Package ‘CRTgeeDR’

July 9, 2017

Title Doubly Robust Inverse Probability Weighted Augmented GEE Estimator

Version 2.0

Maintainer Melanie Prague <mprague@hsph.harvard.edu>

Description Implements a semi-parametric GEE estimator accounting for missing data with Inverse-probability weighting (IPW) and for imbalance in covariates with augmentation (AUG). The estimator IPW-AUG-GEE is Doubly robust (DR).

License GPL (>= 2)

Depends Matrix, MASS, ggplot2, grDevices, graphics, stats, methods

LazyData true

NeedsCompilation no

Author Melanie Prague [aut, cre],
    Paul Gilbert [ctb] (Author of R package numDeriv, which has been acknowledged in numDeriv.R),
    Ravi Varadhan [ctb] (Author of R package numDeriv, which has been acknowledged in numDeriv.R),
    Ming Wang [ctb] (Author of R package geesmv, which has been acknowledged in getFay.R),
    Lee McDaniel [ctb] (Author of R package geeM, which has been modified and references in multiple R files),
    Nick Henderson [ctb] (Author of R package geeM, which has been modified and references in multiple R files)

Repository CRAN

Date/Publication 2017-07-09 07:47:26 UTC

R topics documented:

<table>
<thead>
<tr>
<th>Topic</th>
<th>R Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRTgeeDR</td>
<td>2</td>
</tr>
<tr>
<td>data.sim</td>
<td>2</td>
</tr>
<tr>
<td>fitted.CRTgeeDR</td>
<td>3</td>
</tr>
<tr>
<td>geeDREstimation</td>
<td>3</td>
</tr>
<tr>
<td>getCI</td>
<td>8</td>
</tr>
</tbody>
</table>
The CRTgeeDR package allows you to estimate parameters in a regression model (with possibly a link function). It allows treatment augmentation and IPW for missing data alone.

The only function you’re likely to need from CRTgeeDR is geeDRestimation. Otherwise refer to the help documentation.

A dataset containing the HIV risk scores and presence of risky behaviors (yes/no) and other covarites of 10000 subjects among 100 communities. The variables are as follows:

- IDPAT subject id
- CLUSTER cluster id
- TRT treatment status, 1 is received STI/HIV intervention
- X1 A covariate following a N(0,1)
- JOB employment status
- MARRIED marital status
• AGE age
• HIV.KNOW Score for HIV knowledge
• RELIGION religiosity score
• OUTCOME Binary outcome - 1 if the subject is at high risk of HIV infection, 0 otherwise. NA if missing.
• MISSING 1 if the outcome is missing - 0 otherwise.

fitted.CRTgeeDR

Fit CRTgeeDR object.

Description
Fit CRTgeeDR object to a dataset

Usage
```r
## S3 method for class 'CRTgeeDR'
fitted(object, ...)
```

Arguments
- object CRTgeeDR object
- ... ignored

geedrestimation

Doubly Robust Inverse Probability Weighted Augmented GEE Estimator

Description
This function implements a GEE estimator. It implements classical GEE, IPW-GEE, augmented GEE and IPW-Augmented GEE (Doubly robust).

Usage
geeDREstimation(formula, id, data = parent.frame(), family = gaussian, corstr = "independence", Mv = 1, weights = NULL, aug = NULL, pi.a = 1/2, corr.mat = NULL, init.beta = NULL, init.alpha = NULL, init.phi = 1, scale.fix = FALSE, sandwich = TRUE, maxit = 20, tol = 1e-05, print.log = FALSE, typeweights = "VW", nameTRT = "TRT", model.weights = NULL, model.augmentation.trt = NULL, model.augmentation.ctrl = NULL, stepwise.augmentation = FALSE, stepwise.weights = FALSE, nameMISS = "MISSING", nameY = "OUTCOME", sandwich.nuisance = FALSE, fay.adjustment = FALSE, fay.bound = 0.75)
Arguments

**formula**
An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.

**id**
A vector which identifies the clusters. The length of "id" should be the same as the number of observations. Data are assumed to be sorted so that observations on a cluster are contiguous rows for all entities in the formula.

**data**
An optional data frame, list or environment (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which CRTgeeDR is called.

**family**
A description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)

**corstr**
A character string specifying the correlation structure. The following are permitted: "independence", "exchangeable", "ar1", "unstructured" and "userdefined".

**Mv**
For "m-dependent", the value for m.

**weights**
A vector of weights for each observation. If an observation has weight 0, it is excluded from the calculations of any parameters. Observations with a NA anywhere (even in variables not included in the model) will be assigned a weight of 0.

**aug**
A list of vectors (one for A=1 treated, one for A=0 control) for each observation representing E(Y|X,A=a).

**pi.a**
A number, Probability of treatment attribution P(A=1)

**corr.mat**
The correlation matrix for "fixed". Matrix should be symmetric with dimensions >= the maximum cluster size. If the correlation structure is "userdefined", then this is a matrix describing which correlations are the same.

**init.beta**
An optional vector with the initial values of beta. If not specified, then the intercept will be set to InvLink(mean(response)). init.beta must be specified if not using an intercept.

**init.alpha**
An optional scalar or vector giving the initial values for the correlation. If provided along with Mv>1 or unstructured correlation, then the user must ensure that the vector is of the appropriate length.

**init.phi**
An optional initial overdispersion parameter. If not supplied, initialized to 1.

**scale.fix**
If set to TRUE, then the scale parameter is fixed at the value of init.phi.

**sandwich**
If set to TRUE, the sandwich variance is provided together with the naive estimator of variance.

**maxit**
Maximum number of iterations.

**tol**
Tolerance in calculation of coefficients.

**print.log**
If set to TRUE, a report is printed.

**typeweights**
A character string specifying the weights implementation. The following are permitted: "GENMOD" for $W^{1/2}V^{-1}W^{1/2}$, "WV" for $V^{-1}W$
nameTRT Name of the variable containing information for the treatment
model.weights an object of class "formula" (or one that can be coerced to that class): a symbolic
description of the model to be fitted for the propensity score. Must model the
probability of being observed.
model.augmentation.trt
an object of class "formula" (or one that can be coerced to that class): a symbolic
description of the model to be fitted for the outcome model for the treated group
(A=1).
model.augmentation.ctrl
an object of class "formula" (or one that can be coerced to that class): a symbolic
description of the model to be fitted for the outcome model for the control group
(A=0).
stepwise.augmentation
if set to TRUE, a stepwise for the augmentation model is performed during the
fit of the augmentation model for the OM
stepwise.weights
if set to TRUE, a stepwise for the propensity score is performed during the fit of
the augmentation model for the OM
nameMISS Name of the variable containing information for the Missing indicator
nameY Name of the variable containing information for the outcome
sandwich.nuisance
if set to TRUE, the nuisance adjusted sandwich variance is provided.
fay.adjustment
if set to TRUE, the small-sample nuisance adjusted sandwich variance with
Fay's adjustement is provided.
fay.bound
if set to 0.75 by default, bound value used for Fay's adjustement.

Details

The estimator is founds by solving:

\[ 0 = \sum_{i=1}^{M} \left[ D_i^T V_i^{-1} W_i(X_i, A_i, \eta_W) (Y_i - B(X_i, A_i, \eta_B)) \right. \\
+ \left. \sum_{a=0,1} p^a(1-p)^{1-a} D_i^T V_i^{-1} \left( B(X_i, A_i = a, \eta_B) - \mu_i(\beta, A_i = a) \right) \right] \]

where \( D_i = \frac{\partial \mu_i(\beta, A_i)}{\partial \beta} \) is the design matrix, \( V_i \) is the covariance matrix equal to \( U_i^{1/2} C(\alpha) U_i^{1/2} \) with \( U_i \) a diagonal matrix with elements \( \text{var}(y_{ij}) \) and \( C(\alpha) \) is the working correlation structure with non-diagonal terms \( \alpha \). Parameters \( \alpha \) are estimated using simple moment estimators from the Pearson residuals. The matrix of weights \( W_i(X_i, A_i, \eta_W) = \text{diag} [R_{ij}/\pi_{ij}(X_i, A_i, \eta_W)]_{i=1,...,n} \), where \( \pi_{ij}(X_i, A_i, \eta_W) = P(R_{ij}|X_i, A_i) \) is the Propensity score (PS). The function \( B(X_i, A_i = a, \eta_B) \), which is called the Outcome Model (OM), is a function linking \( Y_{ij} \) with \( X_i \) and \( A_i \). The \( \eta_B \) are nuisance parameters that are estimated. The estimator is most efficient if the OM is equal
to \( E(Y_i|X_i, A_i = a) \) The estimator denoted \( \hat{\beta}_{aug} \) is found by solving the estimating equation. Although analytic solutions sometimes exist, coefficient estimates are generally obtained using an
iterative procedure such as the Newton-Raphson method. Automatic implementation is such that, \( \hat{\eta}_W \) in \( W_i(X_i, A_i, \hat{\eta}_W) \) are obtained using a logistic regression and \( \hat{\eta}_B \) in \( B(X_i, A_i, \hat{\eta}_B) \) are obtained using a linear regression.

The variance of \( \hat{\beta}_{aug} \) is estimated by the sandwich variance estimator. There are two external sources of variability that need to be accounted for: estimation of \( \eta_W \) for the PS and of \( \eta_B \) for the OM. We denote \( \Omega = (\beta, \eta_W, \eta_B) \) the estimated parameters of interest and nuisance parameters. We can stack estimating functions and score functions for \( \Omega \):

\[
U_i(\Omega) = \left( \Phi_i(Y_i, X_i, A_i, \beta, \eta_W, \eta_B) \right)
\[
S^W_i(X_i, A_i, \eta_W)
\]
\[
S^B_i(X_i, A_i, \eta_B)
\]

where \( S^W_i \) and \( S^B_i \) represent the score equations for patients in cluster \( i \) for the estimation of \( \eta_W \) and \( \eta_B \) in the PS and the OM. A standard Taylor expansion paired with Slutsky’s theorem and the central limit theorem give the sandwich estimator adjusted for nuisance parameters estimation in the OM and PS:

\[
Var(\Omega) = E \left[ \frac{\partial U_i(\Omega)}{\partial \Omega} \right]^{-1} E \left[ U_i(\Omega)U_i^T(\Omega) \right] E \left[ \frac{\partial U_i(\Omega)}{\partial \Omega} \right]^{-1}.
\]

Value

An object of type 'CRTgeeDR'

Sbeta Final values for regressors estimates

- $\phi$ scale parameter estimate
- $\alpha$ Final values for association parameters in the working correlation structure when exchangeable
- $\text{coefnames}$ Name of the regressors in the main regression
- $\text{niter}$ Number of iteration done by the algorithm before convergence
- $\text{converged}$ convergence status
- $\text{var.naiv}$ Variance of the estimates model based (naive)
- $\text{var}$ Variance of the estimates sandwich
- $\text{var.nuisance}$ Variance of the estimates nuisance adjusted sandwich
- $\text{var.fay}$ Variance of the estimates nuisance adjusted sandwich with Fay correction for small samples
- $\text{call}$ Call function
geeDREstimation

- $corr$ Correlation structure used
- $clusz$ Number of unit in each cluster
- $SFunList$ List of function associated with the family
- $SX$ design matrix for the main regression
- $Soffset$ Offset specified in the regression
- $Seta$ predicted values
- $Sweights$ Weights vector used in the diagonal term for the IPW
- $Sps.model$ Summary of the regression fitted for the PS if computed internally
- $Som.model.trt$ Summary of the regression fitted for the OM for treated if computed internally
- $Som.model.ctrl$ Summary of the regression fitted for the OM for control if computed internally

Author(s)


References

Details regarding implementation can be found in

- 'Small-Sample Adjustments for Wald-Type Tests Using Sandwich Estimators’ - 2001 - Fay, Michael P and Graubard, Barry I: Biometrics 57(4) - 1198-1206.

Examples

data(data.sim)
# Not run:
##### STANDARD GEE
geeresults<-geeDREstimation(formula=OUTCOME~TRT, 
  id="CLUSTER", data = data.sim, 
  family = "binomial", corstr = "independence")
summary(geeresults)
##### IPW GEE
ipwresults<-geeDREstimation(formula=OUTCOME~TRT, 
  id="CLUSTER", data = data.sim, 
  family = "binomial", corstr = "independence", 
  model.weights=I(MISSING==0)*TRT*AGE)
summary(ipwresults)
### AUGMENTED GEE

```r
augresults <- geeDREstimation(formula = OUTCOME ~ TRT,
   id = "CLUSTER", data = data.sim,
   family = "binomial", corstr = "independence",
   model.augmentation.trt = OUTCOME ~ AGE,
   model.augmentation.ctrl = OUTCOME ~ AGE,
   model.augmentation.crl = OUTCOME ~ AGE,
   stepwise.augmentation = FALSE)

summary(augresults)
```

### DOUBLY ROBUST

```r
drresults <- geeDREstimation(formula = OUTCOME ~ TRT,
   id = "CLUSTER", data = data.sim,
   family = "binomial", corstr = "independence",
   model.weights = I(MISSING == 0) * TRT * AGE,
   model.augmentation.trt = OUTCOME ~ AGE,
   model.augmentation.ctrl = OUTCOME ~ AGE,
   model.augmentation.crl = OUTCOME ~ AGE,
   stepwise.augmentation = FALSE)

summary(drresults)
```

---

### getCI

*Get Mean, Sd and CI for estimates from CRTgeeDR object.*

**Description**

Get the estimates, standard deviations and confidence intervals from an CRTgeeDR object associated with a regressor given in argument.

**Usage**

```r
getCI(object, nameTRT = "TRT", quantile = 1.96)
```

**Arguments**

- `object`: CRTgeeDR
- `nameTRT`: character including the name of the variable of interest (often the treatment)
- `quantile`: value of the normal quantile for the IC. default is 1.96 for 95%CI.

---

### getOMPlot

*Get the observed vs fitted residuals*

**Description**

Get the histogram and some basic statistics for the weights used in the IPW part.

**Usage**

```r
getOMPlot(object, save = FALSE, name = "plotOM", typeplot = 0)
```
getPSPlot

Arguments

- **object**: CRTgeeDR
- **save**: logical if TRUE the plot is saved as a pdf in the current directory
- **name**: name of the plot saved as pdf
- **typeplot**: integer indicating which is the adequation diagnostic plot for the PS. '0', all available in plot.glm are displayed, '1' Residuals vs Fitted, '2' Normal Q-Q, '3' Scale-Location, '4' Cook's distance, '5' Residuals vs Leverage and '6' Cook's dist vs Leverage* h[iii] / (1 - h[iii])

getPSPlot  
Get the histogram of weights for IPW and adequation for the glm weights model

Description

Get the histogram and some basic statistics for the weights used in the IPW part.

Usage

getPSPlot(object, save = FALSE, name = "plotPS", typeplot = NULL)

Arguments

- **object**: CRTgeeDR
- **save**: logical if TRUE the plot is saved as a pdf in the current directory
- **name**: name of the plot saved as pdf
- **typeplot**: integer indicating which is the adequation diagnostic plot for the PS. Default is NULL no output. '0', all available in plot.glm are displayed, '1' Residuals vs Fitted, '2' Normal Q-Q, '3' Scale-Location, '4' Cook's distance, '5' Residuals vs Leverage and '6' Cook's dist vs Leverage* h[iii] / (1 - h[iii])

predict.CRTgeeDR

Predict CRTgeeDR object.

Description

Predict CRTgeeDR object to a dataset

Usage

## S3 method for class 'CRTgeeDR'
predict(object, newdata = NULL, ...)

Arguments

- **object**: CRTgeeDR object
- **newdata**: dataframe, new dataset to which the CRTgeeDR need to be used for prediction
- **...**: ignored

---

print.CRTgeeDR *Prints CRTgeeDR object.*

---

Description

Prints CRTgeeDR object

Usage

```r
## S3 method for class 'CRTgeeDR'
print(x, ...)
```

Arguments

- **x**: CRTgeeDR x
- **...**: ignored

---

print.summary.CRTgeeDR *Print the summarizing CRTgeeDR object.*

---

Description

Print Summary CRTgeeDR object

Usage

```r
## S3 method for class 'summary.CRTgeeDR'
print(x, ...)
```

Arguments

- **x**: summary.CRTgeeDR x
- **...**: ignored
**Summary**

**CRTgeeDR**

**Description**

Summary CRTgeeDR object

**Usage**

```r
## S3 method for class 'CRTgeeDR'
summary(object, ...)
```

**Arguments**

- `object` CRTgeeDR object
- `...` ignored
Index

CRTgeeDR, 2
CRTgeeDR-package (CRTgeeDR), 2

data.sim, 2

fitted (fitted.CRTgeeDR), 3
fitted.CRTgeeDR, 3

geeDREstimation, 2, 3
getCI, 8
getOMPlot, 8
getPSPPlot, 9

predict (predict.CRTgeeDR), 9
predict.CRTgeeDR, 9
print (print.CRTgeeDR), 10
print.CRTgeeDR, 10
print.summary (print.summary.CRTgeeDR), 10
print.summary.CRTgeeDR, 10

summary (summary.CRTgeeDR), 11
summary.CRTgeeDR, 11