Package 'ppmHR'

October 14, 2022

Type Package

Title Privacy-Protecting Hazard Ratio Estimation in Distributed Data
Networks

Version 1.0

Date 2020-01-06

Author Di Shu <shudi1991@gmail.com>, Sengwee Toh <darren_toh@harvardpilgrim.org>

Maintainer Di Shu <shudi1991@gmail.com>

Description An implementation of the one-step privacy-protecting method for estimating the over-
all and site-specific hazard ratios using inverse probability weighted Cox models in dis-
tributed data network studies, as proposed by Shu, Yoshida, Fire-
man, and Toh (2019) <doi:10.1177/0962280219869742>. This method only requires shar-
ing of summary-level riskset tables instead of individual-level data. Both the conventional in-
verse probability weights and the stabilized weights are implemented.

License GPL (>= 2)

Imports nleqslv, stats

RoxygenNote 7.0.2

NeedsCompilation no

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2020-01-07 17:20:02 UTC

R topics documented:

---

ppmHR-package	*Privacy-Protecting Hazard Ratio Estimation in Distributed Data Networks*

Description

This package allows implementation of a one-step, privacy-protecting method for estimating the overall and site-specific hazard ratios using inverse probability weighted Cox models in multi-center distributed data network studies, as proposed by Shu, Yoshida, Fireman, and Toh (2019). The method only requires sharing of summary-level riskset tables instead of individual-level data by data-contributing sites. Both the conventional inverse probability weights and the stabilized weights are implemented.

Details

The ppmHR package implements a one-step, privacy-protecting method for estimating overall and site-specific hazard ratios using inverse probability weighted Cox models, under both the conventional inverse probability weights and the stabilized weights, as proposed by Shu, Yoshida, Fireman, and Toh (2019). The function checkBalanceSite allows the data-contributing sites to check their site-specific covariate balance before and after weighting. The function computeInfoForTable1 allows the data-contributing sites to compute summary-level information that will be used to create the "Table 1" of baseline patient characteristics before and after weighting. The function createTable1 allows the analysis center to create the "Table 1" of baseline patient characteristics before and after weighting, using only summary-level information shared by data-contributing sites. The function createRisksetTable allows the data-contributing sites to create summary-level riskset tables using their site-specific individual-level data. The function estimateStratHR allows the analysis center to estimate the overall hazard ratios and robust sandwich variances using inverse probability weighted Cox models stratified on data-contributing sites, using only the summary-level riskset tables provided by the data-contributing sites. The function estimateSiteHRs allows the analysis center to estimate the site-specific hazard ratios and robust sandwich variances using inverse probability weighted Cox models, using only the summary-level riskset tables provided by the data-contributing sites.

Author(s)

Di Shu and Sengwee Toh

Maintainer: Di Shu <shudi1991@gmail.com>

References

Shu D, Yoshida K, Fireman BH, Toh S (2019). Inverse probability weighted Cox model in multi-site studies without sharing individual-level data. *Statistical Methods in Medical Research* <doi:10.1177/0962280219869742>

---

checkBalanceSite *Checking of site-specific covariate balance before and after weighting*

Description

This function checks for site-specific covariate balance before and after weighting using the mean of baseline covariates in the original unweighted and the inverse probability weighted samples. A logistic regression model is used to estimate the propensity scores with an option to do weight truncation. This is a function for data-contributing sites.

Usage

```
checkBalanceSite(data, indA, indX, truncP = 1)
```

Arguments

data   An individual-level dataset in the form of R data frame.

indA   A column name indicating the exposure/treatment variable.

indX   A vector of column names indicating the baseline covariates to be included in the propensity score model.

truncP   A value between 0 and 1 indicating the percentile used for weight truncation. The default is 1, corresponding to no weight truncation.

Value

A data frame of balance checking results. Each covariate has a row of six numbers: the mean of this covariate for exposed individuals in the original sample, for unexposed individuals in the original sample, for exposed individuals in the weighted sample with conventional weights, for unexposed individuals in the weighted sample with conventional weights, for exposed individuals in the weighted sample with stabilized weights, and for unexposed individuals in the weighted sample with stabilized weights.

Examples

```
#load an example dataset site1.RData in the package
#site1 contains individual-level data of data-contributing site 1
data(site1)
#data-contributing site 1 checks covariate balance before and after weighting,
#with logistic propensity score model A~X1+X2+X3+X4+X5
#no weight truncation
checkBalanceSite(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),truncP=1)
#with truncation: set weights larger than the 90% quantile of original weights to the 90% quantile
checkBalanceSite(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),truncP=0.9)
```

---

computeInfoForTable1	*Computation of summary-level information needed for the creation of a "Table 1" of baseline patient characteristics*

Description

Computation of summary-level information needed for the creation of a "Table 1" that shows the baseline patient characteristics in the original unweighted and the inverse probability weighted samples. A logistic regression model is used to estimate propensity scores with an option to do weight truncation. This is a function for data-contributing sites.

Usage

```
computeInfoForTable1(data, indA, indX, truncP = 1)
```

Arguments

data	An individual-level dataset in the form of R data frame.
indA	A column name indicating the exposure/treatment variable.
indX	A vector of column names indicating the baseline covariates to be included in the propensity score model.
truncP	A value between 0 and 1 indicating the percentile used for weight truncation. The default is 1, corresponding to no weight truncation.

Value

A data frame of summary-level information needed for the creation of a "Table 1" of baseline patient characteristics. Each covariate has a row of 19 numbers: covariate type ("yes" if it is binary, "no" if it is continuous/count), number of exposed individuals (same across rows), covariate mean of exposed individuals, covariate mean square of exposed individuals, total conventional weights of exposed individuals, weighted covariate mean of exposed individuals with conventional weights, weighted covariate mean square of exposed individuals with conventional weights, total stabilized weights of exposed individuals, weighted covariate mean of exposed individuals with stabilized weights, weighted covariate mean square of exposed individuals with stabilized weights, number of unexposed individuals (same across rows), covariate mean of unexposed individuals, covariate mean square of unexposed individuals, total conventional weights of unexposed individuals, weighted covariate mean of unexposed individuals with conventional weights, weighted covariate mean square of unexposed individuals with conventional weights, total stabilized weights of unexposed individuals, weighted covariate mean of unexposed individuals with stabilized weights, and weighted covariate mean square of unexposed individuals with stabilized weights.

Examples

```
#load an example dataset site1.RData in the package
#site1 contains individual-level data of data-contributing site 1
data(site1)
#site 1 computes summary-level information needed for creating the "Table 1"
```

```
#with logistic propensity score model A~X1+X2+X3+X4+X5
#no weight truncation
computeInfoForTable1(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),truncP=1)
#with truncation: set weights larger than the 90% quantile of original weights to the 90% quantile
computeInfoForTable1(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),truncP=0.9)
```

---

createRisksetTable   *Creation of summary-level riskset tables from individual-level data*

Description

This function creates summary-level riskset tables from individual-level data at each data-contributing site, under both the conventional inverse probability weights and the stabilized weights. A logistic regression model is used to estimate the propensity scores with an option to do weight truncation. This is a function for data-contributing sites.

Usage

```
createRisksetTable(
  data,
  indA,
  indX,
  indStatus,
  indTime,
  truncP = 1,
  shareEventTime = "no"
)
```

Arguments

data	An individual-level dataset in the form of R data frame.
indA	A column name indicating the exposure/treatment variable.
indX	A vector of column names indicating the baseline covariates to be included in the propensity score model.
indStatus	A column name indicating the non-censoring status (1 if observed and 0 if censored).
indTime	A column name indicating the outcome variable, i.e., min(true event time, censoring time).
truncP	A value between 0 and 1 indicating the percentile used for weight truncation. The default is 1, corresponding to no weight truncation.
shareEventTime	If the data-contributing site would like to share a column of event times, then set shareEventTime="yes". If not, then set shareEventTime="no". By default, shareEventTime="no".

Value

A list of two summary-level riskset tables to be shared with the analysis center, under both the conventional inverse probability weights (i.e., $ipwTable) and the stabilized weights (i.e., $stabTable). If shareEventTime="no", then a riskset table has 8 columns: total weights of exposed cases or events, total weights of cases or events, total weights of exposed individuals, total weights of unexposed individuals, total squared weights of exposed cases or events, total squared weights of unexposed cases or events, total squared weights of exposed individuals, total squared weights of unexposed individuals. If shareEventTime="yes", then a riskset table further has a column of event times.

Examples

```
#load an example dataset site1.RData in the package
#site1 contains individual-level data of data-contributing site 1
data(site1)
#data-contributing site 1 creates its two summary-level riskset tables
#with logistic propensity score model A~X1+X2+X3+X4+X5
#no weight truncation
#agree to share a column of event times
rsTb1=createRisksetTable(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),
                         indStatus="status",indTime="time",truncP=1,shareEventTime="yes")
#print the first six rows of riskset table using conventional weights
head(rsTb1$ipwTable)
#print the first six rows of riskset table using stabilized weights
head(rsTb1$stabTable)
```

---

createTable1   *Creation of an overall "Table 1" of baseline patient characteristics*
                             before and after weighting

Description

Creation of an overall "Table 1" of baseline patient characteristics in the original unweighted and the inverse probability weighted samples, using only summary-level information shared by data-contributing sites. This is a function for the analysis center.

Usage

```
createTable1(XsummaryList, digits = 2)
```

Arguments

XsummaryList	A list of summary-level information tables shared by data-contributing sites. Each site provides one table, which can be obtained by using the function computeInfoForTable1 in this package.
digits	An integer indicating the number of decimal places for the generated "Table 1". The default is 2.

Value

A data frame of the "Table 1". Each row represents a covariate. The seven columns represent the covariate type ("yes" if it is binary, "no" if it is continuous or count), the exposed group in the original sample, the unexposed group in the original sample, the exposed group in the weighted sample with conventional weights, the unexposed group in the weighted sample with conventional weights, the exposed group in the weighted sample with stabilized weights, and the unexposed group in the weighted sample with stabilized weights. For cells of binary covariates, values are count(percentage). For cells of continuous or count covariates, values are mean(standard deviation).

Examples

```
#load example datasets in the package
#site1-3 contain individual-level data of data-contributing sites 1-3
data(site1)
data(site2)
data(site3)
#sites 1-3 compute summary-level information needed for creating the "Table 1"
#with logistic propensity score model A~X1+X2+X3+X4+X5
#no weight truncation
Xsummary1=computeInfoForTable1(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),truncP=1)
Xsummary2=computeInfoForTable1(data=site2,indA="A",indX=c("X1","X2","X3","X4","X5"),truncP=1)
Xsummary3=computeInfoForTable1(data=site3,indA="A",indX=c("X1","X2","X3","X4","X5"),truncP=1)
#analysis center creates the "Table 1"
#using summary-level information Xsummary1-3 shared by data-contributing sites
#display the table with 3 decimal places
createTable1(list(Xsummary1,Xsummary2,Xsummary3),digits=3)
#analysis center can also generate site-specific "Table 1"
#for example, for site 1
createTable1(list(Xsummary1),digits=3)
```

---

estimateSiteHRs	*Privacy-protecting estimation of site-specific hazard ratios using inverse probability weighted Cox models*

Description

This function allows privacy-protecting estimation of the site-specific hazard ratios using inverse probability weighted Cox models, under both the conventional inverse probability weights and the stabilized weights. The robust sandwich variance estimation method is used for estimating the variance of the log hazard ratio estimates. The Breslow method is used to handle tied events. This is a function for the analysis center.

Usage

```
estimateSiteHRs(tableList, initialHR = 1, endpoint = Inf, confidence = 0.95)
```

Arguments

tableList	A list of summary-level riskset tables shared by data-contributing sites. Each data-contributing site provides a list of two riskset tables; the first using the conventional inverse probability weights and the second using the stabilized weights. Data-contributing sites can obtain their summary-level riskset tables using the function `createRisksetTable` in this package.
initialHR	An initial value for hazard ratio when solving the site-specific weighted partial likelihood score equation. The default is 1.
endpoint	A value of the end of follow-up used to conduct sensitivity analysis. Observed events in the original data that occur after this value will be censored. The default is Inf, which means that we use the original data without conducting sensitivity analysis. For riskset tables that do not provide event times, endpoint should be left as the default Inf.
confidence	A confidence level between 0 and 1. The default is 0.95 corresponding to a 95 per cent confidence interval.

Value

A matrix of inference results from inverse probability weighted Cox models for each data-contributing site. Each site has two rows of results, where the first and the second rows report the log hazard ratio estimate and associated robust sandwich standard error, hazard ratio estimate and associated normality-based confidence interval, under the conventional inverse probability weights and the stabilized weights, respectively.

Examples

```
#load example datasets in the package
#site1-3 contain individual-level data of data-contributing sites 1-3
data(site1)
data(site2)
data(site3)
#data-contributing sites 1-3 create summary-level riskset tables
#with logistic propensity score model A~X1+X2+X3+X4+X5
#no weight truncation
#do not share event times
rsTb1=createRisksetTable(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),
                        indStatus="status",indTime="time",truncP=1,shareEventTime="no")
rsTb2=createRisksetTable(data=site2,indA="A",indX=c("X1","X2","X3","X4","X5"),
                        indStatus="status",indTime="time",truncP=1,shareEventTime="no")
rsTb3=createRisksetTable(data=site3,indA="A",indX=c("X1","X2","X3","X4","X5"),
                        indStatus="status",indTime="time",truncP=1,shareEventTime="no")
#analysis center estimates site-specific hazard ratios in IPW Cox models
#for all data-contributing sites
#using summary-level riskset tables rsTb1-3 shared by data-contributing sites
estimateSiteHRs(list(rsTb1,rsTb2,rsTb3),initialHR=1,endpoint=Inf,confidence=0.95)
```

---

estimateStratHR	*Privacy-protecting estimation of overall hazard ratios using inverse probability weighted Cox models stratified on data-contributing sites*

Description

This function estimates the overall hazard ratios using inverse probability weighted Cox models stratified on data-contributing sites, under both the conventional inverse probability weights and the stabilized weights. The robust sandwich variance estimation method is used for estimating the variance of the log hazard ratio estimates. The Breslow method is used to handle tied events. This is a function for the analysis center.

Usage

```
estimateStratHR(tableList, initialHR = 1, endpoint = Inf, confidence = 0.95)
```

Arguments

tableList	A list of summary-level riskset tables shared by data-contributing sites. Each data-contributing site provides a list of two riskset tables; the first using the conventional inverse probability weights and the second using the stabilized weights. Data-contributing sites can obtain their summary-level riskset tables using the function [createRisksetTable](#) in this package.
initialHR	An initial value for hazard ratio when solving the stratified weighted partial likelihood score equation. The default is 1.
endpoint	A value of the end of follow-up used to conduct sensitivity analysis. Observed events in the original data that occur after this value will be censored. The default is Inf, which means that we use the original data without conducting sensitivity analysis. For riskset tables that do not provide event times, endpoint should be left as the default Inf.
confidence	A confidence level between 0 and 1. The default is 0.95 corresponding to a 95 per cent confidence interval.

Value

A matrix of inference results from inverse probability weighted Cox models stratified on data-contributing sites. The first and the second rows report the log hazard ratio estimate and associated robust sandwich standard error, hazard ratio estimate and associated normality-based confidence interval, under the conventional inverse probability weights and the stabilized weights, respectively.

Examples

```
#load example datasets in the package
#site1-3 contain individual-level data of data-contributing sites 1-3
data(site1)
data(site2)
data(site3)
```

```
#data-contributing sites 1-3 create summary-level riskset tables
#with logistic propensity score model A~X1+X2+X3+X4+X5
#no weight truncation
#agree to share event times
rsTb1=createRisksetTable(data=site1,indA="A",indX=c("X1","X2","X3","X4","X5"),
                         indStatus="status",indTime="time",truncP=1,shareEventTime="yes")
rsTb2=createRisksetTable(data=site2,indA="A",indX=c("X1","X2","X3","X4","X5"),
                         indStatus="status",indTime="time",truncP=1,shareEventTime="yes")
rsTb3=createRisksetTable(data=site3,indA="A",indX=c("X1","X2","X3","X4","X5"),
                         indStatus="status",indTime="time",truncP=1,shareEventTime="yes")
#analysis center estimates hazard ratio in a stratified IPW Cox model
#using summary-level riskset tables rsTb1-3 shared by data-contributing sites
estimateStratHR(list(rsTb1,rsTb2,rsTb3),initialHR=1,endpoint=Inf,confidence=0.95)
#sensitivity analysis at endpoint 20
estimateStratHR(list(rsTb1,rsTb2,rsTb3),initialHR=1,endpoint=20,confidence=0.95)
```

---

site1   *Example individual-level data of data-contributing site 1*

Description

This dataset gives an example of individual-level data of data-contributing site 1.

Usage

```
data(site1)
```

Format

A data frame consisting of 800 individuals (i.e., rows) with variables (i.e., columns) sequence number (id), outcome (time), non-censoring indicator (status), exposure/treatment indicator (A), five baseline covariates (X1-X5), and data-contributing site indicator (indSite).

---

site2   *Example individual-level data of data-contributing site 2*

Description

This dataset gives an example of individual-level data of data-contributing site 2.

Usage

```
data(site2)
```

Format

A data frame consisting of 1000 individuals (i.e.,rows) with variables (i.e., columns) sequence number (id), outcome (time), non-censoring indicator (status), exposure/treatment indicator (A), five baseline covariates (X1-X5), and data-contributing site indicator (indSite).

---

site3	*Example individual-level data of data-contributing site 3*

Description

This dataset gives an example of individual-level data of data-contributing site 3.

Usage

```
data(site3)
```

Format

A data frame consisting of 1200 individuals (i.e.,rows) with variables (i.e., columns) sequence number (id), outcome (time), non-censoring indicator (status), exposure/treatment indicator (A), five baseline covariates (X1-X5), and data-contributing site indicator (indSite).

Index