# Package ‘lmmprobe’

October 25, 2023

**Type** Package

**Title** Sparse high-dimensional linear mixed modeling with a partitioned empirical Bayes ECM algorithm (LMM-PROBE).

**Version** 1.0

**Date** 2023-09-07

**Author** Anja Zgodic, Alex McLain

**Maintainer** Anja Zgodic <azgodic@email.sc.edu>

**Description** Linear Mixed Modeling using the PROBE algorithm. This package is in development.

**License** GPL (>= 2)

**Encoding** UTF-8

**RoxygenNote** 7.2.0

**Depends** R (>= 3.5.0), Rcpp (>= 1.0.8.3), snow (>= 0.4.4), snowfall (>= 1.84.6.1), tidyr (>= 1.2.0), lme4 (>= 1.1.29)

**LinkingTo** Rcpp, RcppArmadillo

**NeedsCompilation** yes

## R topics documented:



---

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

---

### Description

Linear Mixed Modeling using the PROBE algorithm. This package is in development.

## Package Content

Index of help topics:

```
SLE                     High-dimensional dataset for sparse linear
                        mixed modeling.
lmmprobe                Sparse high-dimensional linear mixed modeling
                        with PaRtitiOned empirical Bayes ECM
                        (LMM-PROBE) algorithm.
lmmprobe-package        Sparse high-dimensional linear mixed modeling
                        with a partitioned empirical Bayes ECM
                        algorithm (LMM-PROBE).
```

## Maintainer

Anja Zgodic <azgodic@email.sc.edu>

## Author(s)

Anja Zgodic, Alex McLain

---

`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 matrix containing the outcome `Y`. |
| `Z` | A matrix containing the sparse fixed-effect predictors on which to apply the lmmprobe algorithm. |
| `V` | A 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. |
| `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 $\rho$. 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. |

## Value

A list of the output of the lmmprobe function, including

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

gamma the posterior expectation of the latent $\gamma$ variables,

preds predictions of $Y$,

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

sigma2_est 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.

## References

Zgodic, A. and McLain, A. C. (2023). Sparse high-dimensional linear mixed modeling with a partitioned empirical Bayes ECM algorithm. arXiv preprint arXiv:2310.12285.

## Examples

```
library(lmmprobe)
data(SLE)
ep <- 0.05
alpha <- 0.05
Y = SLE$Y
Z = SLE$Z
V = SLE$V
full_res <- lmmprobe(Y = Y, Z = Z, V = V, ep = ep, alpha = alpha)
```

---

SLE *High-dimensional dataset for sparse linear mixed modeling.*

---

## Description

This is the Systemic Lupus Erythematosus (SLE) data used in the data analysis section of the LMM-PROBE reference. The dataset has 309 observations, 103 subjects, and 15387 predictors.

## Usage

```
data(SLE)
```

## Format

A data frame with 309 observations and the following list elements:

- `Y` First element of list, corresponding to the outcome to use in the lmmprobe function.
- `Z` Second element of list, corresponding to the high-dimesional matrix of sparse predictors for fixed effects.
- `V` Third element of list, corresponding to the low-dimesional matrix of non-sparse predictors for random effects. This matrix has either only one ID column, or one ID column with an additional column for a continuous variable for which a random slope is to be estimated.

## References

Banchereau, R., Hong, S., Cantarel, B., Baldwin, N., Baisch, J., Edens, M., Cepika, A.-M., Acs, P., Turner, J., Anguiano, E., Vinod, P., Khan, S., Obermoser, G., Blankenship, D., Wakeland, E., Nassi, L., Gotte, A., Punaro, M., Liu, Y.-J., Banchereau, J., Rossello-Urgell, J., Wright, T., and Pascual, V. (2016), "Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients," Cell, 165, 551–565.

Zgodic, A. and McLain, A. C. (2023). Sparse high-dimensional linear mixed modeling with a partitioned empirical Bayes ECM algorithm. arXiv preprint arXiv:2310.12285.

## Examples

```
data(SLE)
Y = SLE$Y
Z = SLE$Z
V = SLE$V
```

# Index