Package 'GMeta'

September 28, 2017

Title Implements generalized meta-analysis

Version 0.1

Description GMeta implements the generalized meta-analysis using IRWLS algorithm.

Depends R (>= 3.3.1)

Imports MASS, magic

License GPL

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1.9000

R topics documented:

`GMeta` *Implementing Generalized Meta-analysis*

Description

Generalized Meta-analysis(GMeta) is an approach for combining information on multivariate regression parameters across multiple different studies which have different, but, possibly overlapping information on subsets of covariates. GMeta implements the generalized meta-analysis using IRWLS algorithm.

Usage

```
GMeta(study_info, ref_dat, model, variable_intercepts = FALSE,
  control = list(...))
```

Arguments

`study_info` a list of lists containing information about the studies; the main list contains a list for each study, which must have the fields:

- "Coeff": a named numeric vector containing the estimates of regression parameters (including intercept) where the names identify the covariates. For example, names(study_info$Coeff) <- c("(Intercept)", "Age", "Height", "Weight").
- "Covariance": a matrix containing an estimate of variance-covariance matrix of the regression parameters. This can be NULL if the "Sample_size" is provided.
- "Sample_size": a numeric containing sample size of the study. This can be NULL if the "Covariance" is provided.

`ref_dat` a data matrix containing all the distinct covariates across studies from a reference set of individuals. This is used for estimating joint distribution of the covariates. The data matrix must have the vector of ones as its first column. The column names of the data matrix should match the names of the covariates from the studies.

`model` a description of the type of regression model; this is a character string naming the regression model. The current version is for "logistic" and "linear".

`variable_intercepts` an optional logical (applicable only when the model is "logistic"); if TRUE, the intercepts of the true models for each of the studies are assumed to be different. Default is FALSE.

`control` an optional list containing the epsilon (numeric) and maxiter (numeric) needed for convergence of the algorithm. Default epsilon and maximum iterations are 1e-06 and 1000, respectively. For creating a control argument for GMeta, see `GMeta.control`.

Details

Generalized Meta-analysis (GMeta) is a tool that allows researchers to quickly build models for multivariate meta-analysis in the presence of disparate covariate information across studies. It is implemented based on mainly two input arguments:

- Information on the model parameters from each of the studies.
- Reference data for estimation of the joint distribution of all the distinct covariates across studies.

The software provides flexibility to the users to choose the intercepts to be different (when the model is logistic) across studies through the input argument, variable_intercepts. It also allows estimation of the regression parameters, only from the sample sizes of the studies when it is difficult to obtain estimate of the variance-covariance matrices.

Note: GMeta will not work if both the estimates of the covariance matrix and the sample size are NULL.

When the model is "linear", it is assumed that the outcome is standardized to have unit variance. For more details on the IRWLS, see References.

Value

An object of class "GMeta" is a list containing GMeta estimate, its variance-covariance matrix and estimates the residual variance in the case of "linear" model .

`Est.coeff`	a numeric vector containing the estimated regression coefficients of the maximal model using optimal weighting matrix.
`Est.var.cov`	a matrix containing estimate of variance-covariance matrix of the corresponding GMeta estimator.
`Res.var`	a numeric containing the residual variance of the maximal model when it is linear. It is calculated from the formula : $1 - \hat{\beta}^T_{GMeta} var(X) \hat{\beta}_{GMeta}$ which is derived by assuming the outcomes to have unit variance. $var(X)$ is calculated from reference data. Res.var is NA when the model is "logistic".
`iter`	a numeric containing the number of iterations used in the algorithm
`call`	the matched call

The function `GMeta.summary` can be used to obtain a summary of the results obtained from GMeta.

Author(s)

Prosenjit Kundu, Runlong Tang and Nilanjan Chatterjee.

References

Tang, R., Kundu, P. and Chatterjee, N. (2017) Generalized Meta-Analysis for Multivariate Regression Models Across Studies with Disparate Covariate Information. arXiv:1708.03818v1 [stat.ME].

See Also

`GMeta.summary`, `study_info`, `reference_data`

Examples

```
# This example shows the GMeta implementation on a simulated data set for logistic regression
data(reference_data)
head(reference_data)
data(study_info)
head(study_info)
model <- "logistic"
# When the true intercept parameters of the studies are different.
result_diff <- GMeta(study_info, reference_data, model, variable_intercepts = TRUE)
print(result_diff)
# When the true intercept parameters of the studies are same.
result_same <- GMeta(study_info, reference_data, model)
print(result_same)
```

GMeta.control *Auxiliary for controlling the IRWLS algorithm*

Description

This is an auxiliary function for the iteratively reweighted least squares algorithm for GMeta. This is used internally by the myoptim function, but can be used by the user to create a control argument in the GMeta function

Usage

```
GMeta.control(epsilon = 1e-06, maxit = 1000)
```

Arguments

epsilon	a positive numeric indicating converegence tolerence; the algorithm stops when the absolute difference between the estimates in current and previous step is less than epsilon, i.e, $\|estimate_{new} - estimate_{old}\| < \epsilon$
maxit	a positive numeric indicating the maximum number of iterations to be used in the algorithm. Default is 1000.

Value

A list with components named as the arguments.

Examples

```
control <- GMeta.control(1e-08, 100)
```

GMeta.summary *Summarizing Generalized Meta Analysis*

Description

This function prints the summary of GMeta results.

Usage

```
GMeta.summary(object, signi_digits = 3)
```

Arguments

object	an object of class "GMeta"
signi_digits	an optional numeric indicating the number of significant digits to be shown in the summary. Default is 3.

Examples

```
# This example shows how to obtain the summary of GMeta object.
data(reference_data)
data(study_info)
model <- "logistic"
result_diff <- GMeta(study_info, reference_data, model, variable_intercepts = TRUE)
GMeta.summary(result_diff, signi_digits = 4)
result_same <- GMeta(study_info, reference_data, model)
GMeta.summary(result_same)
```

---

reference_data *Simulated reference data on a set of 50 individuals.*

Description

This is a data set containing simulated covariates, Age, Height and Weight, from a multivariate normal distribution with zero mean vector and variance-covariance matrix, matrix(c(1,0.3,0.6,0.3,1,0.1,0.6,0.1,1),3,3). In other words, the covariates in the underlying population have unit variance with correlation between Age and Height, Height and Weight, and, Height and Age are 0.3, 0.1 and 0.6, respectively. The first column of the data set contains vector of ones indicating the intercept.

Usage

```
data(reference_data)
```

Format

A data matrix containing 50 rows and 4 columns.

References

See the simulation section of the paper by Tang, R., Kundu, P. and Chatterjee, N. (2017) Generalized Meta-Analysis for Multivariate Regression Models Across Studies with Disparate Covariate Information. arXiv:1708.03818v1 [stat.ME].

Examples

```
data(reference_data)
```

---

study_info *Simulated data on study information based.*

Description

The underlying model is assumed to be logistic. Then, outcomes are simulated from a logistic model (maximal) where the covariates(Age, Height and Weight) are simulated from the same multivariate normal distribution (see `reference_data`). In this example, three different studies are considered. Study 1, study 2 and, study 3 have information on Age and Height, Height and Weight, and, Weight and Age, respectively. Sample sizes of the three studies are taken to be 300, 500 and 1000, respectively. Estimates of the regression coefficients are obtained by fitting a logistic model using the glm function. For study 2, estimate of the variance-covariance matrix is set to NULL and the sample size of the corresponding study is provided to reflect the scenario where it is difficult for the user to provide variance-covariance matrix.

NOTE: This is a simulated data on summary-level information (estimates of regression coefficients, variance-covariance matrices) across studies. In real scenario, the users, usually, have those information. They have to just put that information in study_info argument appropriately.

Usage

```
data(study_info)
```

Format

A list of lists containing study information on 3 studies.

References

See the simulation section of the paper by Tang, R., Kundu, P. and Chatterjee, N. (2017) Generalized Meta-Analysis for Multivariate Regression Models Across Studies with Disparate Covariate Information. arXiv:1708.03818v1 [stat.ME].

For reference_data, see `reference_data` .

Examples

```
data(study_info)
```

Index