalized robust approach accounting for hierarchical structures. *Statistics in Medicine*, 37:437–456 doi:10.1002/sim.7518

Wu, C., Cui, Y., and Ma, S. (2014). Integrative analysis of gene–environment interactions under a multi–response partially linear varying coefficient model. *Statistics in Medicine*, 33(28), 4988–4998 doi:10.1002/sim.6287

Wu, C., Zhong, P.S. and Cui, Y. (2013). High dimensional variable selection for gene-environment interactions. *Technical Report. Michigan State University.*

See Also

`mixedBayes`

`data` *simulated data for demonstrating the features of mixedBayes*

Description

Simulated gene expression data for demonstrating the features of mixedBayes.

Format

The data object consists of seven components: y, e, X, g, w ,k and coeff. coeff contains the true values of parameters used for generating Y.

Details

The data and model setting

Consider a longitudinal study on n subjects with k repeated measurement for each subject. Let Y_{ij} be the measurement for the ith subject at each time point $j(1 \leq i \leq n, 1 \leq j \leq k)$ .We use a m-dimensional vector G_{ij} to denote the genetics factors, where $G_{ij} = (G_{ij1}, ..., G_{ijm})^\top$. Also, we use p-dimensional vector E_{ij} to denote the environment factors, where $E_{ij} = (E_{ij1}, ..., E_{ijp})^\top$. $X_{ij} = (1, T_{ij})^\top$, where $T_{ij}^\top$ is a vector of time effects . Z_{ij} is a $h \times 1$ covariate associated with random effects and α_i is a $h \times 1$ vector of random effects. In a typical one-way repeated measure ANOVA with a fixed number (say four) of factor levels, the environment (or treatment) factor is modelled as a group of three dummy variables. Therefore, gene-environment (or treatment) interaction leads to variable selections on individual levels (main effects) and group levels (interaction effect) simultaneously. Considering the genetics factors, environment (or treatment) factors and their interactions that are jointly associated with the longitudinal phenotype, we have the following mixed-effects model:

$$Y_{ij} = X_{ij}^\top \gamma_0 + E_{ij}^\top \gamma_1 + G_{ij}^\top \gamma_2 + (G_{ij} \bigotimes E_{ij})^\top \gamma_3 + Z_{ij}^\top \alpha_i + \epsilon_{ij}.$$

where γ_1,γ_2,γ_3 are p,m and mp dimensional vectors that represent the coefficients of the environment effects, the genetics effects and interactions effects, respectively. Accommodating the Kronecker product of the m - dimensional vector G_{ij} and the p-dimensional vector E_{ij}, the interactions between genetics and environment factors can be expressed as a mp-dimensional vector, denoted as the following form:

$$G_{ij} \bigotimes E_{ij} = [E_{ij1}E_{ij1}, E_{ij2}E_{ij2}, ..., E_{ij1}E_{ijp}, E_{ij2}E_{ij1}, ..., E_{ijm}E_{ijp}]^\top.$$

For random intercept-and-slope model, $Z_{ij}^\top = (1, j)$ and $\alpha_i = (\alpha_{i1}, \alpha_{i2})^\top$. For random intercept model, $Z_{ij}^\top = 1$ and $\alpha_i = \alpha_{i1}$.

See Also

`mixedBayes`

Examples

```
data(data)
length(y)
dim(g)
dim(e)
dim(w)
print(k)
print(X)
print(coeff)
```

`mixedBayes` *fit a Bayesian longitudinal regularized quantile mixed model*

Description

fit a Bayesian longitudinal regularized quantile mixed model

Usage

```
mixedBayes(
  y,
  e,
  X,
  g,
  w,
  k,
  iterations = 10000,
  burn.in = NULL,
  slope = TRUE,
  robust = TRUE,
  quant = 0.5,
  sparse = TRUE,
  structure = c("bi-level", "individual")
)
```

Arguments

`y`	the vector of repeated measured responses. The current version of mixedBayes only supports continuous response.
`e`	the long format matrix of environment (treatment) factors (a group of dummy variables).
`X`	the long format matrix of the intercept and time effects (time effects are optional).
`g`	the long format matrix of predictors (genetic factors) without intercept. Each row should be an observation vector.
`w`	the long format matrix of interactions between genetic factors and environment (treatment) factors.
`k`	the number of repeated measurements.
`iterations`	the number of MCMC iterations.
`burn.in`	the number of iterations for burn-in.
`slope`	logical flag. If TRUE, random intercept-and-slope model will be used.
`robust`	logical flag. If TRUE, robust methods will be used.
`quant`	specify different quantiles when applying robust methods.

`sparse`	logical flag. If TRUE, spike-and-slab priors will be used to shrink coefficients of irrelevant covariates to zero exactly.
`structure`	two choices are available. "bi-level" for selection on both the main and interaction effects corresponding to individual and group levels. "individual" for selections on individual-level only.

Details

Consider the data model described in "`data`":

$$Y_{ij} = X_{ij}^\top\gamma_0 + E_{ij}^\top\gamma_1 + \sum_{l=1}^{p} G_{ijl}\gamma_{2l} + \sum_{l=1}^{p} W_{ijl}^\top\gamma_{3l} + Z_{ij}^\top\alpha_i + \epsilon_{ij}.$$

where γ_{2l} is the main effect of the lth genetic variant. The interaction effects is corresponding to the coefficient vector $\gamma_{3l} = (\gamma_{3l1}, \gamma_{3l2}, \ldots, \gamma_{3lm})^\top$, with $W_{ij} = G_{ij} \bigotimes E_{ij}$.

When ‘structure="bi-level"‘, bi-level selection will be conducted. If ‘structure="individual"‘, individual-level selection will be conducted.

When ‘slope=TRUE‘ (default), random intercept-and-slope model will be used as the mixed effects model.

When ‘sparse=TRUE‘ (default), spike-and-slab priors are imposed to identify important main and interaction effects. Otherwise, Laplacian shrinkage will be used.

When ‘robust=TRUE‘ (default), the distribution of ϵ_{ij} is defined as an asymmetric Laplace distribution with density.

$f(\epsilon_{ij}|\theta, \tau) = \theta(1-\theta)\exp\{-\tau\rho_\theta(\epsilon_{ij})\}$, $(i = 1, \ldots, n, j = 1, \ldots, k)$, which leads to a Bayesian formulation of quantile regression. If ‘robust=FALSE‘, ϵ_{ij} follows a normal distribution.

Please check the references for more details about the prior distributions.

Value

an object of class ‘mixedBayes’ is returned, which is a list with component:

`posterior`	the posteriors of coefficients.
`coefficient`	the estimated coefficients.
`burn.in`	the total number of burn-ins.
`iterations`	the total number of iterations.

See Also

`data`

Examples

```
data(data)

## default method (robust sparse bi-level selection under random intercept-and-slope model)
fit = mixedBayes(y,e,X,g,w,k,structure=c("bi-level"))
```

```
fit$coefficient

## Compute TP and FP
b = selection(fit,sparse=TRUE)
index = which(coeff!=0)
pos = which(b != 0)
tp = length(intersect(index, pos))
fp = length(pos) - tp
list(tp=tp, fp=fp)

## alternative: robust sparse individual level selections under random intercept-and-slope model
fit = mixedBayes(y,e,X,g,w,k,structure=c("individual"))
fit$coefficient

## alternative: non-robust sparse bi-level selection under random intercept-and-slope model
fit = mixedBayes(y,e,X,g,w,k,robust=FALSE, structure=c("bi-level"))
fit$coefficient

## alternative: robust sparse bi-level selection under random intercept model
fit = mixedBayes(y,e,X,g,w,k,slope=FALSE, structure=c("bi-level"))
fit$coefficient
```

`selection` *Variable selection for a mixedBayes object*

Description

Variable selection for a mixedBayes object

Usage

```
selection(obj, sparse)
```

Arguments

`obj`	mixedBayes object.
`sparse`	logical flag. If TRUE, spike-and-slab priors will be used to shrink coefficients of irrelevant covariates to zero exactly..

Details

If sparse, the median probability model (MPM) (Barbieri and Berger, 2004) is used to identify predictors that are significantly associated with the response variable. Otherwise, variable selection is based on 95% credible interval. Please check the references for more details about the variable selection.

Value

an object of class ‘selection’ is returned, which is a list with component:

`inde`	a vector of indicators of selected effects.

References

Ren, J., Zhou, F., Li, X., Ma, S., Jiang, Y. and Wu, C. (2023). Robust Bayesian variable selection for gene-environment interactions. *Biometrics*,79(2),684-694 doi:10.1111/biom.13670

Barbieri, M.M. and Berger, J.O. (2004). Optimal predictive model selection. *Ann. Statist*, 32(3):870–897

See Also

`mixedBayes`

Examples

```
data(data)
## sparse
fit = mixedBayes(y,e,X,g,w,k,structure=c("bi-level"))
selected=selection(fit,sparse=TRUE)
selected


## non-sparse
fit = mixedBayes(y,e,X,g,w,k,sparse=FALSE,structure=c("bi-level"))
selected=selection(fit,sparse=FALSE)
selected
```

Index