

# Package ‘sMS’

August 29, 2019

**Type** Package

**Title** Semi-parametric estimation of the Receiver  
Operating-Characteristic (ROC) curve for cohort studies

**Version** 0.1.0

**Author** Susana Díaz-Coto <U0266718@uniovi.es>

**Maintainer** Susana Díaz-Coto <U0266718@uniovi.es>

**Description**

Provides estimations of ROC curve and the Area under the Curve (AUC) based on the semi-parametric Mixed Subjects (sMS) Approach.

**License** GPL

**Encoding** UTF-8

**Imports** survival, rms, pROC

**LazyData** true

**RoxygenNote** 6.0.1

**NeedsCompilation** no

## R topics documented:

binsROC . . . . .	1
timesROC . . . . .	4

<b>Index</b>	7
--------------	---

---

binsROC	<i>sMS ROC Curve estimation for binary outcomes</i>
---------	---

---

## Description

Provides estimations of the Receiver Operating-Characteristic (**ROC**) curve and the corresponding Area Under the Curve (**AUC**) based on the semi-parametric Mixed Subjects (**sMS**) **Approach**, when the logistic regression model is assumed as predictive model.

Confidence intervals for the AUC are also computed at the desired confidence level through three different estimation methods: the empirical approximation, the asymptotic distribution of the AUC and bootstrapping.

## Usage

```
binsROC(X, D, meth, grid, ci, cl, ci_meth, nboost)
```

## Arguments

X	vector of the (bio)marker/test values.
D	response vector with two different levels. The highest one stands for the subjects having the characteristic in study (positive). The lowest, for those who do not (negative).
meth	method for approximating the binary predictive model, $P(D X = x)$ , <ul style="list-style-type: none"> <li>• “L” for <b>L</b>inear logistic regression,</li> <li>• “S” for <b>S</b>mooth logistic regression (splines).</li> </ul>
grid	grid size for computing the Area Under the ROC Curve. Default value 1000.
ci	boolean variable indicating if a confidence interval for the AUC will be computed. The default value is FALSE.
cl	confidence level at which the confidence interval for the AUC will be provided. The default value is 95%. This parameter is ignored when ci is set to FALSE.
ci_meth	method for computing the AUC confidence interval. There are three options: <ul style="list-style-type: none"> <li>• “E” for the <b>E</b>mpirical estimation.</li> <li>• “V” for the <b>V</b>ariance of the asymptotic AUC distribution.</li> <li>• “B” for the <b>B</b>ootstrap approximation.</li> </ul> <p>The empirical method <b>E</b> is taken as default value and the parameter is ignored too when ci value is FALSE.</p>
nboost	number of bootstrap samples to be run when <b>B</b> ootstrap is set as ci_meth parameter. The default value is 500 and, like in the previous cases, it is not taken into account when no confidence interval is computed.

## Details

Marker and response vectors X and D are mandatory parameters with no default values. They must be the same length. Besides, the response vector D should have two different levels. In case of having more, only the two lowest will be taken into account.

The predictive model  $P(D|X = x)$  represents the relationship between the biomarker values and the binary response variable. If the **L**inear option is chosen as meth parameter, it will be approximated by a linear logistic regression model:

$$P(D|X = x) = 1/(1 + \exp\{-(\beta_0 + \beta_1 x)\}),$$

with  $\beta_0, \beta_1 \in \mathcal{R}$  and, when the **S**mooth one is selected, by the smooth logistic regression,

$$P(D|X = x) = 1/(1 + \exp\{-s(x)\}),$$

being  $s(\cdot)$  the smooth function (splines). Notice that the predictive model allows to compute the probability of being positive/negative even when the actual group is missing.

When the option of computing a confidence interval for the AUC is chosen (parameter ci set to TRUE), three different methods are offered to obtain the interval. The **E**mpirical approximation arises from the fact that, the AUC variance computed by the SMS Approach, is similar to the one based on the empirical estimator of the observed subjects, when a logistic regression model is assumed as predictive model. The **V**ariance method lies in the use of the variance from the asymptotic approximation of the AUC distribution. Finally, the **B**ootstrap approximation computes the confidence intervals for the AUC, via bootstrapping, by the generation of the chosen nboost samples.

**Value**

A list of class `sMSROC` is returned with the following values:

<code>th</code>	vector of thresholds at which the sensitivity and specificity are computed.
<code>TP</code>	vector of true positive (or sensitivity) at the considered thresholds <code>th</code> .
<code>FP</code>	vector of false positive (or complement to specificity) calculated at the thresholds <code>th</code> .
<code>Pi</code>	vector with the predictive model estimates. Notice that they are the probabilities of having the target characteristic, given the corresponding biomarker value.
<code>u</code>	sequence of points at which the ROC curve is evaluated.
<code>ROC</code>	ROC curve values computed at each point of the vector <code>u</code> .
<code>meth</code>	estimation method used for approximating the predictive model.
<code>t</code>	null. No value is returned.
<code>auc</code>	area under the corresponding ROC curve estimation.
<code>cl</code>	confidence level at which the confidence interval for the auc has been computed.
<code>ci_meth</code>	estimation method by which the confidence interval for the auc has been calculated.
<code>ci</code>	vector with two components: the lower and upper edges of the auc confidence interval, computed at confidence level <code>cl</code> and by <code>ci_meth</code> the method.
<code>mt</code>	two columns matrix with the warning messages from the parameters checking auxiliary functions. The first column stands for the parameter and the second one for the text of the message.

**Author(s)**

Susana Diaz-Coto, <U0266718@uniovi.es> and

Pablo Martinez Camblor, <pablo.martinez.camblor@dartmouth.edu>

**References**

Diaz-Coto, S.; Martinez-Camblor, P. and Corral-Blanco, N. *Semiparametric receiver operating-characteristic ROC curve estimation for cohort studies*.

Balakrishnan, N. (1991). Handbook of the Logistic Distribution, CRC Press, Statistics: A Series of Textbooks and Monographs, Boca Raton, FL.

**Examples**

```
require(pROC)
data(aSAH)

dt <- as.data.frame(cbind(aSAH$s100b, aSAH$outcome))
colnames(dt) <- c("s100b", "outcome")
R1 <- binsROC(X = dt$s100b, D = dt$outcome, meth = "S", ci=TRUE, ci_meth = "V")
```

timesROC

*sMS ROC Curve estimation for time-dependent outcomes*

## Description

This function provides estimations of the Cumulative/Dynamic(C/D) **ROC** curve at time **t** and the corresponding Area Under the Curve (**AUC**) based on the semi-parametric Mixed Subjects (**sMS**) **Approach**, when a proportional hazards model is assumed as predictive model.

Confidence intervals for the AUC are also computed at the desired confidence level through three different estimation methods: the empirical approximation, the asymptotic distribution of the AUC and bootstrapping.

## Usage

```
timesROC(X, D, Time, t, meth, grid, ci, cl, ci_meth, nboost)
```

## Arguments

X	vector of the (bio)marker/test values.
D	response vector with two different levels. The highest one stands for the subjects who have experienced the event in study at the observed time. The lowest, for those who still do not.
Time	vector with the observed times for each subject.
t	point of time at which the C/D ROC curve will be estimated.
meth	method for approximating the binary predictive model, $P(D X = x)$ , <ul style="list-style-type: none"> <li>• “L” for a proportional hazards Cox Regression model (<b>L</b>inear).</li> <li>• “S” for a proportional hazards Cox Regression model (<b>S</b>mooth).</li> </ul>
grid	grid size for computing the Area Under the ROC Curve. Default value 1000.
ci	boolean variable indicating if a confidence interval for the AUC will be computed. The default value is FALSE.
cl	confidence level at which the confidence interval for the AUC will be provided. The default value is 95%. This parameter is ignored when ci is set to FALSE.
ci_meth	method for computing the AUC confidence interval. There are three options: <ul style="list-style-type: none"> <li>• “E” for the <b>E</b>mpirical estimation.</li> <li>• “V” for the <b>V</b>ariance of the asymptotic AUC distribution.</li> <li>• “B” for the <b>B</b>ootstrap approximation.</li> </ul> <p>The empirical method <b>E</b> is taken as default value and the parameter is ignored too when the ci is FALSE.</p>
nboost	number of bootstrap samples to be run when <b>B</b> ootstrap is set as parameter ci_meth parameter. The default value is 500 and, like in the previous cases, it is not taken into account when no confidence interval is computed.

## Details

Marker, response and time vectors  $X$ ,  $D$  and  $T$  are mandatory parameters with no default values. They must be the same length. Besides, the response vector  $D$  should have two different levels. In case of having more, only the two lowest will be taking into account.

The predictive model  $P(D|X = x)$  represents the relationship between the biomarker values and the outcome, which is a time-dependent variable. Then, in this case,  $D = T \leq t$ , being  $T$  the Time variable and the predictive model takes the form  $P(T \leq t | X = x)$ . If the **Linear** option is chosen as **meth** parameter, it will be approximated by a proportional hazard Cox regression model:

$$P(T \leq t | X = x) = 1 - \exp\{-\Delta_0(t) \cdot \exp\{\beta_0 + \beta_1 \cdot \log(x)\}\},$$

where  $\Delta_0(\cdot)$  is the baseline hazard function and  $\beta_0, \beta_1 \in \mathcal{R}$ . When the **Smooth** one is selected, the approximation is done by

$$P(T \leq t | X = x) = 1 - \exp\{-\Delta_0(t) \cdot \exp\{s(x)\}\}$$

being  $s(\cdot)$  the smooth function (penalized splines).

When the option of computing a confidence interval for the AUC is chosen (parameter **ci** set to **TRUE**), three different methods are offered to obtain the interval. The **Empirical** approximation issues from the fact that the AUC variance computed by the **sMS Approach** is similar to the one based on the empirical estimator of the observed subjects, when a logistic regression model is assumed as predictive model. The **Variance** method lies in the use of the variance from the asymptotic approximation of the AUC distribution. Finally, the **Bootstrap** approximation computes the confidence intervals for the AUC, via bootstrapping, by the generation of the chosen **nboost** samples.

## Value

A list of class **sMSROC** is returned with the following values:

<b>th</b>	vector of thresholds at which the sensitivity and specificity are computed.
<b>TP</b>	vector of true positive (or sensitivity) at the considered thresholds <b>th</b> .
<b>FP</b>	vector of false positive (or complement to specificity) calculated at the thresholds <b>th</b> .
<b>Pi</b>	vector with the predictive model estimates. Notice that they are the probabilities of having the target characteristic, given the corresponding biomarker value.
<b>u</b>	sequence of points at which the ROC curve is evaluated.
<b>ROC</b>	c/D ROC curve values computed at each point of the vector <b>u</b> .
<b>meth</b>	estimation method used for approximating the predictive model.
<b>t</b>	point of time at which the C/D ROC curve has been computed.
<b>auc</b>	area under the corresponding ROC curve estimation.
<b>cl</b>	confidence level at which the confidence interval for the auc has been computed.
<b>ci_meth</b>	estimation method by which the confidence interval for the auc has been calculated.
<b>ci</b>	vector with two components: the lower and upper edges of the auc confidence interval, computed at confidence level <b>cl</b> and by <b>ci_meth</b> the method.
<b>mt</b>	two columns matrix with the warning messages from the parameters checking auxiliary functions. The first column stands for the parameter and the second one for the text of the message.

**Author(s)**

Susana Diaz-Coto, <U0266718@uniovi.es> and  
Pablo Martinez Camblor, <pablo.martinez.camblor@dartmouth.edu>

**References**

Diaz-Coto, S.; Martinez-Camblor, P. and Corral-Blanco, N. *Semiparametric receiver operating-characteristic ROC curve estimation for cohort studies*.

Cox. D.R. *Regression models and life-tables*. Journal of the Royal Statistical Society. Series B (Methodological), 34(2):187 (220, 1972).

Hurvich, C.M.; Simonoff J.S. and Tsai, C-L. *Smoothing parameter selection in non-parametric regression using an improved akaike information criterion*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 60(2):271(293, 1998).

**Examples**

```
require(survival)
data(pbc)
DT <- pbc
t_0 <- 3*365
DT$status1 <- ifelse(pbc$status==2,1,0)
R1 <- timesROC(X = DT$bili, D = DT$status1, Time = DT$time,
               t = t_0, meth = "L", ci = TRUE, ci_meth = "V")
```

# Index

binsROC, [1](#)

timesROC, [4](#)