

Package ‘apeglm’

April 11, 2018

Title Approximate posterior estimation for GLM coefficients

Version 1.0.3

Author Anqi Zhu, Joseph G. Ibrahim, Michael I. Love

Maintainer Anqi Zhu <anqizhu@live.unc.edu>

Description apeglm provides Bayesian shrinkage estimators for effect sizes for a variety of GLM models, using approximation of the posterior for individual coefficients.

VignetteBuilder knitr

Imports emdbook, SummarizedExperiment, GenomicRanges, methods, stats, utils

Suggests DESeq2, airway, BiocStyle, knitr, rmarkdown, testthat

License GPL-2

Encoding UTF-8

LazyData true

biocViews Sequencing, RNASeq, DifferentialExpression, GeneExpression, Bayesian

RoxygenNote 6.0.1

NeedsCompilation no

R topics documented:

apeglm	1
bbEstDisp	4
logLikNB	5
svalue	6
Index	7

apeglm *Approximate posterior estimation for GLM coefficients*

Description

apeglm provides Bayesian shrinkage estimators for effect sizes in GLM models, using approximation of the posterior for individual coefficients.

Usage

```
apeglm(Y, x, log.lik, param = NULL, coef = NULL, mle = NULL,
       no.shrink = FALSE, interval.type = c("laplace", "HPD", "credible"),
       interval.level = 0.95, threshold = NULL, contrasts, weights = NULL,
       offset = NULL, flip.sign = TRUE, prior.control, multiplier = 1,
       ngrid = 50, nse = 5, ngrid.nuis = 5, nse.nuis = 2, log.link = TRUE,
       param.sd = NULL, optim.method = "BFGS", bounds = c(-Inf, Inf))
```

Arguments

Y	the observations, which can be a matrix or SummarizedExperiment, with columns for samples and rows for "features" (e.g. genes in a genomic context). If Y is a SummarizedExperiment, apeglm will return, in addition to other list items, a GRanges or GRangesList ranges with the estimated coefficients as metadata columns.
x	design matrix, with intercept in the first column
log.lik	the log of likelihood function, specified by the user. For Negative Binomial distribution, user can use logLikNB provided within the package.
param	the other parameter(s) to be used in the likelihood function, e.g. the dispersion parameter for a negative binomial distribution. this can be a vector or a matrix (with columns as parameters)
coef	(optional) the index of the coefficient for which to generate posterior estimates (FSR and intervals)
mle	(optional) a 2 column matrix giving the MLE and its standard error of coef. this will be used to adapt the scale of the prior (empirical Bayes). This overrides the prior.scale specified by prior.control and sets no.shrink to all coefficients other than coef. Note that these MLE's and SE's should be on the natural log scale for a log link GLM.
no.shrink	logical, if TRUE, apeglm won't perform shrinkage (default is FALSE)
interval.type	(optional) can be "laplace", "HPD", or "credible", which specifies the type of Bayesian interval that the user wants to output; "laplace" represents the Laplace approximation of the posterior mode
interval.level	(optional) default is 0.95
threshold	(optional) a threshold for integrating posterior probabilities, see details under 'Value'. Note that this should be on the natural log scale for a log link GLM.
contrasts	(optional) contrast matrix, same number of rows as x
weights	(optional) weights matrix, same shape as Y
offset	(optional) offsets matrix, same shape as Y. Note that this should be on the natural log scale for a log link GLM.
flip.sign	whether to flip the sign of threshold value when MAP is negative, default is TRUE (threshold must then be positive)
prior.control	see Details
multiplier	a positive number, when the prior is adapted to the mle matrix provided, this parameter connects the scale of the estimated distribution of betas to the scale of the prior. the default value was chosen based on FSR and error analysis of simulated data
ngrid	the number of grid points for grid integration of intervals

<code>nse</code>	the number of Laplace standard errors to set the left and right edges of the grid around the MAP
<code>ngrid.nuis</code>	the number of grid points for nuisance parameters
<code>nse.nuis</code>	the number of Laplace standard errors to set the left and right edges of the grid around the MAP of the nuisance parameters
<code>log.link</code>	whether the GLM has a log link (default = TRUE)
<code>param.sd</code>	(optional) potential uncertainty measure on the parameter <code>param</code> . this should only be a vector, used when <code>param</code> specifies a single parameter
<code>optim.method</code>	the method passed to <code>optim</code>
<code>bounds</code>	the bounds for the numeric optimization

Details

`prior.control` is a list of parameters that will be passed to determine the prior distribution. Users are allowed to have a Normal prior on the intercept, and a t prior on the non-intercept coefficients (similar to `bayesglm` in the `arm` package. The following are defaults:

- `no.shrink = 1`: index of the coefficient(s) not to shrink
- `prior.mean = 0`: mean of t prior
- `prior.scale = 1`: scale of t prior
- `prior.df = 1`: df of t prior
- `prior.no.shrink.mean = 0`: mean of Normal
- `prior.no.shrink.scale = 15`: scale of Normal

So without specifying `prior.control`, the following is set inside `apeglm`:

```
prior.control <- list(no.shrink=1,prior.mean=0,prior.scale=1, prior.df=1,prior.no.shrink.mean
```

Note that the prior should be defined on the natural log scale for a log link GLM.

Value

a list of matrices containing the following components:

- `map`: a matrix of MAP estimates, columns for coefficients and rows for features
- `se`: a matrix of SE estimates, same shape as `map` (note: in future versions of `apeglm`, this slot is renamed `sd`).
- `prior.control`: a list with details on the prior
- `fsr`: a vector of the false sign rate for `coef`
- `interval`: a matrix of either HPD or credible interval for `coef`
- `thresh`: a vector of the posterior probability that the estimated parameter is smaller than the threshold value specified in `threshold` when MAP is positive (or greater than $-1 * \text{threshold}$ value when MAP is negative and `flip.sign` is TRUE)
- `diag`: a matrix of diagnostics
- `contrast.map`: a vector of MAP corresponding to the contrast when contrast is given
- `contrast.se`: a vector of SE corresponding to the contrast when contrast is given
- `ranges`: a `GRanges` or `GRangesList` with the estimated coefficients, if `Y` was a Summarized-Experiment.

Note that all parameters associated with coefficients, e.g. `map`, `se`, etc., are returned on the natural log scale for a log link GLM.

Examples

```

# Simulate RNA-Seq read counts data

# 5 samples for each of the two groups
# a total of 100 genes
n.per.group <- 5
n <- n.per.group * 2
m <- 100

# The design matrix includes one column of intercept
# and one column indicating samples that belong to the second group
condition <- factor(rep(letters[1:2], each = n.per.group))
x <- model.matrix(~condition)

# Specify the standard deviation of beta (LFC between groups)
beta.sd <- 2
beta.cond <- rnorm(m, 0, beta.sd)
beta.intercept <- runif(m, 2, 6)
beta.mat <- cbind(beta.intercept, beta.cond)

# Generate the read counts
mu <- exp(t(x) %*% t(beta.mat))
Y <- matrix(rnbinom(m*n, mu=mu, size=1/.1), ncol = n)

# Here we will use the negative binomial log likelihood
# which is an exported function. See 'logLikNB' for details.
# For the NB:
# 'param' is the dispersion estimate (1/size)
# 'offset' can be used to adjust for size factors (log of size factors)
param <- matrix(0.1, nrow = m, ncol = 1)
offset <- matrix(0, nrow = m, ncol = n)

# Shrinkage estimator of betas:
# (for adaptive shrinkage, 'apeglm' requires 'mle' coefficients
# estimated with another software, or by first running 'apeglm'
# setting 'no.shrink=TRUE'.)
res <- apeglm(Y = Y, x = x,
              log.lik = logLikNB,
              param = param,
              offset = offset,
              coef = 2)

head(res$map)
plot(beta.mat[,2], res$map[,2])
abline(0,1)

```

bbEstDisp

Simple line-search estimator for dispersion of a beta binomial

Description

Uses R's optimize function to find the maximum likelihood estimate of dispersion for a beta binomial distribution (theta for the dbetabinom function in the emdbook package). The counts, size,

and beta are matrices, such that each row could be treated as a beta-binomial GLM problem.

Usage

```
bbEstDisp(success, size, x, beta, minDisp, maxDisp)
```

Arguments

success	the observed successes (a matrix)
size	the total trials (a matrix)
x	the design matrix, as many rows as columns of success and size
beta	a matrix of MLE coefficients, as many rows as success and size
minDisp	the minimum dispersion value
maxDisp	the maximum dispersion value

Value

a vector of estimated dispersions

Examples

```
library(emdbook)
n <- 100
m <- 100
size <- matrix(rnbinom(n*m, mu=100, size=10), ncol=m)
success <- matrix(rbetabinom(n*m, prob=.5, size=size, theta=100), ncol=m)
x <- matrix(rep(1,m), ncol=1)
beta <- matrix(rep(0,n), ncol=1)
theta <- bbEstDisp(success, size, x, beta, 1, 500)
summary(theta)
```

logLikNB

Log likelihood for Negative Binomial

Description

This is a simple function to be passed to `apeglm` as a log likelihood for the negative binomial distribution.

Usage

```
logLikNB(y, x, beta, param, offset)
```

Arguments

y	the counts
x	a design matrix
beta	the coefficient vector (natural log scale)
param	the overdispersion (alpha in DESeq2 notation)
offset	an offset matrix (natural log scale)

Value

the log likelihood for each sample in y

Examples

```
# this function is used by 'apeglm' to specify
# a negative binomial log likelihood.
# so it's only passed as an argument, not for use on its own.
# we can show its output nevertheless:

y <- rnbinom(10, mu=100, size=1/.1)
x <- cbind(rep(1,10),rep(0:1,each=5))
beta <- c(log(100),0)
param <- .1
offset <- rep(0, 10)
logLikNB(y, x, beta, param, offset)
```

svalue

Local FSR to svalue

Description

Convert local FSR to svalue by taking the cumulative mean of increasing local FSRs. This is used within `apeglm` to generate the svalue table, but provided as convenience function.

Usage

```
svalue(lfsr)
```

Arguments

lfsr the local false sign rates

Value

s-values

Examples

```
# Example 1 -- simulated local FSR data
local.fsr <- runif(1000)
sval <- svalue(local.fsr)

# Example 2 -- first runs example from 'apeglm'
example("apeglm")
local.fsr <- res$fsr[1, ]
sval <- svalue(local.fsr)
```

Index

`apeglm`, 1

`bbEstDisp`, 4

`logLikNB`, 5

`svalue`, 6