Package ‘truncAIPW’

July 22, 2025

Type Package

Title Doubly Robust Estimation under Covariate-Induced Dependent Left Truncation

Version 1.0.1

Author Yuyao Wang [cre, aut],
Andrew Ying [aut],
Ronghui Xu [aut]

Maintainer Yuyao Wang <yuw079@ucsd.edu>

Description
Doubly robust estimation for the mean of an arbitrarily transformed survival time under covariate-induced dependent left truncation and noninformative right censoring. The functions truncAIPW(), truncAIPW_cen1(), and truncAIPW_cen2() compute the doubly robust estimators under the scenario without censoring and the two censoring scenarios, respectively. The package also contains three simulated data sets 'simu', 'simu_c1', and 'simu_c2', which are used to illustrate the usage of the functions in this package.
Reference: Wang, Y., Ying, A., Xu, R. (2022) ``Doubly robust estimation under covariate-induced dependent left truncation'' <doi:10.48550/arXiv.2208.06836>.

Depends R (>= 3.5.0)

License GPL-3

Encoding UTF-8

URL https://arxiv.org/pdf/2208.06836.pdf

LazyData true

Imports stats, survival, survPen

RoxygenNote 7.2.3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-08-31 13:50:02 UTC

Contents

F_est 2
G_est 3
simu 5
simu_c1 6
simu_c2 7
truncAIPW 7
truncAIPW_cen1 9
truncAIPW_cen2 11

Index **13**

`F_est` *Estimate the Conditional CDF of the Event Time given Covariates*

Description

Estimate the conditional cumulative distribution function (CDF) of the event time given covariates evaluated at given time points. The options implemented in this function are: Cox proportional hazards regression using function `coxph()` from R package 'survival', and the hazard model with penalized splines using function `survPen()` from R package 'survPen'.

Usage

```
F_est(
  dat.fit,
  dat.est = dat.fit,
  time.eval,
  model,
  time.name,
  Q.name,
  event.name,
  cov.names,
  trim = 0,
  formula.survPen = NA
)
```

Arguments

`dat.fit`	data frame that is used to fit the model for the full data conditional distribution of the event time given the covariates.
`dat.est`	data frame that contains the subjects for which the estimated conditional CDF is computed.
`time.eval`	vector of time points at which the conditional CDF is evaluated.

`model` method used to estimate the conditional CDF. The options available are "Cox" and "spline", corresponding to Cox proportional hazards regression using function `coxph()` from R package ‘survival’, and the hazard model with penalized splines using function `survPen()` from R package ‘survPen’, respectively.

`time.name` name of the event time variable.

`Q.name` name of the left truncation time variable.

`event.name` name of the event indicator.

`cov.names` vector of the names of covariates.

`trim` constant for bounding the estimated conditional CDF from 1.

`formula.survPen` the formula when applying the hazard model with penalized splines implemented in `survPen::survPen`.

Value

`F_est()` returns a matrix of the estimated conditional CDF for subjects in ‘`data.est`’ evaluated at the time points in the vector ‘`time.eval`’. Each row corresponds to a subject and each column corresponds to a time point. The column names of the matrix are the times in ‘`time.eval`’.

See Also

G_est

Examples

```
data("simu")
u = c(1, 1.5, 2, 2.5, 3, 3.5, 4)
Fuz.mx = F_est(simu, simu[1:10,], u, "Cox", "time", "Q", "delta", c("Z1","Z2"))
```

G_est *Estimate the Conditional CDF for the Left Truncation Time given Covariates*

Description

Estimate the conditional cumulative distribution function (CDF) of the left truncation time given covariates evaluated at given time points. The options implemented in this function are: Cox proportional hazards regression using function `coxph()` from R package ‘survival’, and the hazard model with penalized splines using function `survPen()` from R package ‘survPen’.

Usage

```
G_est(
  dat.fit,
  dat.est = dat.fit,
  time.eval,
  model,
  time.name,
  Q.name,
  event.name,
  cov.names,
  trim = 0,
  weights = rep(1, nrow(dat.fit)),
  formula.survPen = NA
)
```

Arguments

`dat.fit`	data frame that is used to fit the model for the full data conditional distribution of the event time given the covariates.
`dat.est`	data frame that contains the subjects for which the estimated conditional CDF is computed.
`time.eval`	vector of time points at which the conditional CDF is evaluated.
`model`	method used to estimate the conditional CDF. The options available are "Cox" and "spline", corresponding to Cox proportional hazards regression using function `coxph()` from R package 'survival', and the hazard model with penalized splines using function `survPen()` from R package 'survPen', respectively.
`time.name`	name of the event time variable.
`Q.name`	name of the left truncation time variable.
`event.name`	name of the event indicator.
`cov.names`	vector of the names of covariates.
`trim`	constant for bounding the estimated conditional CDF from 0.
`weights`	vector of case weights.
`formula.survPen`	the formula when applying the hazard model with penalized splines implemented in `survPen::survPen`.

Value

`G_est()` returns a matrix of the estimated conditional CDF for subjects in '`data.est`' evaluated at the time points in the vector '`time.eval`'. Each row corresponds to a subject and each column corresponds to a time point. The column names of the matrix are the times in '`time.eval`'.

See Also

`F_est`

Examples

```
data("simu")
v = c(0.5, 1, 1.5, 2, 2.5, 3)
Gvz.mx = G_est(simu, simu[1:10,], v, "Cox", "time", "Q", "delta", c("Z1","Z2"))
```

---

simu *A Simulated Data Set under Left Truncation but No Right Censoring*

Description

A simulated data set under left truncation but no right censoring. The left truncation time and the event time are dependent via the two covariates 'Z1' and 'Z2'. Under the data generating mechanism, the conditional distribution of the event time given covariates follows a Cox proportional hazards model in the full data, and the conditional distribution of the left truncation time given covariates follows a Cox proportional hazards model on the reversed time scale in the full data. The truncation rate is 29.5%; and the truth $P^*(T > 3) = 0.5755753$.

Usage

```
data(simu)
```

Format

A data frame with 500 subjects (rows) and 5 variables (columns).

time the event time

Q the left truncation time

delta the event indicator

Z1 a continous covariate

Z2 a binary covariate

Source

Simulated

Examples

```
data(simu)
```

simu_c1 *A Simulated Data Set under Left Truncation and Right Censoring where Censoring can be before Truncation*

Description

A simulated data set under left truncation and right censoring where censoring can be before left truncation. The left truncation time and the event time are dependent via the two covariates 'Z1' and 'Z2'. Under the data generating mechanism, the conditional distribution of the censored event time X given covariates follows a Cox proportional hazards model in the full data, and the conditional distribution of the left truncation time given covariates follows a Cox proportional hazards model on the reversed time scale in the full data. The truncation rate is 29.5%; the censoring rate is 16.5%; $P^*(C < Q) = 0.6057$; and the truth $P^*(T > 3) = 0.623955$.

Usage

```
data(simu_c1)
```

Format

A data frame with 500 rows and 5 variables.

X the censored event time X = min(T,C)

Q the left truncation time

delta the event indicator

Z1 a continous covariate

Z2 a binary covariate

Source

Simulated

Examples

```
data(simu_c1)
```

simu_c2 — *A Simulated Data Set under Left Truncation and Right Censoring where Censoring is always after Truncation*

Description

A simulated data set under left truncation and right censoring where censoring is always after left truncation. The left truncation time and the event time are dependent via two covariates 'Z1' and 'Z2'. Under the data generating mechanism, the conditional distributions of the event time given covariates follows a Cox model in the full data, and the conditional distribution of the left truncation time given covariates follows a Cox model on the reversed time scale in the full data. The truncation rate is 29.5%; the censoring rate is 27.1%; and the truth $P^*(T > 3) = 0.576547$.

Usage

```
data(simu_c2)
```

Format

A data frame with 500 rows and 5 variables.

X the censored event time X = min(T,C)

Q the left truncation time

delta the event indicator

Z1 a continous covariate

Z2 a binary covariate

Source

Simulated

Examples

```
data(simu_c2)
```

truncAIPW — *Doubly Robust Estimation under Covariate-induced Dependent Left Truncation and No Censoring*

Description

Doubly robust estimation for the mean of an arbitrarily transformed survival time under covariate-induced dependent left truncation and no right censoring.

Usage

```
truncAIPW(dat, nu, Fuz.mx, Gvz.mx, T.name, Q.name, trim = 1e-07)
```

Arguments

- `dat` data frame that contains the data for constructing the estimating equation.
- `nu` transformation that defines the parameter of interest.
- `Fuz.mx` matrix for the estimated conditional CDF of the event time given covariates. Each row corresponds to a subject, and each column corresponds to a time point. The column names of the matrix are the time points. See `F_est` for an example of computing this conditional CDF matrix.
- `Gvz.mx` matrix for the estimated conditional CDF of the truncation time given covariates. Each row corresponds to a subject, and each column corresponds to a time point. The column names of the matrix are the time points. See `G_est` for an example of computing this conditional CDF matrix.
- `T.name` name of the event time variable.
- `Q.name` name of the left truncation time variable.
- `trim` constant that is used to bound from below for the denominators involved in the computation.

Value

`truncAIPW()` returns a list of estimators (‘dr’, ‘IPW.Q’, ‘Reg.T1’, ‘Reg.T2’), and the model-based standard errors for the ‘dr’ and ‘IPW.Q’ estimators.

- `dr` doubly robust estimator ‘dr’.
- `IPW.Q` inverse probability of truncation weighted estimator ‘IPW.Q’.
- `Reg.T1` regression based estimator ‘Reg.T1’.
- `Reg.T2` regression based estimator ‘Reg.T2’.
- `SE_dr` standard error of the ‘dr’ estimator based on the efficient influence function.
- `SE_IPW.Q` standard error of the ‘IPW.Q’ estimator computed from the robust sandwich variance estimator assuming the truncation weights are known.

References

Wang, Y., Ying, A., Xu, R. (2022) "Doubly robust estimation under covariate-induced dependent left truncation" <arXiv:2208.06836>.

See Also

See `truncAIPW_cen1`, `truncAIPW_cen2` for the estimations also under noninformative right censoring. See `F_est`, `G_est` for examples of computing the input matrices for the conditional CDF’s.

Examples

```
data("simu")
nu <- function(t){ return(as.numeric(t>3)) }
u = c(min(simu$time)-1e-10, sort(simu$time), max(simu$time)+1e-10)
v = c(min(simu$Q)-1e-10, sort(simu$Q), max(simu$Q)+1e-10)
Fuz.mx = F_est(simu, simu, u, "Cox", "time", "Q", "delta", c("Z1","Z2"))
Gvz.mx = G_est(simu, simu, v, "Cox", "time", "Q", "delta", c("Z1","Z2"))

est = truncAIPW(simu, nu, Fuz.mx, Gvz.mx, "time", "Q", trim = 1e-7)
est
```

---

truncAIPW_cen1 — *Doubly Robust Estimation under Covariate-induced Dependent Left Truncation and Noninformative Right Censoring where Censoring can be before Left Truncation*

Description

Doubly robust estimation of the mean of an arbitrarily transformed survival time under covariate-induced dependent left truncation and noninformative right censoring where censoring can be before left truncation. Inverse probability of censoring weighting is used to handle the right censoring.

Usage

```
truncAIPW_cen1(
  dat,
  nu,
  Fuz.mx,
  Gvz.mx,
  Sc,
  X.name,
  Q.name,
  status.name,
  trim = 1e-07
)
```

Arguments

`dat`	data frame that contains the data for constructing the estimating equation.
`nu`	transformation that defines the parameter of interest.
`Fuz.mx`	matrix for the estimated conditional CDF of the event time given covariates. Each row corresponds to a subject, and each column corresponds to a time point. The column names of the matrix are the time points. See `F_est` for an example of computing this input matrix for the conditional CDF.
`Gvz.mx`	matrix for the estimated conditional CDF of the truncation time given covariates. Each row corresponds to a subject, and each column corresponds to a time point. The column names of the matrix are the time points. See `G_est` for an example of computing this input matrix for the conditional CDF.

Sc	a function for the censoring survival curve $S_c(\cdot)$.
X.name	name of the censored event time variable X = min(T, C).
Q.name	name of the left truncation time variable.
status.name	name of the event time indicator.
trim	constant that is used to bound from below for the denominators involved in the computation.

Value

truncAIPW_cen1() returns a list of estimators ('dr', 'IPW.Q', 'Reg.T1', 'Reg.T2').

dr	doubly robust estimator 'dr'.
IPW.Q	inverse probability of truncation weighted estimator 'IPW.Q'.
Reg.T1	regression based estimator 'Reg.T1'.
Reg.T2	regression based estimator 'Reg.T2'.

References

Wang, Y., Ying, A., Xu, R. (2022) "Doubly robust estimation under covariate-induced dependent left truncation" <arXiv:2208.06836>.

See Also

See also truncAIPW for estimation under no censoring, and truncAIPW_cen2 for estimation under another type of noninformative right censoring. See also F_est, G_est as examples for computing the input matrices of the conditional CDF's.

Examples

```
library(survival)
data("simu_c1")
simu_c1$delta.1 = 1

nu <- function(t){ return(as.numeric(t>3)) }
u = c(min(simu_c1$X)-1e-10, sort(simu_c1$X), max(simu_c1$X)+1e-10)
v = c(min(simu_c1$Q)-1e-10, sort(simu_c1$Q), max(simu_c1$Q)+1e-10)

Fuz.mx = F_est(simu_c1, simu_c1, u, "Cox", "X", "Q", "delta.1", c("Z1","Z2"))
Gvz.mx = G_est(simu_c1, simu_c1, v, "Cox", "X", "Q", "delta.1", c("Z1","Z2"))

# KM curve for Sc
kmfit.C = survfit(Surv(Q, X, 1-delta)~1, data = simu_c1, type = "kaplan-meier")
Sc = stepfun(kmfit.C$time,  c(1, kmfit.C$surv))

est = truncAIPW_cen1(simu_c1, nu, Fuz.mx, Gvz.mx, Sc, "X", "Q", "delta", trim = 1e-7)
est
```

truncAIPW_cen2 *Doubly Robust Estimation under Covariate-induced Dependent Left Truncation and Noninformative Right Censoring where Censoring is always after Left Truncation*

Description

Doubly robust estimation of the mean of an arbitrarily transformed survival time under covariate-induced dependent left truncation and noninformative right censoring where censoring is always after left truncation. Inverse probability of censoring weighting is used to handle the right censoring.

Usage

```
truncAIPW_cen2(
  dat,
  nu,
  Fuz.mx,
  Gvz.mx,
  wd,
  X.name,
  Q.name,
  status.name,
  trim = 1e-07
)
```

Arguments

dat	data frame that contains the data for constructing the estimating equation.
nu	transformation that defines the parameter of interest.
Fuz.mx	matrix for the estimated conditional CDF of the event time given covariates. Each row corresponds to a subject, and each column corresponds to a time point. The column names of the matrix are the time points. See F_est for an example of computing this input matrix for the conditional CDF.
Gvz.mx	matrix for the estimated conditional CDF of the truncation time given covariates. Each row corresponds to a subject, and each column corresponds to a time point. The column names of the matrix are the time points. See G_est for an example of computing this input matrix for the conditional CDF.
wd	vector for the inverse probability of residual censoring weights $\Delta/\hat{S}_D(X-Q)$.
X.name	name of the censored event time variable X = min(T, C).
Q.name	name of the left truncation time variable.
status.name	name of the event time indicator.
trim	constant that is used to bound from below for the denominators involved in the computation.

Value

`truncAIPW_cen2()` returns a list of estimators ('dr', 'IPW.Q', 'Reg.T1', 'Reg.T2').

dr	doubly robust estimator 'dr'.
IPW.Q	inverse probability of truncation weighted estimator 'IPW.Q'.
Reg.T1	regression based estimator 'Reg.T1'.
Reg.T2	regression based estimator 'Reg.T2'.

References

Wang, Y., Ying, A., Xu, R. (2022) "Doubly robust estimation under covariate-induced dependent left truncation" <arXiv:2208.06836>.

See Also

See also truncAIPW for estimation under no censoring, and truncAIPW_cen1 for estimation under another type of noninformative right censoring. See also F_est, G_est as examples for computing the input matrices of the conditional CDF's.

Examples

```
library(survival)
data("simu_c2")
nu <- function(t){ return(as.numeric(t>3)) }

u = c(min(simu_c2$X)-1e-10, sort(simu_c2$X), max(simu_c2$X)+1e-10)
v = c(min(simu_c2$Q)-1e-10, sort(simu_c2$Q), max(simu_c2$Q)+1e-10)

kmfit.D = survfit(Surv(X-Q, 1-delta)~1, data = simu_c2, type = "kaplan-meier")
Sd = stepfun(kmfit.D$time,  c(1, kmfit.D$surv))
wd = rep(0, nrow(simu_c2))
wd[which(simu_c2$delta == 1)] = 1/Sd(simu_c2$X - simu_c2$Q)[which(simu_c2$delta == 1)]

simu_c2$wd = wd
simu_c2.1 = simu_c2[simu_c2$delta==1,]
wd_1 = simu_c2.1$wd

Fuz.mx = F_est(simu_c2, simu_c2, u, "Cox", "X", "Q", "delta", c("Z1","Z2"))
Gvz.mx = G_est(simu_c2.1, simu_c2, v, "Cox", "X", "Q", "delta", c("Z1","Z2"), weights = wd_1)

est = truncAIPW_cen2(simu_c2, nu, Fuz.mx, Gvz.mx, wd, "X", "Q", "delta", trim = 1e-7)
est
```

Index

$*$ **datasets**
    simu, 5
    simu_c1, 6
    simu_c2, 7

F_est, 2, *4*, *8–12*

G_est, *3*, 3, *8–12*

simu, 5
simu_c1, 6
simu_c2, 7

truncAIPW, 7, *10*, *12*
truncAIPW_cen1, *8*, 9, *12*
truncAIPW_cen2, *8*, *10*, 11