Package ‘MultiRobust’

October 12, 2022

Type Package

Title Multiply Robust Methods for Missing Data Problems

Version 1.0.5

Author Shixiao Zhang and Peisong Han

Maintainer Shixiao Zhang <praetere@gmail.com>

Description Multiply robust estimation for population mean (Han and Wang 2013) <doi:10.1093/biomet/ass087>, regression analysis (Han 2014) <doi:10.1080/01621459.2014.880058> (Han 2016) <doi:10.1111/sjos.12177> and quantile regression (Han et al. 2019) <doi:10.1111/rssb.12309>.

License GPL (>= 2)

Encoding UTF-8

RoxygenNote 6.1.1

Imports stats

Suggests MASS, SuppDists, quantreg

NeedsCompilation no

Repository CRAN

Date/Publication 2019-06-03 22:30:03 UTC

R topics documented:

glm.work — *Define a Generalized Linear Model*

Description

Define a generalized linear model. All the arguments in glm are allowed except for data. Supported types of family include gaussian, binomial, poisson, Gamma and inverse.gaussian.

Usage

```
glm.work(formula, family = gaussian, weights = NULL, ...)
```

Arguments

formula	The formula of the model to be fitted.
family	The distribution of the response variable and the link function to be used in the model.
weights	The prior weights to be used in the model.
...	Addition arguments for the function glm.

See Also

glm.

Examples

```
# A logistic regression with response R and covariates X1 and X2
mis1 <- glm.work(formula = R ~ X1 + X2, family = binomial(link = logit))
```

MR.mean — *Multiply Robust Estimation of the Marginal Mean*

Description

MR.mean() is used to estimate the marginal mean of a variable which is subject to missingness. Multiple missingness probability models and outcome regression models can be accommodated.

Usage

```
MR.mean(response, reg.model = NULL, mis.model = NULL, moment = NULL,
  order = 1, data, bootstrap = FALSE, bootstrap.size = 300,
  alpha = 0.05)
```

Arguments

response	The response variable of interest whose marginal mean is to be estimated.
reg.model	A list of outcome regression models defined by glm.work.
mis.model	A list of missingness probability models defined by glm.work. The dependent variable is always specified as R.
moment	A vector of auxiliary variables whose moments are to be calibrated.
order	A numeric value. The order of moments up to which to be calibrated.
data	A data frame with missing data encoded as NA.
bootstrap	Logical. If bootstrap = TRUE, the bootstrap will be applied to calculate the standard error and construct a Wald confidence interval.
bootstrap.size	A numeric value. Number of bootstrap resamples generated if bootstrap = TRUE.
alpha	Significance level used to construct the 100(1 - alpha)% Wald confidence interval.

Value

mu	The estimated value of the marginal mean.
SE	The bootstrap standard error of mu when bootstrap = TRUE.
CI	A Wald-type confidence interval based on mu and SE when bootstrap = TRUE.
weights	The calibration weights if any reg.model, mis.model or moment is specified.

References

Han, P. and Wang, L. (2013). Estimation with missing data: beyond double robustness. *Biometrika*, **100**(2), 417–430.

Han, P. (2014). A further study of the multiply robust estimator in missing data analysis. *Journal of Statistical Planning and Inference*, **148**, 101–110.

Examples

```
# Simulated data set
set.seed(123)
n <- 400
gamma0 <- c(1, 2, 3)
alpha0 <- c(-0.8, -0.5, 0.3)
X <- runif(n, min = -2.5, max = 2.5)
p.mis <- 1 / (1 + exp(alpha0[1] + alpha0[2] * X + alpha0[3] * X ^ 2))
R <- rbinom(n, size = 1, prob = 1 - p.mis)
a.x <- gamma0[1] + gamma0[2] * X + gamma0[3] * exp(X)
Y <- rnorm(n, a.x, sd = sqrt(4 * X ^ 2 + 2))
dat <- data.frame(X, Y)
dat[R == 0, 2] <- NA

# Define the outcome regression models and missingness probability models
reg1 <- glm.work(formula = Y ~ X + exp(X), family = gaussian)
```

```
reg2 <- glm.work(formula = Y ~ X + I(X ^ 2), family = gaussian)
mis1 <- glm.work(formula = R ~ X + I(X ^ 2), family = binomial(link = logit))
mis2 <- glm.work(formula = R ~ X + exp(X), family = binomial(link = cloglog))
MR.mean(response = Y, reg.model = list(reg1, reg2),
        mis.model = list(mis1, mis2), data = dat)
MR.mean(response = Y, moment = c(X), order = 2, data = dat)
```

MR.quantile *Multiply Robust Estimation of the Marginal Quantile*

Description

MR.quantile() is used to estimate the marginal quantile of a variable which is subject to missingness. Multiple missingness probability models and imputation models are allowed.

Usage

```
MR.quantile(response, tau = 0.5, imp.model = NULL, mis.model = NULL,
  moment = NULL, order = 1, L = 30, data, bootstrap = FALSE,
  bootstrap.size = 300, alpha = 0.05)
```

Arguments

response	The response variable of interest whose marginal quantile is to be estimated.
tau	A numeric value in (0,1). The quantile to be estimated.
imp.model	A list of imputation models defined by glm.work.
mis.model	A list of missingness probability models defined by glm.work. The dependent variable is always specified as R.
moment	A vector of auxiliary variables whose moments are to be calibrated.
order	A numeric value. The order of moments up to which to be calibrated.
L	Number of imputations.
data	A data frame with missing data encoded as NA.
bootstrap	Logical. If bootstrap = TRUE, the bootstrap will be applied to calculate the standard error and construct a Wald confidence interval.
bootstrap.size	A numeric value. Number of bootstrap resamples generated if bootstrap = TRUE.
alpha	Significance level used to construct the 100(1 - alpha)% Wald confidence interval.

Value

q	The estimated value of the marginal quantile.
SE	The bootstrap standard error of q when bootstrap = TRUE.
CI	A Wald-type confidence interval based on q and SE when bootstrap = TRUE.
weights	The calibration weights if any imp.model, mis.model or moment is specified.

References

Han, P., Kong, L., Zhao, J. and Zhou, X. (2019). A general framework for quantile estimation with incomplete data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **81**(2), 305–333.

Examples

```
# Simulated data set
set.seed(123)
n <- 400
gamma0 <- c(1, 2, 3)
alpha0 <- c(-0.8, -0.5, 0.3)
X <- runif(n, min = -2.5, max = 2.5)
p.mis <- 1 / (1 + exp(alpha0[1] + alpha0[2] * X + alpha0[3] * X ^ 2))
R <- rbinom(n, size = 1, prob = 1 - p.mis)
a.x <- gamma0[1] + gamma0[2] * X + gamma0[3] * exp(X)
Y <- rnorm(n, a.x, sd = sqrt(4 * X ^ 2 + 2))
dat <- data.frame(X, Y)
dat[R == 0, 2] <- NA

# Define the outcome regression models and missingness probability models
imp1 <- glm.work(formula = Y ~ X + exp(X), family = gaussian)
imp2 <- glm.work(formula = Y ~ X + I(X ^ 2), family = gaussian)
mis1 <- glm.work(formula = R ~ X + I(X ^ 2), family = binomial(link = logit))
mis2 <- glm.work(formula = R ~ X + exp(X), family = binomial(link = cloglog))
MR.quantile(response = Y, tau = 0.25, imp.model = list(imp1, imp2),
            mis.model = list(mis1, mis2), L = 10, data = dat)
MR.quantile(response = Y, tau = 0.25, moment = c(X), order = 2, data = dat)
```

MR.quantreg *Multiply Robust Estimation for Quantile Regression*

Description

MR.quantreg() is used for quantile regression with missing responses and/or missing covariates. Multiple missingness probability models and imputation models are allowed.

Usage

```
MR.quantreg(formula, tau = 0.5, imp.model = NULL, mis.model = NULL,
  moment = NULL, order = 1, L = 30, data, bootstrap = FALSE,
  bootstrap.size = 300, alpha = 0.05, ...)
```

Arguments

formula	The formula of the linear quantile regression model of interest.
tau	A numeric value in (0,1). The quantile to be estimated.

imp.model	A list of possibly multiple lists of the form list(list.1, list.2, ..., list.K), where K is the total number of different imputation models. For the *k*-th imputation model, list.k is a list of possibly multiple models, each of which is defined by glm.work and imputes one single missing variable marginally. See details.
mis.model	A list of missingness probability models defined by glm.work. The dependent variable is always specified as R.
moment	A vector of auxiliary variables whose moments are to be calibrated.
order	A numeric value. The order of moments up to which to be calibrated.
L	Number of imputations.
data	A data frame with missing data encoded as NA.
bootstrap	Logical. If bootstrap = TRUE, the bootstrap will be applied to calculate the standard error and construct a Wald confidence interval.
bootstrap.size	A numeric value. Number of bootstrap resamples generated if bootstrap = TRUE.
alpha	Significance level used to construct the 100(1 - alpha)% Wald confidence interval.
...	Addition arguments for the function rq.

Details

The function MR.quantreg() currently deals with data with one missingness pattern. When multiple variables are subject to missingness, their values are missing simultaneously. The method in Han et al. (2019) specifies an imputation model by modeling the joint distribution of the missing variables conditional on the fully observed variables. In contrast, the function MR.quantreg() specifies an imputation model by separately modeling the marginal distribution of each missing variable conditional on the fully observed variables. These marginal distribution models for different missing variables constitute one joint imputation model. Different imputation models do not need to model the marginal distribution of each missing variable differently.

Value

coefficients	The estimated quantile regression coefficients.
SE	The bootstrap standard error of coefficients when bootstrap = TRUE.
CI	A Wald-type confidence interval based on coefficients and SE when bootstrap = TRUE.

References

Han, P., Kong, L., Zhao, J. and Zhou, X. (2019). A general framework for quantile estimation with incomplete data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **81**(2), 305–333.

See Also

rq.

Examples

```
# Simulated data set
set.seed(123)
n <- 400
gamma0 <- c(1, 2, 3)
alpha0 <- c(-0.8, -0.5, 0.3)
S <- runif(n, min = -2.5, max = 2.5) # auxiliary variables
X1 <- rbinom(n, size = 1, prob = 0.5) # covariate X1
X2 <- rexp(n) # covariate X2
p.obs <- 1 / (1 + exp(alpha0[1] + alpha0[2] * S + alpha0[3] * S ^ 2)) # non-missingness probability
R <- rbinom(n, size = 1, prob = p.obs)
a.x <- gamma0[1] + gamma0[2] * X1 + gamma0[3] * X2
Y <- rnorm(n, a.x)
dat <- data.frame(S, X1, X2, Y)
dat[R == 0, c(2, 4)] <- NA # X1 and Y may be missing

# marginal imputation models for X1
impX1.1 <- glm.work(formula = X1 ~ S, family = binomial(link = logit))
impX1.2 <- glm.work(formula = X1 ~ S + X2, family = binomial(link = cloglog))
# marginal imputation models for Y
impY.1 <- glm.work(formula = Y ~ S, family = gaussian)
impY.2 <- glm.work(formula = Y ~ S + X2, family = gaussian)
# missingness probability models
mis1 <- glm.work(formula = R ~ S + I(S ^ 2), family = binomial(link = logit))
mis2 <- glm.work(formula = R ~ I(S ^ 2), family = binomial(link = cloglog))
# this example considers the following K = 3 imputation models for imputing the missing (X1, Y)
imp1 <- list(impX1.1, impY.1)
imp2 <- list(impX1.1, impY.2)
imp3 <- list(impX1.2, impY.1)

results <- MR.quantreg(formula = Y ~ X1 + X2, tau = 0.75, imp.model = list(imp1, imp2, imp3),
                       mis.model = list(mis1, mis2), L = 10, data = dat)
results$coefficients
MR.quantreg(formula = Y ~ X1 + X2, tau = 0.75,
            moment = c(S, X2), order = 2, data = dat)$coefficients
```

MR.reg *Multiply Robust Estimation for (Mean) Regression*

Description

MR.reg() is used for (mean) regression under generalized linear models with missing responses and/or missing covariates. Multiple missingness probability models and imputation models are allowed.

Usage

```
MR.reg(formula, family = gaussian, imp.model = NULL,
  mis.model = NULL, moment = NULL, order = 1, L = 30, data,
  bootstrap = FALSE, bootstrap.size = 300, alpha = 0.05, ...)
```

Arguments

formula	The formula of the regression model of interest.
family	A description of the error distribution and link function to be used for the GLM of interest.
imp.model	A list of possibly multiple lists of the form list(list.1, list.2, ..., list.K), where K is the total number of different imputation models. For the *k*-th imputation model, list.k is a list of possibly multiple models, each of which is defined by glm.work and imputes one single missing variable marginally. See details.
mis.model	A list of missingness probability models defined by glm.work. The dependent variable is always specified as R.
moment	A vector of auxiliary variables whose moments are to be calibrated.
order	A numeric value. The order of moments up to which to be calibrated.
L	Number of imputations.
data	A data frame with missing data encoded as NA.
bootstrap	Logical. If bootstrap = TRUE, the bootstrap will be applied to calculate the standard error and construct a Wald confidence interval.
bootstrap.size	A numeric value. Number of bootstrap resamples generated if bootstrap = TRUE.
alpha	Significance level used to construct the 100(1 - alpha)% Wald confidence interval.
...	Addition arguments for the function glm.

Details

The function MR.reg() currently deals with data with one missingness pattern. When multiple variables are subject to missingness, their values are missing simultaneously. The methods in Han (2016) and Zhang and Han (2019) specify an imputation model by modeling the joint distribution of the missing variables conditional on the fully observed variables. In contrast, the function MR.reg() specifies an imputation model by separately modeling the marginal distribution of each missing variable conditional on the fully observed variables. These marginal distribution models for different missing variables constitute one joint imputation model. Different imputation models do not need to model the marginal distribution of each missing variable differently.

Value

coefficients	The estimated regression coefficients.
SE	The bootstrap standard error of coefficients when bootstrap = TRUE.
CI	A Wald-type confidence interval based on coefficients and SE when bootstrap = TRUE.

References

Han, P. (2014). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association*, **109**(507), 1159–1173.

Han, P. (2016). Combining inverse probability weighting and multiple imputation to improve robustness of estimation. *Scandinavian Journal of Statistics*, **43**, 246–260.

Zhang, S. and Han, P. (2019). A simple implementation of multiply robust estimation for GLMs with missing data. Unpublished manuscript.

See Also

glm.

Examples

```
# Simulated data set
set.seed(123)
n <- 400
gamma0 <- c(1, 2, 3)
alpha0 <- c(-0.8, -0.5, 0.3)
S <- runif(n, min = -2.5, max = 2.5) # auxiliary variables
X1 <- rbinom(n, size = 1, prob = 0.5) # covariate X1
X2 <- rexp(n) # covariate X2
p.obs <- 1 / (1 + exp(alpha0[1] + alpha0[2] * S + alpha0[3] * S ^ 2)) # non-missingness probability
R <- rbinom(n, size = 1, prob = p.obs)
a.x <- gamma0[1] + gamma0[2] * X1 + gamma0[3] * X2
Y <- rnorm(n, a.x)
dat <- data.frame(S, X1, X2, Y)
dat[R == 0, c(2, 4)] <- NA # X1 and Y may be missing

# marginal imputation models for X1
impX1.1 <- glm.work(formula = X1 ~ S, family = binomial(link = logit))
impX1.2 <- glm.work(formula = X1 ~ S + X2, family = binomial(link = cloglog))
# marginal imputation models for Y
impY.1 <- glm.work(formula = Y ~ S, family = gaussian)
impY.2 <- glm.work(formula = Y ~ S + X2, family = gaussian)
# missingness probability models
mis1 <- glm.work(formula = R ~ S + I(S ^ 2), family = binomial(link = logit))
mis2 <- glm.work(formula = R ~ I(S ^ 2), family = binomial(link = cloglog))
# this example considers the following K = 3 imputation models for imputing the missing (X1, Y)
imp1 <- list(impX1.1, impY.1)
imp2 <- list(impX1.1, impY.2)
imp3 <- list(impX1.2, impY.1)

results <- MR.reg(formula = Y ~ X1 + X2, family = gaussian, imp.model = list(imp1, imp2, imp3),
                  mis.model = list(mis1, mis2), L = 10, data = dat)
results$coefficients
MR.reg(formula = Y ~ X1 + X2, family = gaussian,
       moment = c(S, X2), order = 2, data = dat)$coefficients
```

Index