Package ‘cenROC’

March 4, 2020

Type Package

Title Estimation of Time-Dependent ROC Curve and AUC for Censored Survival Data

Version 1.0.0

Author Kassu Mehari Beyene and Anouar El Ghouch

Maintainer Kassu Mehari Beyene <kasu.beyene@uclouvain.be>

Description Contains functions to estimate smoothed and non-smoothed (empirical) time-dependent ROC curve and the corresponding area under the ROC curve (AUC) for the right censored survival data.

Depends R(>= 3.6.0), condSURV

License GPL(>=2)

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

R topics documented:

`cenROC`	*Estimation of time-dependent ROC curve for right censored survival data*

Description

a

This function computes the time-dependent ROC curve for the right censored survival data using the cumulative sensitivity and dynamic specificity definitions. The ROC curves can be either empirical (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", len = 151, method = "tra",
    ktype = "normal", ktype1 = "normal", 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 ~~we want~~ to compute the time-dependent ROC curve.
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 using the method of Sheather and Jones (1991).
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.~~
~~len~~	~~The length of the grid points. The default is 151.~~
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.
plot	The logical parameter to see the ROC curve plot. The default is TRUE.

equidistant

for the ROC itself

the value obtained by

this is confusing for the user

It can also be a user-specified numerical

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 to ~~use numeric value~~. ~~The time-dependent AUC can be computed either using the direct method (i.e. $\widehat{AUC}_t = 1 - n^{-1}\sum_{i=1}^{n} \hat{W}_i \hat{Z}_i$) or using the numerical integration (i.e. $\widehat{AUC}_t = \int_0^1 \widehat{ROC}_t(u)du$). The details about these methods can be found in the paper of~~ Beyene and El Ghouch (2020).

introduce a numerical 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 0 and 1.

the

at U for the given t.

I suggest removing this.

`U` The vector of grid points used.

`AUC` The estimated area under the ROC curve at a given `t` ~~using direct method~~.

~~`AUC1` The estimated area under the ROC curve at a given `t` using numerical integration method.~~

`bw` The computed value of bandwidth. For the empirical method this is always 1.

What about h!!?

Author(s)

Kassu Mehari Beyene, Catholic University of Louvain. `<kasu.beyene@uclouvain.be>`

Anouar El Ghouch, Catholic University of Louvain. `<anouar.elghouch@uclouvain.be>`

References

(should be completed)

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent ROC curves for right-censored survival data.

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)
data <- mayo[ ,c( "time","censor","mayoscore5" )]
t <- 365*6
```

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

```
resu <- cenROC(Y=data$time, M=data$mayoscore5, censor=data$censor, t=t, U=NULL,
        len=151, bw="PI", method="tra", plot=TRUE)

resu$AUC
```

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

Description

using the CV method of Beyene ...

This

This function computes the data-driven bandwidth for smoothing the ROC (or distribution) function. ~~It~~ is an extension of the classical (unweighted) cross-validation bandwith selection method ~~for the case of weighted data. If all the weights are equal, then the value is the same as the one obtained using the classical cross-validation method proposed by~~ Bowman et al (1998).

of

Usage

```
CV(X, wt, ktype = "normal", n_pts = 100, seq_bws = NULL)
```

I realy think that this should also be removed as the purpose of this function is to get a bandwidth from the data.

You should delet this and use the R integrate function and not your integ function. I don't see any reason to do not use integrate function !!.

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.
~~`n_pts`~~	~~The number of points to be used in the integration. The default is `100`~~.
`seq_bws`	The sequence of bandwidths in which to compute the cross-validation function.

numerical

256

default ?

Details

!!!

!!!

This code has been adapted from Quintela-del-Rio and Estevez-Perez (2015). Bowman et al (1998) proposed the cross-validation bandwidth selection method for unweighted kernal smoothed distribution function and it 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.

Try to improve this. It is not well written.

Value

Returns the computed value for the bandwith parameter.

Author(s)

Kassu Mehari Beyene, Catholic University of Louvain. `<kasu.beyene@uclouvain.be>`

Anouar El Ghouch, Catholic University of Louvain. `<anouar.elghouch@uclouvain.be>`

References

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent ROC curves for right-censored survival data.

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
#bandCV <- CV(X = X, wt = wt)

#bandCV$bw
```

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

using the NR method of Beyene ....

This function computes the data-driven bandwidth for smoothing the ROC (or distribution) function. This ~~It~~ is an extension of the classical (unweighted) normal reference bandwith selection method ~~for~~ to the case of weighted data. If all the weights are equal, then the value is the same as the one obtained using the classical normal reference method.

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

This mehod is derived from the asymptotic mean square error of the time-dependent smoothed ROC function. It require estimating the unknown quantity $\kappa(ROC) = \int_{-\infty}^{\infty} \left\{ROC''(u)\right\}^2 du$, where $ROC''(.)$ is the second derivative of the ROC function. To estimate this it is assumed that the random variable follows a normal distribution and the weighted nature is accounted in the estimation of the unknown parameters of the assumed normal distribution. ~~For smoothing the ROC function the following kernel can be considered: `"normal"`, `"epanechnikov"`, `"biweight"`, or `"triweight"`.~~

Not clear. Try to rewrite this

I don't think that you need to show any mathematical formula here. Interested readers should check the paper.

Value

Returns the computed value for the bandwith parameter.

Author(s)

Kassu Mehari Beyene, Catholic University of Louvain. <kasu.beyene@uclouvain.be>

Anouar El Ghouch, Catholic University of Louvain. <anouar.elghouch@uclouvain.be>

References

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent ROC curves for right-censored survival data.

Examples

```
library(cenROC)

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

## Normal reference bandwidth selection
bandNR <- NR(X = X, wt = wt)

bandNR$bw
```

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

same as above

Description

This function computes the data-driven bandwidth for smoothing the ROC (or distribution) function. It is an extension of the classical (unweighted) direct plug-in bandwith selection method for the case of weighted data. If all the weights are equal, then the value is the same as the one obtained using the classical plug-in method.

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.

same as above

Details

This mehod is derived from the asymptotic mean square error of the time-dependent smoothed ROC function. It is the modified version of the plug-in bandwidth selection method for the kernel smoothed distribution function estimation. The quantity $\kappa(ROC) = \int_{-\infty}^{\infty} \left\{ROC^{''}(u)\right\}^2 du$, where $ROC^{''}(.)$ is the second derivative of the ROC function, is estimated nonparametricly using the relation $\kappa(ROC) = -E\left\{WROC^{(3)}(Z)\right\}$, where $ROC^{(3)}(.)$ is the third derivative of the ROC function . For the pilot bandwidth needed for estimating the quantity $E\left\{WROC^{(3)}(Z)\right\}$, the normal reference method is used.

Value

Returns the computed value for the bandwith parameter.

Author(s)

Kassu Mehari Beyene, Catholic University of Louvain. `<kasu.beyene@uclouvain.be>`

Anouar El Ghouch, Catholic University of Louvain. `<anouar.elghouch@uclouvain.be>`

References

Beyene, K. M. and El Ghouch A. (2020). Smoothed time-dependent ROC curves for right-censored survival data.

Examples

```
library(cenROC)

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

## Plug-in bandwidth selection
bandPI <- PI(X = X, wt = wt)

bandPI$bw
```

Index