# Package 'BayesQVGEL'

April 12, 2023

**Type** Package

**Title** Bayesian Quantile Variable Selection for Gene - Environment Interactions in Longitudinal Studies

**Version** 0.1

**Date** 2023-04-03

**Description** In longitudinal studies, the same subjects are measured repeatedly over time, leading to correlations among the repeated measurements. Properly accounting for the intra-cluster correlations in the presence of data heterogeneity and long tailed distributions of the disease phenotype is challenging, especially in the context of high dimensional regressions. Here, we aim at developing novel Bayesian regularized quantile mixed effect models to tackle these challenges. We have proposed a Bayesian variable selection in the mixed effect models for longitudinal genomics studies. To dissect important gene - environment interactions, our model can simultaneously identify important main and interaction effects on the individual and group level, which have been facilitated by imposing the spike- and -slab priors through Laplacian shrinkage in the Bayesian quantile hierarchical models. The within - subject dependence among data can be accommodated by incorporating the random effects. 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++.

**Depends** R (>= 4.2.0)

**License** GPL-2

**Encoding** UTF-8

**URL** https://github.com/kunfa/BayesQVGEL

**Imports** Rcpp

**LinkingTo** Rcpp, RcppArmadillo

**RoxygenNote** 7.2.3

# R topics documented:

| BayesQVGEL-package | _Bayesian Quantile Variable Selection for Gene - Environment Interactions in Longitudinal Studies_ |
|---|---|

## Description

In this package, we provide a set of Bayesian quantile variable selection methods in the mixed effect models (random intercept and slope model, random intercept model) to dissect important gene - environment interactions for longitudinal studies. A Bayesian quantile regression has been adopted to accommodate data contamination and heavy-tailed distributions in the response/ phenotype. The default method (the proposed method) conducts variable selection by accounting the group level selection on the interaction effects under random intercept and slope model. In particular, the spike–and–slab priors are imposed on both individual and group levels to identify important main and interaction effects. In addition to the default method, users can also choose different selection structures for the interaction effects (group-level or individual-level), random intercept model, methods without spike–and–slab priors and non-robust methods. In total, _BayesQVGEL_ provides 16 different methods (8 robust and 8 non-robust) under both mixed effects models. Among them, robust methods with spike–and–slab priors and the robust method for both individual level selection and group level selection under both mixed effects models have been developed for the first time. Please read the Details below for how to configure the method used.

## Details

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

| | |
|---:|---|
| slope: | whether to use random intercept and slope model. |
| robust: | whether to use robust methods for modelling. |
| quant: | to specify different quantiles when using robust methods. |
| structure: | structure for interaction effects. |
| sparse: | whether to use the spike-and-slab priors to impose sparsity. |

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

## References

Fan, K., Wang, W. and Wu, C. (2022). Bayesian Quantile Variable Selection with Mixed Effects Model in the Longitudinal Study. (to be submitted)

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

Ren, J., Zhou, F., Li, X., Ma, S., Jiang, Y. and Wu, C. (2022). Robust Bayesian variable selection for gene-environment interactions. _Biometrics_, (in press) doi:10.1111/biom.13670

Ren, J., Zhou, F., Li, X., Ma, S., Jiang, Y. and Wu, C. (2020). roben: Robust Bayesian Variable Selection for Gene-Environment Interactions. R package version 0.1.1. `https://CRAN.R-project.org/package=roben`

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 `https://link.springer.com/protocol/10.1007/978-1-0716-0947-7_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

Ren, J., Zhou, F., Li, X., Wu, C. and Jiang, Y. (2019) spinBayes: Semi-Parametric Gene-Environment Interaction via Bayesian Variable Selection. R package version 0.1.0. `https://CRAN.R-project.org/package=spinBayes`

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

[BayesQVGEL](BayesQVGEL)

---

| BayesQVGEL | *fit a Bayesian quantile variable selection for gene - environment interactions in longitudinal studies* |

---

## Description

fit a Bayesian quantile variable selection for gene - environment interactions in longitudinal studies

## Usage

```
BayesQVGEL(
  y,
  e,
  C,
  g,
  w,
  k,
  iterations = 10000,
  burn.in = NULL,
  slope = TRUE,
  robust = TRUE,
  quant = 0.5,
```

```
    sparse = TRUE,
    structure = c("group", "individual")
)
```

## Arguments

| | |
|---|---|
| y | the matrix of response variable. The current version of BayesQVGEL only supports continuous response. |
| e | the matrix of a group of dummy environmental factors variables. |
| C | the matrix of time effects. |
| g | the matrix of predictors (genetic factors) without intercept. Each row should be an observation vector. |
| w | the matrix of interactions between genetic factors and environmental factors. |
| k | the total number of time points. |
| 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 | structure for interaction effects, two choices are available. "group" for selection on group-level only. "individual" for selection on individual-level only. |

## Details

Consider the data model described in "[data](data)":

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

where $\gamma_{2l}$ is the main effect of the $l$th genetic variant. The interaction effects is corresponding to the coefficient vector $\gamma_{3l} = (\gamma_{3l1}, \gamma_{3l2}, \ldots, \gamma_{3lm})^\top$.

When structure="group", group-level selection will be conducted on $||\gamma_{3l}||_2$. If structure="individual", individual-level selection will be conducted on each $\gamma_{3lq}$, $(q = 1, \ldots, m)$.

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 on individual and/or group levels to identify important main and interaction effects. Otherwise, Laplacian shrinkage will be used.

When robust=TRUE (default), the distribution of $\epsilon_i$ is defined as a 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, $\epsilon_{ij}$ follows a normal distribution.

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

## Value

an object of class 'BayesQVGEL' 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
fit=BayesQVGEL(y,e,C,g,w,k,structure=c("group"))
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 individual selection
fit=BayesQVGEL(y,e,C,g,w,k,structure=c("individual"))
fit$coefficient

## alternative: non-robust group selection
fit=BayesQVGEL(y,e,C,g,w,k,robust=FALSE, structure=c("group"))
fit$coefficient

## alternative: robust group selection under random intercept model
fit=BayesQVGEL(y,e,C,g,w,k,slope=FALSE, structure=c("group"))
fit$coefficient
```

---

data | *simulated data for demonstrating the features of BayesQVGEL*

---

## Description

Simulated gene expression data for demonstrating the features of BayesQVGEL.

## Format

The data object consists of seven components: y, e, C, 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 $i$th 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 $G_{ij} = (G_{ij1}, ..., G_{ijm})^T$. Also, we use $p$-dimensional vector $E_{ij}$ to denote the environment factors, where $E_{ij} = (E_{ij1}, ..., E_{ijp})^T$. $X_{ij} = (1, C_{ij})^T$, where $C_{ij}^T$ is a vector of time effects . $Z_{ij}$ is a $h \times 1$ covariate associated with random effects and $\alpha_i$ is a $h \times 1$ vector of random effects. At the beginning, the interaction effects is modeled as the product of genomics features and environment factors with 4 different levels. After representing the environment factors as three dummy variables, the identification of the gene by environment interaction needs to be performed as group level. Combing the genetics factors, environment factors and their interactions that associated with the longitudinal phenotype, we have the following mixed-effects model:

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

where $\gamma_1, \gamma_2, \gamma_3$ 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:

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

When $h = 1$, the model becomes a mixed-effects model with random intercept only.

## See Also

BayesQVGEL

## Examples

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

---

| selection | *Variable selection for a BayesQVGEL object* |
|---|---|

---

## Description

Variable selection for a BayesQVGEL object

## Usage

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

## Arguments

| | |
|---|---|
| `obj` | BayesQVGEL 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. (2022). Robust Bayesian variable selection for gene-environment interactions. *Biometrics*, (in press) 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

BayesQVGEL

## Examples

```
data(data)
## sparse
fit=BayesQVGEL(y,e,C,g,w,k,structure=c("group"))
selected=selection(fit,sparse=TRUE)
selected


## non-sparse
fit=BayesQVGEL(y,e,C,g,w,k,sparse=FALSE,structure=c("group"))
selected=selection(fit,sparse=FALSE)
selected
```

# Index