Package 'mixedBayes'

May 8, 2026

Type Package

Title Bayesian Longitudinal Regularized Quantile Mixed Model

Version 0.2.5

Date 2026-04-23

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

BugReports https://github.com/kunfa/mixedBayes/issues

Imports Rcpp

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.3.3

NeedsCompilation yes

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

Maintainer Kun Fan <fzt0428@gmail.com>

Repository CRAN

Date/Publication 2026-04-23 06:30:02 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 group 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 quantile levels 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*, 74(5), 1372–1394 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 data for demonstrating the features of mixedBayes

Format

The data object consists of seven components: y, e, X, g, k, and coeff. The response y and the covariates e, X, and g are all provided in long format and share a common row ordering. The component coeff contains the true parameter values (main and interaction effects) used to generate 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 the m-dimensional vector $\boldsymbol{G_{ij}}$ to denote measurements of genetic factors for the ith subject at time point j, where $\boldsymbol{G_{ij}} = (G_{ij1}, ..., G_{ijm})^\top$. Also, we use p-dimensional vector $\boldsymbol{E_{ij}}$ to denote the environmental/treatment factors, where $\boldsymbol{E_{ij}} = (E_{ij1}, ..., E_{ijp})^\top$. $\boldsymbol{X_{ij}} = (1, \boldsymbol{T_{ij}^\top})^\top$, where $\boldsymbol{T_{ij}}$ is a vector of time effects . $\boldsymbol{Z_{ij}}$ is a $h \times 1$ covariate associated with random effects and $\boldsymbol{\alpha_{i,\theta}}$ 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 environmental (or treatment) factor is modeled 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 effects) simultaneously. Considering the genetics factors, environmental (or treatment) factors and their interactions that are jointly associated with the longitudinal phenotype, we have the following mixed-effects model at a given quantile level θ, $(0 < \theta < 1)$:

$$Y_{ij} = \boldsymbol{X_{ij}^\top \gamma_{0,\theta}} + \boldsymbol{E_{ij}^\top \gamma_{1,\theta}} + \boldsymbol{G_{ij}^\top \gamma_{2,\theta}} + (\boldsymbol{G_{ij}} \bigotimes \boldsymbol{E_{ij}})^\top \boldsymbol{\gamma_{3,\theta}} + \boldsymbol{Z_{ij}^\top \alpha_{i,\theta}} + \epsilon_{ij,\theta}.$$

where $\boldsymbol{\gamma_{1,\theta}}$,$\boldsymbol{\gamma_{2,\theta}}$,$\boldsymbol{\gamma_{3,\theta}}$ are p,m and mp dimensional vectors that represent the coefficients of the environmental effects, the genetic effects and interaction effects, respectively. In addition, $\boldsymbol{\gamma_{0,\theta}}$ is the coefficient vector for $\boldsymbol{X_{ij}}$. The gene–environment interactions that can be expressed as a Kronecker product between the two types of main effects as a mp-dimensional vector:

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

The above model also includes $\boldsymbol{Z_{ij}}$ with random effects $\boldsymbol{\alpha_{i,\theta}}$ to account for intra-correlations among repeated measurements. The $h \times 1$ vector $\boldsymbol{Z_{ij}}$ corresponds to the random intercept–slope model and random intercept model under $h = 2$ and 1, respectively. The model error $\epsilon_{ij,\theta}$'s are independent with the θth quantile being zero.

Without loss of generality, we suppress the subscript θ of the regression coefficient vectors for both fixed and random effects from now on for simplicity of notation.

In this example, we generate data under random intercept-and-slope model.

See Also

`mixedBayes`

Examples

```
data(data)
length(y)
```

```
dim(g)
dim(e)
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,
  k,
  iterations = 10000,
  burn.in = 5000,
  slope = TRUE,
  robust = TRUE,
  quant = 0.5,
  sparse = TRUE,
  structure = "bilevel"
)
```

Arguments

`y`	a numeric vector of repeated-measure responses in long format. The current version only supports continuous response.
`e`	the long-format matrix for environmental/treatment effects. In applications, this corresponds to a set of dummy variables encoding treatment levels.
`X`	the long-format matrix, including an intercept term and optionally time covariates.
`g`	the long-format matrix of genetic predictors.
`k`	a positive integer. Number of repeated measurements per subject.
`iterations`	the number of MCMC iterations. The default value is 10,000.
`burn.in`	the number of iterations for burn-in. If NULL, no burn-in is applied and all MCMC samples are retained. The default value is 5,000.
`slope`	logical flag. If TRUE, random intercept-and-slope model will be used. Otherwise, random intercept model will be used. The default value is TRUE.

robust	logical flag. If TRUE, robust methods will be used. Otherwise, non-robust methods will be used. The default value is TRUE.
quant	the quantile level specified by users. Required when robust = TRUE. Ignored (set to NULL) when robust = FALSE. The default value is 0.5.
sparse	logical flag. If TRUE, spike-and-slab priors will be adopted to impose exact sparsity on regression coefficients. Otherwise, Laplacian shrinkage will be adopted. The default value is TRUE.
structure	two choices are available. "bilevel" performs selection on both main effects and interaction effects corresponding to individual and group levels, whereas "individual" performs selections only on individual levels by ignoring the group structure.

Details

Data layout

Consider a longitudinal study with repeated measurements per subject, where each subject has the same number of repeated measurements.

The response vector y and the design matrices X, e and g must all be provided in long format and share the same row ordering. In practice, each row corresponds to one observation from a particular subject at a particular time point.

The interaction terms between genetic and environmental factors (G×E) are constructed internally within the function and therefore do not need to be provided by the user.

Model

Consider the data model described in "data":

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

Here $\boldsymbol{\gamma_0}$ is the coefficient vector for $\boldsymbol{X_{ij}}$, $\boldsymbol{\gamma_1}$ is the coefficient vector for $\boldsymbol{E_{ij}}$, $\boldsymbol{\gamma_2}$ is the coefficient vector for the genetic variants, and $\boldsymbol{\gamma_3}$ is the coefficient vector for the interactions of the genetic variants with environment factors.

where $\boldsymbol{\gamma_1} = (\gamma_{11}, \ldots, \gamma_{1p})^\top$, $\boldsymbol{\gamma_2} = (\gamma_{21}, \ldots, \gamma_{2m})^\top$, $\boldsymbol{\gamma_3} = (\boldsymbol{\gamma_{31}}, \ldots, \boldsymbol{\gamma_{3m}})^\top$ where $\boldsymbol{\gamma_{3l}} = (\gamma_{3l1}, \ldots, \gamma_{3lp})^\top$ for $l = 1, \ldots, m$.

The subject-specific random effects $\boldsymbol{\alpha_i}$ capture within-subject correlation. For random intercept-and-slope model, $\boldsymbol{Z_{ij}^\top} = (1, j)$ and $\boldsymbol{\alpha_i} = (\alpha_{i1}, \alpha_{i2})^\top$. For random intercept model, $Z_{ij} = 1$ and $\alpha_i = \alpha_{i1}$.

When 'structure=bilevel'(default), bi-level selection on main and interaction effects will be conducted corresponding to individual and group levels, respectively. When 'structure="individual"', selections only on individual levels by ignoring the group structure will be performed.

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

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. Otherwise, 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 components:

`posterior`	Posterior samples for fixed effects and random effects.
`coefficient`	Posterior median estimates of fixed effects and random effects.
`sparse`	Logical value indicating whether spike-and-slab priors were used.
`robust`	Logical value indicating whether the robust (quantile regression) model was used.
`slope`	Logical value indicating whether a random intercept-and-slope model was used.
`quant`	Quantile level used when `robust = TRUE`; otherwise `NULL`.
`structure`	Character string specifying the selection structure ("bilevel" or "individual").
`k`	Number of repeated measurements per subject.
`iterations`	Total number of MCMC iterations.
`burn.in`	Number of burn-in 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,k,structure="bilevel")
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,k,structure="individual")
fit$coefficient

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

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

`predict_mixedBayes` *Make predictions from a mixedBayes object*

Description

Make predictions from a mixedBayes object

Usage

```
predict_mixedBayes(object, y, X, e, g, k, slope, loss)
```

Arguments

`object`	a mixedBayes object.
`y`	a numeric vector of repeated-measure responses in long format. The current version only supports continuous response.
`X`	the long-format design matrix, including an intercept and optionally time-related covariates.
`e`	the long-format design matrix for environment/treatment effects. In applications, this is a group of dummy variables encoding treatment levels.
`g`	the long-format matrix of genetic predictors. The interaction terms between genetic and environmental factors (G×E) are constructed internally within the function and therefore do not need to be provided by the user.
`k`	integer. Number of repeated measurements per subject.
`slope`	logical flag. If TRUE, random intercept-and-slope model will be used.
`loss`	character string specifying the prediction loss function. "L1" for mean absolute error; "L2" for mean squared error.

Value

an object of class 'mixedBayes.pred' is returned, which is a list with components:

`pred_error`	prediction error.
`y_hat`	predicted values of the repeated measured responses.

See Also

`mixedBayes`

Examples

```
data(data)

fit <- mixedBayes(y, e, X, g, k, structure = "bilevel")
pred1 <- predict_mixedBayes(fit, y, X, e, g, k, slope = TRUE, loss = "L1")
print(pred1$pred_error)
fit <- mixedBayes(y, e, X, g, k, robust =FALSE, quant =NULL,structure = "bilevel")
pred2 <- predict_mixedBayes(fit, y, X, e, g, k, slope = TRUE, loss = "L2")
print(pred2$pred_error)
```

---

reformat *This function changes the format of the longitudinal data from wide format to long format*

Description

This function changes the format of the longitudinal data from wide format to long format

Usage

```
reformat(k,data,type="r")
```

Arguments

k	the number of repeated measurement.
data	either the response matrix or the predictor matrix.
type	"r" for response matrix, "d" for design matrix.

Value

the reformatted response vector or predictor matrix.

---

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:

index	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,k,structure="bilevel")
selected=selection(fit,sparse=TRUE)
selected


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

Index