Package: lmmprobe (via r-universe)

October 12, 2026

Type Package

Title Sparse High-Dimensional Linear Mixed Modeling with a Partitioned Empirical Bayes ECM Algorithm

Version 0.1.0

Date 2026-02-27

Description Implements a partitioned Empirical Bayes Expectation Conditional Maximization (ECM) algorithm for sparse high-dimensional linear mixed modeling as described in Zgodic, Bai, Zhang, and McLain (2025) <doi:10.1007/s11222-025-10649-z>. The package provides efficient estimation and inference for mixed models with high-dimensional fixed effects.

License GPL (>= 2)

URL https://github.com/anjazgodic/lmmprobe

BugReports https://github.com/anjazgodic/lmmprobe/issues

Encoding UTF-8

RoxygenNote 7.3.3

Depends R (>= 3.5.0)

Imports Rcpp (>= 1.0.8.3), lme4 (>= 1.1-29), future.apply (>= 1.10.0)

LinkingTo Rcpp, RcppArmadillo

Suggests testthat (>= 3.0.0), knitr, rmarkdown, MASS

VignetteBuilder knitr

NeedsCompilation yes

Author Anja Zgodic [aut, cre], Ray Bai [aut] (ORCID: <https://orcid.org/0000-0001-7150-7229>), Jiajia Zhang [aut] (ORCID: <https://orcid.org/0000-0001-8281-8532>), Alex McLain [aut] (ORCID: <https://orcid.org/0000-0002-5475-0670>), Peter Olejua [aut] (ORCID: <https://orcid.org/0000-0002-2478-0908>)

Maintainer Anja Zgodic <anja.zgodic@gmail.com>

Repository https://cran.r-universe.dev

Date/Publication 2026-03-12 09:00:09 UTC

RemoteUrl https://github.com/cran/lmmprobe

RemoteRef HEAD

RemoteSha cbdfa860d269375cabd357b6c67cd06123d01e56

Contents

lmmprobe 2
SLE 4

Index **5**

`lmmprobe` *Sparse high-dimensional linear mixed modeling with PaRtitiOned empirical Bayes ECM (LMM-PROBE) algorithm.*

Description

Sparse high-dimensional linear mixed modeling with PaRtitiOned empirical Bayes ECM (LMM-PROBE) algorithm. Currently, the package offers functionality for two scenarios. Scenario 1: only a random intercept, each unit has the same number of observations; Scenario 2: a random intercept and a random slope, each unit has the same number of observations. We are actively expanding the package for more flexibility and scenarios.

Arguments

`Y`	A training-data matrix containing the outcome `Y`.
`Z`	A training-data matrix containing the sparse fixed-effect predictors on which to apply the lmmprobe algorithm. The first columns should be the "id" column.
`V`	A training-data matrix containing non-sparse predictors for the random effects. This matrix is currently only programmed for two scenarios. Scenario 1: only a random intercept, where V is a matrix with one column containing ID's and each unit has the same number of observations. Scenario 2: a random intercept and a random slope, where V is a matrix with two columns. The first column is ID and the second column is a continuous variable (e.g. time) for which a random slope is to be estimated. Each unit has the same number of observations.
`ID_data`	A factor vector of IDs for subjects in the training set.
`Y_test`	A testing-data matrix containing the outcome `Y`. Default is NULL.
`Z_test`	A testing-data matrix containing the sparse fixed-effect predictors. Default is NULL.
`V_test`	A testing-data matrix containing non-sparse predictors for the random effects, structured the same as `V`. Default is NULL.
`ID_test`	A factor vector of IDs for subjects in the testing set. Default is NULL.
`alpha`	Type I error; significance level.
`ep`	Value against which to compare convergence criterion, we recommend 0.05.

B	The number of groups to categorize estimated coefficients in to calculate correlation ρ. We recommend five.
adj	A factor multiplying Silverman's 'rule of thumb' in determining the bandwidth for density estimation, same as the 'adjust' argument of R's density function. Default is three.
maxit	Maximum number of iterations the algorithm will run for. Default is 10000.
cpus	The number of CPUS user would like to use for parallel computations. Default is four.
LR	A learning rate parameter r. Using zero corresponds to the implementation described in Zgodic et al.
C	A learning rate parameter c. Using one corresponds to the implementation described in Zgodic et al.
sigma_init	An initial value for the residual variance parameter. Default is NULL which corresponds to the sample variance of Y.

Value

A list containing:

beta MAP estimates of the posterior expectation of the prior mean (β) of the regression coefficients assuming $\gamma = 1$,

beta_var posterior variance of β,

gamma the posterior expectation of the latent γ variables,

preds predictions of Y,

PI_lower, PI_upper lower and upper prediction intervals for the predictions,

residual_var MAP estimate of the residual variance,

random_var MAP estimate of the random effect(s) variance,

random_intercept estimated random intercept terms,

random_slope estimated random slope terms, if applicable,

c_coefs calibration regression coefficients,

p_vals p-values for the fixed-effect coefficients,

count number of iterations until convergence.

References

Zgodic, A., Bai, R., Zhang, J. et al. (2025). Sparse high-dimensional linear mixed modeling with a partitioned empirical Bayes ECM algorithm. Stat Comput 35, 109. https://doi.org/10.1007/s11222-025-10649-z

Examples

```
set.seed(1)
n_subj <- 10
n_obs <- 5
N <- n_subj * n_obs
```

```
Y <- matrix(rnorm(N), ncol = 1)
Z <- matrix(rnorm(N * 20), nrow = N, ncol = 20)
V <- matrix(rep(1:n_subj, each = n_obs), ncol = 1)
ID_data <- rep(1:n_subj, each = n_obs)
result <- lmmprobe(Y = Y, Z = Z, V = V, ID_data = ID_data, maxit = 3)

data(SLE)
Y <- matrix(real_data[, "y"], ncol = 1)
Z <- real_data[, 4:ncol(real_data)]
V <- matrix(real_data[, "id"], ncol = 1)
ID_data <- as.numeric(as.character(real_data$id))
full_res <- lmmprobe(Y = Y, Z = Z, V = V, ID_data = ID_data)
```

SLE *Systemic Lupus Erythematosus (SLE) Gene Expression Data*

Description

A subset of longitudinal gene expression data from a pediatric Systemic Lupus Erythematosus (SLE) study. The full dataset contains 15,378 Illumina HumanHT-12 V4.0 probes; this subset includes 500 probes plus 16 clinical variables for a total of 519 columns. Loading this dataset creates an object named `real_data`.

Usage

```
data(SLE)
```

Format

A data frame with 353 observations on 519 variables:

id Subject ID (integer).

y Response variable (continuous).

intercept Intercept column (all ones).

ILMN_* 500 Illumina gene expression probes (numeric).

AGE, WBC, NEUTROPHIL_COUNT, ESR Continuous clinical predictors.

female, nonwhite Demographic indicators.

ARTHRITIS, URINARY_CASTS, HEMATURIA, PROTEINURIA, PYURIA, NEW_RASH, MUCOSAL_ULCERS, SLEDAI clinical components.

Source

Banchereau, R., Hong, S., Cantarel, B., et al. (2016). Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell*, 165(3), 551–565. doi:10.1016/j.cell.2016.05.057. Gene Expression Omnibus accession GSE65391.

Index

lmmprobe, 2

real_data (SLE), 4

SLE, 4