Package ‘brlrmr’

July 17, 2017

Title  Bias Reduction with Missing Binary Response
Version  0.1.5
Date  2016-11-23
Description  Provides two main functions, il() and fil(). The il() function implements the EM algorithm developed by Ibrahim and Lipsitz (1996) <DOI:10.2307/2533068> to estimate the parameters of a logistic regression model with the missing response when the missing data mechanism is nonignorable. The fil() function implements the algorithm proposed by Maity et. al. (2017+) <https://github.com/arnabkrmaity/brlrmr> to reduce the bias produced by the method of Ibrahim and Lipsitz (1996) <DOI:10.2307/2533068>.

Depends  R (>= 2.10)
Imports  boot, brglm, MASS, profileModel, Rcpp, stats
License  GPL-3
Encoding  UTF-8
LazyData  true
RoxygenNote  6.0.1
NeedsCompilation  no
Author  Arnab Maity [aut, cre],
        Vivek Pradhan [ctb],
        Ujjwal Das [ctb]
Maintainer  Arnab Maity <akumar@stat.tamu.edu>
Repository  CRAN
Date/Publication  2017-07-17 17:51:55 UTC

R topics documented:

em.fil .................................................. 2
em.fil.interaction ........................................ 3
em.il .................................................. 3
em.il.interaction ........................................ 4
fil .................................................. 5
il .................................................. 7
description

It is called by the main function fil and is for internal use.

usage

em.fil(parameter, X, full.missing.data, observed.data, full.data, family)

arguments

parameter The starting values of the parameters as (β, α) where β is the parameters of original model and α is the for the missing data model.
X The design matrix with the intercept column.
full.missing.data The augmented response with design matrix and missing indicator 1 for missing data.
observed.data The observed response with design matrix and missing indicator 0 for observed data.
full.data The observed response, augmented response with corresponding design matrix and missing indicator 0 for observed data and 1 for missing data.
family as in glm. brlrmr currently supports only the "binomial" family with links

references

Bias Reduction in Logistic Regression with Missing Responses when the Missing Data Mechanism is Nonignorable.
Description

It is called by the main function fil and is for internal use.

Usage

em.fil.interaction(parameter, X, full.missing.data, observed.data, full.data, k, family)

Arguments

parameter: The starting values of the parameters as $(\beta, \alpha)$ where $\beta$ is the parameters of original model and $\alpha$ is the for the missing data model.

X: The design matrix with the intercept column.

full.missing.data: The augmented response with design matrix and missing indicator 1 for missing data.

observed.data: The observed response with design matrix and missing indicator 0 for observed data.

full.data: The observed response, augmented response with corresponding design matrix and missing indicator 0 for observed data and 1 for missing data.

k: If interaction is present in the missing data model, then the k is the column number of covariate matrix which has interaction with the response.

family: as in glm. brlrmr currently supports only the “binomial” family with links

References

Bias Reduction in Logistic Regression with Missing Responses when the Missing Data Mechanism is Nonignorable.

Description

It is called by the main function fil and is for internal use.

Usage

em.il(parameter, X, full.missing.data, observed.data, full.data, family)
Arguments

parameter The starting values of the parameters as \((\beta, \alpha)\) where \(\beta\) is the parameters of original model and \(\alpha\) is the for the missing data model.

X The design matrix with the intercept column.

full.missing.data The augmented response with design matrix and missing indicator 1 for missing data.

observed.data The observed response with design matrix and missing indicator 0 for observed data.

full.data The observed response, augmented response with corresponding design matrix and missing indicator 0 for observed data and 1 for missing data.

family as in \texttt{glm}. \texttt{brlrmr} currently supports only the “binomial” family with links

References

Bias Reduction in Logistic Regression with Missing Responses when the Missing Data Mechanism is Nonignorable.

em.il.interaction

Description

It is called by the main function \texttt{fil} and is for internal use.

Usage

em.il.interaction(parameter, X, full.missing.data, observed.data, full.data, k, family)

Arguments

parameter The starting values of the parameters as \((\beta, \alpha)\) where \(\beta\) is the parameters of original model and \(\alpha\) is the for the missing data model.

X The design matrix with the intercept column.

full.missing.data The augmented response with design matrix and missing indicator 1 for missing data.

observed.data The observed response with design matrix and missing indicator 0 for observed data.

full.data The observed response, augmented response with corresponding design matrix and missing indicator 0 for observed data and 1 for missing data.

k If interaction is present in the missing data model, then the \(k\) is the column number of covariate matrix which has interaction with the response.

family as in \texttt{glm}. \texttt{brlrmr} currently supports only the “binomial” family with links
References

Bias Reduction in Logistic Regression with Missing Responses when the Missing Data Mechanism is Nonignorable.

Description

This provides the estimates using IL method and FIL method as described in the reference.

Usage

fil(formula, data, parameter = NULL, family = binomial, alpha = 0.05, interaction = FALSE, k = NULL, na.action)

Arguments

- formula as in \texttt{lm}. The missing values of response are NA.
- data as in \texttt{lm}. The first column of data is binary missing response. The missing observations are denoted by NA. The rest of the columns are covariates or explanatory variables.
- parameter The starting values of the parameters as \((\beta, \alpha)\) where \(\beta\) is the parameters of original model and \(\alpha\) is the for the missing data model.
- family as in \texttt{glm}. \texttt{brlr} currently supports only the “binomial” family with links.
- alpha This is used for upper 100(1 - alpha)% point of standard Normal distribution. The default is 1.96.
- interaction TRUE or FALSE, whether to consider interaction in the missing data model. Currently only one interaction between response and covariates is supported. FALSE by default.
- k Which covariate has interaction with response. Takes integer values. User must assign a value if interaction = TRUE.
- na.action as in \texttt{lm}. Always set to \texttt{na.pass}. Note that setting any other value to na.action will remove the NA’s from response and hence will break the code as this package is only intended for missing response data.

Value

- \(n\) number of observations.
- \(n\_missing\) the number of missing observations.
- missing.proportion proportion of missing observations.
- beta.hat parameter estimate of logistic regression of \(y\) on \(x\) using FIL method.
beta.se.hat  standard error using FIL method.
z.value  Wald Z value using FIL method.
p.value  p value using FIL method.

significance.beta.firth  indicator output whether regressors are significant using FIL method, 1 if significant and 0 if not significant.

LCL  Lower Confidence Limits of 100(1 - alpha)% Confidence Intervals.
UCL  Upper Confidence Limits of 100(1 - alpha)% Confidence Intervals.

alpha.hat  parameter estimate due to missing model using FIL.
alpha.se.hat  standard error of the them.
z.value.alpha  Wald Z value for them.
p.value.alpha  p values for them.

References

Bias Reduction in Logistic Regression with Missing Responses when the Missing Data Mechanism is Nonignorable.

Examples

```r
## Not run:
#########################################
#### Simulated Example ###############
#########################################
data(simulated.data) # load simulated data

# parameter definition
beta0 <- 1
beta1 <- 1
beta2 <- 1
beta3 <- 1
beta4 <- 1

# parameter definition for missing indicator
alpha0 <- -1.1
alpha1 <- -1
alpha2 <- -1
alpha3 <- -1
alpha4 <- -1
alpha5 <- -1

parameter <- c(beta0, beta1, beta2, beta3, beta4,
               alpha0, alpha1, alpha2, alpha3, alpha4, alpha5)

fil(y ~ x1 + x2 + x3 + x4, data = simulated.data, parameter,
    family = binomial(link = "logit"), na.action = na.pass)

## End(Not run)
```
### Real data example with separation

data(nhanes)  # load nhanes data
data(incontinence)  # load nhanes data

```r
# load nhanes data
fil(hyp ~ age2 + age3, data = nhanes, family = binomial(link = "logit"), na.action = na.pass)

# load nhanes data
fil(y ~ x1 + x2 + x3, data = incontinence, family = binomial(link = "logit"), na.action = na.pass)
```

---

#### Description

This provides the estimates using IL method as described in the reference.

#### Usage

```r
il(formula, data, parameter = NULL, family = binomial, alpha = 0.05, interaction = FALSE, k = NULL, na.action)
```

#### Arguments

- **formula** as in `lm`. The missing values of response are NA.
- **data** as in `lm`. The first column of data is binary missing response. The missing observations are denoted by NA. The rest of the columns are covariates or explanatory variables.
- **parameter** The starting values of the parameters as \((\beta, \alpha)\) where \(\beta\) is the parameters of original model and \(\alpha\) is the for the missing data model.
- **family** as in `glm`. `brlrmr` currently supports only the "binomial" family with links.
- **alpha** This is used for upper 100(1 - alpha)% point of standard Normal distribution. The default is 1.96.
- **interaction** TRUE or FALSE, whether to consider interaction in the missing data model. Currently only one intercation between response and covariates is supported. FALSE by default.
- **k** Which covariate has interaction with response. Takes integer values. User must assign a value if interaction = TRUE.
- **na.action** as in `lm`. Always set to na.pass. Note that setting any other value to na.action will remove the NA’s from response and hence will break the code as this package is only intended for missing response data.
**Value**

- `n`: number of observations.
- `nmissing`: the number of missing observations.
- `missing.proportion`: proportion of missing observations.
- `beta.hat`: parameter estimate of logistic regression of y on x using IL method.
- `beta.se.hat`: standard error using IL method.
- `z.value`: Wald Z value using IL method.
- `p.value`: p value using IL method.
- `significance.beta`: is indicator output whether regressors are significant using IL method, 1 if significant and 0 if not significant.
- `LCL`: Lower Confidence Limits of 100(1 - alpha)% Confidence Intervals.
- `UCL`: Upper Confidence Limits of 100(1 - alpha)% Confidence Intervals.
- `alpha.hat`: parameter estimate due to missing model using IL.
- `alpha.se.hat`: standard error of the them.
- `z.value.alpha`: Wald Z value for them.
- `p.value.alpha`: p values for them.
- `sep`: separation indicator = 1 if separation, = 0 otherwise

**References**


**Examples**

```r
## Not run:
#############################################################
# Simulated Example
#############################################################
# Simulated data
simulated.data <- read.csv("simulated.data")
# load simulated data

# parameter definition
beta0 <- 1
beta1 <- 1
beta2 <- 1
beta3 <- 1
beta4 <- 1

# parameter definition for missing indicator
alpha0 <- -1.1
alpha1 <- -1
alpha2 <- 1
alpha3 <- 1
alpha4 <- 1
```
incontinence

alpha5 <- -1

parameter <- c(beta0, beta1, beta2, beta3, beta4,
 alpha0, alpha1, alpha2, alpha3, alpha4, alpha5)

il(y ~ x1 + x2 + x3 + x4, data = simulated.data, parameter,
 family = binomial(link = "logit"), na.action = na.pass)

## End(Not run)

## Not run:

#########################################################

Real data example with separation
#########################################################

data(nhanes) # load nhanes data

il(hyp ~ age2 + age3, data = nhanes, family = binomial(link = "logit"), na.action = na.pass)
# IL method encounters separation

## End(Not run)

incontinence Incontinence example.

Description

A urinary incontinence study.

Usage

incontinence

Format

A data frame with 21 observations on the following 4 variables:

- **y** Response (1 = continent, 0 = otherwise)
- **x1** Lower urinary tract measure
- **x2** Lower urinary tract measure
- **x3** Lower urinary tract measure

Source

### nhanes

*Subset of original NHANES data used in mice package.*

**Description**

A small data set with missing values.

**Usage**

nhanes

**Format**

A data frame with 25 observations on the following 2 variables:

- **hyp**: Hypertensive (0 = no, 1 = yes)
- **age2**: Age group (1 = 40-59, 0 = otherwise)
- **age3**: Age group (1 = 60+, 0 = otherwise)

**Source**


### simulated.data

*We simulate this data for the purpose of illustration of the package*

**Description**

A dataset containing the 100 observations and 4 covariates. The covariates are generated from standard normal distribution. The missing binary response is generated using the simulation process as described in the reference.

**Usage**

simulated.data

**Format**

A data frame with 100 observations with 28 missing responses:
Index

*Topic **datasets**
  - incontinence, 9
  - nhanes, 10
  - simulated.data, 10

- em.fil, 2
- em.fil.interaction, 3
- em.il, 3
- em.il.interaction, 4

- fil, 5
- glm, 2–5, 7
- il, 7
- incontinence, 9
- lm, 5, 7
- nhanes, 10
- simulated.data, 10