Package 'DTDA'

February 19, 2020

Type Package

Title Doubly Truncated Data Analysis

Version 2.1-2

Author Carla Moreira, Jacobo de Uña-Álvarez and Rosa Crujeiras

Maintainer Carla Moreira <carlamgmm@gmail.com>

Description Implementation of different algorithms for analyzing randomly truncated data, one-sided and two-sided (i.e. doubly) truncated data. Two real data sets are included.

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2020-02-19 13:00:02 UTC

R topics documented:

DTDA-package *Doubly truncated data analysis*

Description

This package implements different algorithms for analyzing randomly truncated data, one-sided and two-sided (i.e. doubly) truncated data. It incorporates the iterative methods introduced by Efron and Petrosian (1999) and Shen (2008). Estimation of the lifetime distribution function and truncation times distributions is possible, together with the corresponding pointwise confidence limits based on bootstrap methods. Plots of cumulative distributions and survival functions are provided. Two real data sets are included: right-truncated AIDS data and doubly truncated data on quasar luminosities.

Details

Package:	DTDA
Type:	Package
Version:	2.1-2
Date:	2010-11-14
License:	GPL-2
LazyLoad:	yes

Missing data are allowed. Registries with missing data are simply removed. This package incorporates the functions `efron.petrosian`, `lynden` and `shen`, which call the iterative methods introduced by Efron and Petrosian (1999) and Shen (2008).

For a complete list of functions, use `library(help="DTDA")`.

Acknowledgements

Work supported by the research grants MTM 2008-02139 and MTM 2008-0310 of the Spanish Ministry of Science and Innovation, and Xunta de Galicia grants PGIDIT07PXIB300191PR and PGIDIT06PXIB207009RR.

Author(s)

Carla Moreira, Jacobo de Uña-Álvarez and Rosa Crujeiras

Maintainer: Carla Moreira <carlamgmm@gmail.com>

References

Efron, B. and Petrosian, V.(1999) Nonparametric methods for doubly truncated data. *Journal of the American Statistical Association* **94**, 824-834.

Lynden-Bell, D. (1971) A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monograph National Royal Astronomical Society* **155**, 95-118.

Shen, P-S. (2008) Nonparametric analysis of doubly truncated data. *Annals of the Institute of Statistical Mathematics* DOI 10.1007/s10463-008-0192-2.

AIDS *Time to AIDS*

Description

The data include information on the infection and induction times for 258 adults who were infected with HIV virus and developed AIDS by June 30, 1996. The data consist on the time in years, measured from April 1, 1978, when adults were infected by the virus from a contaminated blood transfusion, and the waiting time to development of AIDS, measured from the date of infection.

Usage

```
data(AIDS)
```

Format

A data frame with 258 observations on the following 3 variables.

`INFTime` a numeric vector, the infection time.

`INDTime` a numeric vector, the induction time.

`V` a numeric vector, the time from HIV infection to the end of the study.

Source

J.P. Klein and M.L.Moeschberger.

References

Lagakos SW and Barraj LM and de Gruttola V (1988) Nonparametric Analysis of Truncated Survival Data, with Applications to AIDS. *Biometrika* **75**, 515–523.

Examples

```
data(AIDS)
str(AIDS)
```

efron.petrosian *Doubly truncated data analysis with the first Efron-Petrosian algorithm*

Description

This function computes the NPMLE of a lifetime distribution function observed under one-sided (right or left) and two-sided (double) truncation. It provides bootstrap pointwise confidence limits too.

Usage

```
efron.petrosian(X, U = NA, V = NA, wt = NA, error = NA,
 nmaxit = NA, boot = TRUE, B = NA, alpha = NA,
 display.F = FALSE, display.S = FALSE)
```

Arguments

X	Numeric vector with the times of ultimate interest.
U	Numeric vector with the left truncation times. If there are no truncation times from the left, put U=NA.
V	Numeric vector with the right truncation times. If there are no truncation times from the left, put V=NA.
wt	Numeric vector of non-negative initial solution, with the same length as X. Default value is set to 1/n, being n the length of X.
error	Numeric value. Maximum pointwise error when estimating the density associated to X (f) in two consecutive steps. If this is missing, it is $1e-06$.
nmaxit	Numeric value. Maximum number of iterations. If this is missing, it is set to nmaxit =100 .
boot	Logical. If TRUE (default), the simple bootstrap method is applied to lifetime distribution estimation. Pointwise confidence bands are provided.
B	Numeric value. Number of bootstrap resamples . The default NA is equivalent to B =500 .
alpha	Numeric value. (1-alpha) is the nominal coverage for the pointwise confidence intervals.
display.F	Logical. Default is FALSE. If TRUE, the estimated cumulative distribution function associated to X, (F) is plotted.
display.S	Logical. Default is FALSE. If TRUE, the estimated survival function associated to X, (S) is plotted.

Details

The NPMLE of the lifetime is computed by the first algorithm proposed in Efron and Petrosian (1999). This is an alternative algorithm which converges to the NMPLE after a number of iterations. If the second (respectively third) argument is missing, computation of the Lynden-Bell estimator for right-truncated (respectively left-truncated) data is obtained. Note that individuals with NAs in the three first arguments will be automatically excluded.

Value

A list containing the following values:

time	The timepoint on the curve.
n.event	The number of events that ocurred at time t.
events	The total number of events.
density	The estimated density values.

`cumulative.df`	The estimated cumulative distribution values.
`truncation.probs`	The probabilities of truncation values, in each region.
`S0`	`error` reached in the algorithm.
`Survival`	The estimated survival values.
`n.iterations`	The number of iterations used by this algorithm.
`B`	Number of bootstrap resamples computed.
`alpha`	The nominal level used to construct the confidence intervals.
`upper.df`	The estimated upper limits of the confidence intervals for F.
`lower.df`	The estimated lower limits of the confidence intervals for F.
`upper.Sob`	The estimated upper limits of the confidence intervals for S.
`lower.Sob`	The estimated lower limits of the confidence intervals for S.

Author(s)

Carla Moreira, Jacobo de Uña-Álvarez and Rosa Crujeiras

References

Efron, B. and Petrosian, V.(1999) Nonparametric methods for doubly truncated data. *Journal of the American Statistical Association* **94**, 824-834.

Lynden-Bell, D. (1971) A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monograph National Royal Astronomical Society* **155**, 95-118.

See Also

`lynden`

Examples

```
##  Generating data which are doubly truncated

n<-25
X<-runif(n,0,1)
U<-runif(n,0,0.5)
V<-runif(n,0.5,1)

for (i in 1:n){
while (X[i]<U[i]|X[i]>V[i]){
U[i]<-runif(1,0,0.5)
X[i]<-runif(1,0,1)
V[i]<-runif(1,0.5,1)
   }
 }
efron.petrosian(X=X,U=U,V=V,boot=FALSE,display.F=TRUE,display.S=TRUE)
```

lynden *Doubly truncated data analysis with the second Efron-Petrosian algorithm*

Description

This function computes the NPMLE of a lifetime distribution function observed under one-sided (right or left) and two-sided (double) truncation. It provides bootstrap pointwise confidence limits too.

Usage

```
lynden(X, U = NA, V = NA, error = NA, nmaxit = NA,
 boot = TRUE, B = NA, alpha = NA,
display.F = FALSE, display.S = FALSE)
```

Arguments

X	Numeric vector with the times of ultimate interest.
U	Numeric vector with the left truncation times. If there are no truncation times from the left, put U=NA.
V	Numeric vector with the right truncation times. If there are no truncation times from the left, put V=NA.
error	Numeric value. Maximum pointwise error when estimating the density associated to X (f) in two consecutive steps. If this is missing, it is $1e-06$.
nmaxit	Numeric value. Maximum number of iterations. If this is missing, it is set to nmaxit =100 .
boot	Logical. If TRUE (default), the simple bootstrap method is applied to lifetime distribution estimation. Pointwise confidence bands are provided.
B	Numeric value. Number of bootstrap resamples . The default NA is equivalent to B =500 .
alpha	Numeric value. (1-alpha) is the nominal coverage for the pointwise confidence intervals.
display.F	Logical. Default is FALSE. If TRUE, the estimated cumulative distribution function associated to X, (F) is plotted.
display.S	Logical. Default is FALSE. If TRUE, the estimated survival function associated to X, (S) is plotted.

Details

The NPMLE of the lifetime is computed by the second algorithm proposed in Efron and Petrosian (1999). This is an alternative algorithm which converges to the NMPLE after a number of iterations. If the second (respectively third) argument is missing, computation of the Lynden-Bell estimator for right-truncated (respectively left-truncated) data is obtained. Note that individuals with NAs in the three first arguments will be automatically excluded.

Value

A list containing the following values:

`time`	The timepoint on the curve.
`n.event`	The number of events that ocurred at time `t`.
`events`	The total number of events.
`NJ`	The number of individuals in risk considering the left truncation times.
`density`	The estimated density values.
`cumulative.df`	The estimated cumulative distribution values.
`truncation.probs`	The probabilities of truncation values, in each region.
`hazard`	The estimated hazard values.
`S0`	`error` reached in the algorithm.
`Survival`	The estimated survival values.
`n.iterations`	The number of iterations used by this algorithm.
`B`	Number of bootstrap resamples computed.
`alpha`	The nominal level used to construct the confidence intervals.
`upper.df`	The estimated upper limits of the confidence intervals for F.
`lower.df`	The estimated lower limits of the confidence intervals for F.
`upper.Sob`	The estimated upper limits of the confidence intervals for S.
`lower.Sob`	The estimated lower limits of the confidence intervals for S.

Author(s)

Carla Moreira, Jacobo de Uña-Álvarez and Rosa Crujeiras

References

Efron, B. and Petrosian, V.(1999) Nonparametric methods for doubly truncated data. *Journal of the American Statistical Association* **94**, 824-834.

Lynden-Bell, D. (1971) A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monograph National Royal Astronomical Society* **155**, 95-118.

See Also

`efron.petrosian`

Examples

```
#  Generating data which are doubly truncated
n<-25
X<-runif(n,0,1)
U<-runif(n,0,0.25)
V<-runif(n,0.75,1)
```

```
for (i in 1:n){
while (X[i]<U[i]|X[i]>V[i]){
U[i]<-runif(1,0,0.25)
X[i]<-runif(1,0,1)
V[i]<-runif(1,0.75,1)
   }
 }
res<-lynden(X=X, U=U, V=V, boot=FALSE, display.F=TRUE, display.S=TRUE)

# Generating data which are right truncated

n<-25
X<-runif(n,0,1)
V<-runif(n,0.75,1)

for (i in 1:n){
while (X[i]>V[i]){
X[i]<-runif(1,0,1)
V[i]<-runif(1,0.75,1)
     }
 }
res<-lynden(X=X,U=NA, V=V, boot=FALSE)
```

`Quasars` *Quasars*

Description

The original data set studied by Efron and Petrosian (1999), comprised independently collected quadruplets $(z_i; m_i; a_i; b_i), i = 1, \ldots n$, where z_i is the redshift of the ith quasar and m_i is the apparent magnitude. Due to experimental constraints, the distribution of each luminosity in the log-scale $(y_i = t(z_i, m_i))$ is truncated to a known interval $[a_i; b_i]$, where t represents a transformation which depends on the cosmological model assumed (see Efron and Petrosian (1999) for details). Quasars with apparent magnitude above b_i were too dim to yield dependent redshifts, and hence they were excluded from the study. The lower limit a_i was used to avoid confusion with non quasar stellar objects.

Usage

```
data(Quasars)
```

Format

A data frame with 210 observations on the following 3 variables.

`y (adj lum)` a numeric vector, the log lominosity values.

`u (lower)` a numeric vector, lower truncation limits.

`v (upper)` a numeric vector, upper truncation limits.

Source

Vahé Petrosian and Bradley Efron.

References

Boyle, B.J. and Fong, R. and Shanks, T. and Peterson, B.A. (1990) *Monograph National Royal Astronomical Society* **243**, 1.

Efron, B. and Petrosian, V.(1999) Nonparametric methods for doubly truncated data. *Journal of the American Statistical Association* **94**, 824-834.

Examples

```
data(Quasars)
 str(Quasars)
```

---

shen *Doubly truncated data analysis with the Shen algorithm*

Description

This function computes the NPMLE of a lifetime distribution function observed under one-sided (right or left) and two-sided (double) truncation. The NPMLE of the joint distribution of the truncation times along with its marginal distributions are also computed. It provides bootstrap pointwise confidence limits too.

Usage

```
shen(X, U = NA, V = NA, wt = NA, error = NA,
 nmaxit = NA, boot = TRUE, boot.type = "simple",
 B = NA, alpha = NA, display.FS = FALSE,
display.UV = FALSE, plot.joint = FALSE, plot.type = NULL)
```

Arguments

X	Numeric vector with the times of ultimate interest.
U	Numeric vector with the left truncation times. If there are no truncation times from the left, put U=NA.
V	Numeric vector with the right truncation times. If there are no truncation times from the left, put V=NA.
wt	Numeric vector of non-negative initial solution, with the same length as X. Default value is set to 1/n, being n the length of X.
error	Numeric value. Maximum pointwise error when estimating the density associated to X (f) in two consecutive steps. If this is missing, it is $1e-06$.
nmaxit	Numeric value. Maximum number of iterations. If this is missing, it is set to nmaxit =100 .

boot	Logical. If TRUE (default), the simple bootstrap method is applied to lifetime and truncation times distributions estimation. Pointwise confidence bands are provided.
boot.type	A character string giving the bootstrap type to be used. This must be one of "simple" or "obvious", with default "simple".
B	Numeric value. Number of bootstrap resamples . The default NA is equivalent to B =500 .
alpha	Numeric value. (1-alpha) is the nominal coverage for the pointwise confidence intervals.
display.FS	Logical. Default is FALSE. If TRUE, the estimated cumulative distribution function and the estimated survival function associated to X, (F) and (S) respectively, are plotted.
display.UV	Logical. Default is FALSE. If TRUE, the marginal distributions of U (fU) and V (fV), are plotted.
plot.joint	Logical. Default is FALSE. If TRUE, the joint distribution of the truncation times is plotted.
plot.type	A character string giving the plot type to be used to represent the joint distribution of the truncation times. This must be one of "image" or "persp", with default NULL.

Details

The NPMLE of the lifetime is computed by a single algorithm proposed in Shen (2008). This is an alternative algorithm which converges to the NMPLE after a number of iterations. Initial solutions are given by the ordinary empirical distribution functions. If the second (respectively third) argument is missing, computation of the Lynden-Bell estimator for right-truncated (respectively left-truncated) data is obtained. Note that individuals with NAs in the three first arguments will be automatically excluded.

Value

A list containing the following values:

time	The timepoint on the curve.
n.event	The number of events that ocurred at time t.
events	The total number of events.
density	The estimated density values associated to X.
cumulative.df	The estimated cumulative distribution values of X.
truncation.probs	The probabilities of truncation values, in each region.
S0	error reached in the algorithm.
Survival	The estimated survival values.
density.joint	The estimated joint densities values associated to (U,V).
marginal.U	The estimated cumulative univariate marginal values of the U.

`marginal.V`	The estimated cumulative univariate marginal values of the `V`.
`cumulative.joint`	The estimated joint cumulative distribution values.
`n.iterations`	The number of iterations used by this algorithm.
`biasf`	The estimated probabilities of observing the lifetimes.
`Boot`	The type of bootstrap method applied.
`B`	Number of bootstrap resamples computed.
`alpha`	The nominal level used to construct the confidence intervals.
`upper.df`	The estimated upper limits of the confidence intervals for F.
`lower.df`	The estimated lower limits of the confidence intervals for F.
`upper.Sob`	The estimated upper limits of the confidence intervals for S.
`lower.Sob`	The estimated lower limits of the confidence intervals for S.
`upper.fU`	The estimated upper limits of the confidence intervals for `fU`.
`lower.fU`	The estimated lower limits of the confidence intervals for `fU`.
`upper.fV`	The estimated upper limits of the confidence intervals for `fV`.
`lower.fV`	The estimated lower limits of the confidence intervals for `fV`.

Author(s)

Carla Moreira, Jacobo de Uña-Álvarez and Rosa Crujeiras

References

Lynden-Bell, D. (1971) A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monograph National Royal Astronomical Society* **155**, 95-118. Astronomical Society, 155, 95-118.

Shen, P-S. (2008) Nonparametric analysis of doubly truncated data. *Annals of the Institute of Statistical Mathematics* DOI 10.1007/s10463-008-0192-2.

See Also

`lynden`

Examples

```
##  Generating data which are doubly truncated

n<-25
X<-runif(n,0,1)
U<-runif(n,0,0.67)
V<-runif(n,0.33,1)
for (i in 1:n){
while (X[i]<U[i]|X[i]>V[i]){
U[i]<-runif(1,0,0.67)
X[i]<-runif(1,0,1)
```

```
V[i]<-runif(1,0.33,1)
}
}

res<-shen(X,U,V,boot=FALSE, plot.joint=TRUE, plot.type="persp")
```

Index