Package 'cenROC'

March 27, 2023

Type Package

Title Estimating Time-Dependent ROC Curve and AUC for Censored Data

Version 2.0.0

Description

Contains functions to estimate a smoothed and a non-smoothed (empirical) time-dependent
receiver operating characteristic curve and the corresponding area under the receiver
operating characteristic curve and the optimal cutoff point for the right and interval
censored survival data. See Beyene and El Ghouch (2020)<doi:10.1002/sim.8671> and Beyene
and El Ghouch (2022) <doi:10.1002/bimj.202000382>.

License GPL (>= 2)

Depends R (>= 3.5.0)

Imports Rcpp (>= 1.0.0), icenReg, condSURV, survival, stats, graphics,
methods

LinkingTo Rcpp, RcppEigen

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation yes

Author Kassu Mehari Beyene [aut, cre],
Anouar El Ghouch [aut, ths]

Maintainer Kassu Mehari Beyene <kassu.mehari@wu.edu.et>

Repository CRAN

Date/Publication 2023-03-27 08:10:05 UTC

R topics documented:

---

cenROC   *Estimation of the time-dependent ROC curve for right censored sur-*
                             vival data

Description

This function computes the time-dependent ROC curve for right censored survival data using the
cumulative sensitivity and dynamic specificity definitions. The ROC curves can be either empir-
ical (non-smoothed) or smoothed with/wtihout boundary correction. It also calculates the time-
dependent area under the ROC curve (AUC).

Usage

```
cenROC(Y, M, censor, t, U = NULL, h = NULL, bw = "NR", method = "tra",
    ktype = "normal", ktype1 = "normal", B = 0, alpha = 0.05, plot = "TRUE")
```

Arguments

Y	The numeric vector of event-times or observed times.
M	The numeric vector of marker values for which the time-dependent ROC curves is computed.
censor	The censoring indicator, 1 if event, 0 otherwise.
t	A scaler time point at which the time-dependent ROC curve is computed.
U	The vector of grid points where the ROC curve is estimated. The default is a sequence of 151 numbers between 0 and 1.
h	A scaler for the bandwidth of Beran's weight calculaions. The defualt is the value obtained by using the method of Sheather and Jones (1991).
bw	A character string specifying the bandwidth estimation method for the ROC itself. The possible options are "NR" for the normal reference, the plug-in "PI" and the cross-validation "CV". The default is the "NR" normal reference method. The user can also introduce a numerical value.
method	The method of ROC curve estimation. The possible options are "emp" emperical metod; "untra" smooth without boundary correction and "tra" is smooth ROC curve estimation with boundary correction. The default is the "tra" smooth ROC curve estimate with boundary correction.
ktype	A character string giving the type kernel distribution to be used for smoothing the ROC curve: "normal", "epanechnikov", "biweight", or "triweight". By default, the "normal" kernel is used.

ktype1	A character string specifying the desired kernel needed for Beran weight calculation. The possible options are `"normal"`, `"epanechnikov"`, `"tricube"`, `"boxcar"`, `"triangular"`, or `"quartic"`. The defaults is `"normal"` kernel density.
B	The number of bootstrap samples to be used for variance estimation. The default is `0`, no variance estimation.
alpha	The significance level. The default is `0.05`.
plot	The logical parameter to see the ROC curve plot. The default is `TRUE`.

Details

The empirical (non-smoothed) ROC estimate and the smoothed ROC estimate with/without boundary correction can be obtained using this function. The smoothed ROC curve estimators require selecting two bandwidth parametrs: one for Beran's weight calculation and one for smoothing the ROC curve. For the latter, three data-driven methods: the normal reference `"NR"`, the plug-in `"PI"` and the cross-validation `"CV"` were implemented. To select the bandwidth parameter needed for Beran's weight calculation, by default, the plug-in method of Sheather and Jones (1991) is used but it is also possible introduce a numeric value. See Beyene and El Ghouch (2020) for details.

Value

Returns the following items:

ROC The vector of estimated ROC values. These will be numeric numbers between zero

and one.

U The vector of grid points used.

AUC A data frame of dimension 1×4. The columns are: AUC, standard error of AUC, the lower

and upper limits of bootstrap CI.

bw The computed value of bandwidth. For the empirical method this is always `NA`.

Dt The vector of estimated event status.

M The vector of Marker values.

Author(s)

Kassu Mehari Beyene and Anouar El Ghouch

References

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent receiver operating characteristic curve for right censored survival data. *Statistics in Medicine*. 39: 3373– 3396.

Sheather, S. J. and Jones, M. C. (1991). A Reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*. Series B (Methodological) 53(3): 683–690.

Examples

```
library(cenROC)

data(mayo)

est = cenROC(Y=mayo$time, M=mayo$mayoscore5, censor=mayo$censor, t=365*6)
est$AUC
```

CV   *The cross-validation bandwidth selection for weighted data*

Description

This function computes the data-driven bandwidth for smoothing the ROC (or distribution) function using the CV method of Beyene and El Ghouch (2020). This is an extension of the classical (unweighted) cross-validation bandwith selection method to the case of weighted data.

Usage

```
CV(X, wt, ktype = "normal")
```

Arguments

X	The numeric data vector.
wt	The non-negative weight vector.
ktype	A character string giving the type kernel to be used: "normal", "epanechnikov", "biweight", or "triweight". By default, the "normal" kernel is used.

Details

Bowman et al (1998) proposed the cross-validation bandwidth selection method for unweighted kernal smoothed distribution function. This method is implemented in the R package kerdiest. We adapted this for the case of weighted data by incorporating the weight variable into the cross-validation function of Bowman's method. See Beyene and El Ghouch (2020) for details.

Value

Returns the computed value for the bandwith parameter.

Author(s)

Kassu Mehari Beyene and Anouar El Ghouch

References

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent receiver operating characteristic curve for right censored survival data. *Statistics in Medicine*. 39: 3373– 3396.

Bowman A., Hall P. and Trvan T.(1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* 85:799-808.

Quintela-del-Rio, A. and Estevez-Perez, G. (2015). kerdiest: Nonparametric kernel estimation of the distribution function, bandwidth selection and estimation of related functions. R package version 1.2.

Examples

```
library(cenROC)

X <- rnorm(100) # random data vector
wt <- runif(100) # weight vector

## Cross-validation bandwidth selection
CV(X = X, wt = wt)$bw
```

---

hds	*NASA Hypobaric Decompression Sickness Marker Data*

Description

This data contains the marker values with the left and right limits of the observed time for the subjects in NASA Hypobaric Decompression Sickness Data.

Usage

```
data(hds)
```

Format

This is a data frame with 238 observations and 3 variables: L (left limit of the observed time), R (right limit of the observed time) and M (marker). The marker is a score derived by combining the covariates Age, Sex, TR360, and Noadyn.

References

Beyene, K. M. and El Ghouch A. (2022). Time-dependent ROC curve estimation for interval-censored data. *Biometrical Journal*, 64, 1056– 1074.

IntROC   *Time-dependent ROC curve estimation for interval-censored survival data*

Description

This function computes the time-dependent ROC curve for interval censored survival data using the cumulative sensitivity and dynamic specificity definitions. The ROC curves can be either empirical (non-smoothed) or smoothed with/without boundary correction. It also calculates the time-dependent AUC.

Usage

```
IntROC(L, R, M, t, U = NULL, method = "emp", method2 = "pa", dist = "weibull",
       bw = NULL, ktype = "normal", len = 151, B = 0, alpha = 0.05, plot = "TRUE")
```

Arguments

L	The numericvector of left limit of observed time. For left censored observations $L == 0$.
R	The numericvector of right limit of observed time. For right censored observation $R ==$ inf.
M	The numeric vector of marker values.
t	A scaler time point used to calculate the ROC curve.
U	The numeric vector of cutoff values.
method	The method of ROC curve estimation. The possible options are "emp" empirical metod; "untra" smooth without boundary correction and "tra" is smooth ROC curve estimation with boundary correction. The default is the "emp" empirical method.
method2	A character indication type of modeling. This include nonparametric "np", parmetric "pa" and semiparametric "sp". The default is the "np" parametric with weibull distribution.
dist	A character incating the type of distribution for parametric model. This includes are "exponential", "weibull", "gamma", "lnorm", "loglogistic" and "generalgamma".
bw	A character string specifying the bandwidth estimation method. The possible options are "NR" for the normal reference, the plug-in "PI" and the cross-validation "CV". The default is the "NR" normal reference method. It is also possible to use a numeric value.
ktype	A character string giving the type kernel distribution to be used for smoothing the ROC curve: "normal", "epanechnikov", "biweight", or "triweight". By default, the "normal" kernel is used.
len	The length of the grid points for ROC estimation. Default is 151.

B	The number of bootstrap samples to be used for variance estimation. The default is `0`, no variance estimation.
alpha	The significance level. The default is `0.05`.
plot	The logigal parameter to see the ROC curve plot. Default is `TRUE`.

Details

This function implments time-dependent ROC curve and the corresponding AUC using the model-band and nonparametric for the estimation of conditional survival function. The empirical (non-smoothed) ROC estimate and the smoothed ROC estimate with/without boundary correction can be obtained using this function. The smoothed ROC curve estimators require selecting a bandwidth parametr for smoothing the ROC curve. To this end, three data-driven methods: the normal reference `"NR"`, the plug-in `"PI"` and the cross-validation `"CV"` were implemented. See Beyene and El Ghouch (2020) for details.

Value

Returns the following items:

ROC The vector of estimated ROC values. These will be numeric numbers between zero

and one.

U The vector of grid points used.

AUC A data frame of dimension 1×4. The columns are: AUC, standard error of AUC, the lower

and upper limits of bootstrap CI.

bw The computed value of bandwidth. For the empirical method this is always `NA`.

Dt The vector of estimated event status.

M The vector of Marker values.

References

Beyene, K. M. and El Ghouch A. (2022). Time-dependent ROC curve estimation for interval-censored data. *Biometrical Journal*, 64, 1056–1074.

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent receiver operating characteristic curve for right censored survival data. *Statistics in Medicine*. 39: 3373–3396.

Examples

```
library(cenROC)

data(hds)

est = IntROC(L=hds$L, R=hds$R, M=hds$M, t=2)
est$AUC
```

---

mayo   *Mayo Marker Data*

Description

Two marker values with event time and censoring status for the subjects in Mayo PBC data.

Usage

```
data(mayo)
```

Format

A data frame with 312 observations and 4 variables: time (event time/censoring time), censor (censoring indicator), mayoscore4, mayoscore5. The two scores are derived from 4 and 5 covariates respectively.

References

Heagerty, P. J., and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics*, 61(1), 92-105.

NR   *The normal reference bandwidth selection for weighted data*

Description

This function computes the data-driven bandwidth for smoothing the ROC (or distribution) function using the NR method of Beyene and El Ghouch (2020). This is an extension of the classical (unweighted) normal reference bandwith selection method to the case of weighted data.

Usage

```
NR(X, wt, ktype = "normal")
```

Arguments

X	The numeric data vector.
wt	The non-negative weight vector.
ktype	A character string giving the type kernel to be used: "normal", "epanechnikov", "biweight", or "triweight". By default, the "normal" kernel is used.

Details

See Beyene and El Ghouch (2020) for details.

Value

Returns the computed value for the bandwith parameter.

Author(s)

Kassu Mehari Beyene and Anouar El Ghouch

References

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent receiver operating characteristic curve for right censored survival data. *Statistics in Medicine*. 39: 3373– 3396.

Examples

```
library(cenROC)

X <- rnorm(100) # random data vector
wt <- runif(100) # weight vector

## Normal reference bandwidth selection
NR(X = X, wt = wt)$bw
```

PI   *The plug-in bandwidth selection for weighted data*

Description

This function computes the data-driven bandwidth for smoothing the ROC (or distribution) function using the PI method of Beyene and El Ghouch (2020). This is an extension of the classical (unweighted) direct plug-in bandwith selection method to the case of weighted data.

Usage

```
PI(X, wt, ktype = "normal")
```

Arguments

X	The numeric vector of random variable.
wt	The non-negative weight vector.
ktype	A character string giving the type kernel to be used: "normal", "epanechnikov", "biweight", or "triweight". By default, the "normal" kernel is used.

Details

See Beyene and El Ghouch (2020) for details.

Value

Returns the computed value for the bandwith parameter.

Author(s)

Kassu Mehari Beyene and Anouar El Ghouch

References

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent receiver operating character-istic curve for right censored survival data. *Statistics in Medicine*. 39: 3373– 3396.

Examples

```
library(cenROC)

X <- rnorm(100) # random data vector
wt <- runif(100) # weight vector

## Plug-in bandwidth selection
PI(X = X, wt = wt)$bw
```

---

youden   *Computes optimal cutoff point using the Youden index criteria*

Description

This function computes the optimal cutoff point using the Youden index criteria of both right and interval censored time-to-event data. The Youden index estimator can be either empirical (non-smoothed) or smoothed with/without boundary correction.

Usage

```
youden(est, plot = "FALSE")
```

Arguments

est   The object returned either by cenROC or IntROC.

plot   The logical parameter to see the ROC curve plot along with the Youden inex.
               The default is TRUE.

Details

In medical decision-making, obtaining the optimal cutoff value is crucial to identify subject at high risk of experiencing the event of interest. Therefore, it is necessary to select a marker value that classifies subjects into healthy and diseased groups. To this end, in the literature, several methods for selecting optimal cutoff point have been proposed. In this package, we only included the Youden index criteria.

Value

Returns the following items:

Youden.index The maximum Youden index value.

cutopt The optimal cutoff value.

sens The sensitivity corresponding to the optimal cutoff value.

spec The specificity corresponding to the optimal cutoff value.

References

Beyene, K. M. and El Ghouch A. (2022). Time-dependent ROC curve estimation for interval-censored data. *Biometrical Journal*, 64, 1056– 1074.

Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer* 3, 32–35.

Examples

```
library(cenROC)

# Right censored data
data(mayo)

resu <- cenROC(Y=mayo$time, M=mayo$mayoscore5, censor=mayo$censor, t=365*6, plot="FALSE")
youden(resu,  plot="TRUE")

# Interval censored data
data(hds)

resu1 = IntROC(L=hds$L, R=hds$R, M=hds$M, t=2)
youden(resu1,  plot="TRUE")
```

Index