Package 'APtools'

January 20, 2025

Type Package

Title Average Positive Predictive Values (AP) for Binary Outcomes and
Censored Event Times

Version 6.8.8

Depends graphics, stats, utils, survival, cmprsk

Author Hengrui Cai <hengruicai@gmail.com>, Yan Yuan <yyuan@ualberta.ca>, Qian
Michelle Zhou <qz70@msstate.edu>, Bingying Li<dorisli1120@gmail.com>

Maintainer Hengrui Cai <hengruicai@gmail.com>

Description We provide tools to estimate two prediction accuracy metrics,
the average positive predictive values (AP) as well as the well-known AUC
(the area under the receiver operator characteristic curve) for risk scores.
The outcome of interest is either binary or censored event time.
Note that for censored event time, our functions' estimates, the AP and the
AUC, are time-dependent for pre-specified time interval(s). A function that
compares the APs of two risk scores/markers is also included. Optional
outputs include positive predictive values and true positive fractions at
the specified marker cut-off values, and a plot of the time-dependent AP
versus time (available for event time data).

License LGPL-3

Encoding UTF-8

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2018-09-21 04:40:09 UTC

Contents

APBinary   *Estimating the AP and the AUC for Binary Outcome Data.*

Description

This function calculates the estimates of the AP and AUC for binary outcomes as well as their confidence intervals using the perturbation or the nonparametric bootstrap resampling method.

Usage

```
APBinary(status, marker, cut.values = NULL,
   method = "none", alpha = 0.95, B = 1000, weight = NULL)
```

Arguments

status	Binary indicator, 1 indicates case / the class of prediction interest and 0 otherwise.
marker	Numeric risk score. Data can be continuous or ordinal.
cut.values	risk score values to use as a cut-off for calculation of positive predictive values (PPV) and true positive fractions (TPF). The default value is NULL.
method	Method to obtain confidence intervals. The default is method = "none", in which case only point estimates will be given without confidence intervals. If method= "perturbation", then perturbation based CI will be calculated. If method = "bootstrap", then nonparametric bootstrap based CI will be calculated.
alpha	Confidence level. The default level is 0.95.
B	Number of resampling to obtain confidence interval. The default value is 1000.
weight	Optional. The default weight is 1, same object length as the "status" and "marker" object. Users can use their own weights, and the length of weight is required to be the same as the length of status.

Value

an object of class "APBinary" which is a list with components:

ap_summary	Summary of the AP, including the proportion of cases, a point estimate of AP, and their corresponding confidence intervals.
auc_summary	Summary of the AUC, including a point estimate of AUC with a confidence interval.
PPV	Available object, positive predictive values at the unique risk score in the data.
TPF	Available object, true positive fractions at the unique risk score in the data.

References

Yuan, Y., Su, W., and Zhu, M. (2015). Threshold-free measures for assessing the performance of medical screening tests. Frontiers in Public Health, 3.57.

Bingying Li (2015) Threshold-free Measure for Assessing the Performance of Risk Prediction with Censored Data, MSc. thesis, Simon Fraser University, Canada

Examples

```
status=c(rep(1,10),rep(0,1),rep(1,18),rep(0,11),rep(1,25),
rep(0,44),rep(1,85),rep(0,176))
marker=c(rep(7,11),rep(6,29),rep(5,69),rep(4,261))
cut.values=sort(unique(marker)[-1])
out1 <- APBinary(status,marker,cut.values)
out1
out2 <- APBinary(status,marker,method="perturbation",
alpha=0.90,B=1500)
out2
```

APSurv	*Estimating the Time-dependent AP and AUC for Censored Time to Event Outcome Data.*

Description

This function calculates the estimates of the AP and AUC for censored time to event data as well as their confidence intervals using the perturbation or the nonparametric bootstrap resampling method. The estimation method is based on Yuan, Y., Zhou, Q. M., Li, B., Cai, H., Chow, E. J., Armstrong, G. T. (2018). A threshold-free summary index of prediction accuracy for censored time to event data. Statistics in medicine, 37(10), 1671-1681.

Usage

```
APSurv(stime, status, marker, t0.list, cut.values = NULL,
  method = "none", alpha = 0.95, B = 1000,
  weight = NULL, Plot = TRUE)
```

Arguments

stime	Censored event time.
status	Binary indicator of censoring. 1 indicates observing event of interest, 0 otherwise. Other values will be treated as competing risk event.
marker	Numeric risk score. Data can be continuous or ordinal.
t0.list	Prediction time intervals of interest. It could be one numerical value or a vector of numerical values, which must be in the range of stime.
cut.values	Risk score values to use as a cut-off for calculation of time-dependent positive predictive values (PPV) and true positive fractions (TPF). The default value is NULL.
method	Method to obtain confidence intervals. The default is method = "none", in which case only point estimates will be given without confidence intervals. If method= "perturbation", then perturbation based CI will be calculated. If method = "bootstrap", then nonparametric bootstrap based CI will be calculated.
alpha	Confidence level. The default level is 0.95.

B	Number of resampling to obtain a confidence interval. The default value is 1000.
weight	Optional. The default value is NULL, in which case the observations are weighted by the inverse of the probability that their respective time-dependent event status (whether the event occurs within a specified time period) is observed. In estimating the probability, the survival function of the censoring time is estimated by a Kaplan-Meier estimator under the assumption that the censoring time is independent of both the event time and risks score. Users can use their own weights, in which case the t0.list should be a scalar and the length of weight is required to be the same as the length of status.
Plot	Whether to plot the time-dependent AP versus the prediction time intervals. The default value is TRUE, in which case the AP is evaluated at the time points which partition the range of the event times of the data into 100 intervals.

Value

An object of class "APsurv" which is a list with components:

ap_summary	Summary of estimated AP(s) at the specified prediction time intervals of interest. For each prediction time interval, the output includes the estimated event rate, a point estimate of the AP, the estimated scaled AP (ratio of the AP versus event rate), and their corresponding confidence intervals.
auc_summary	Summary of AUC at the specified prediction time intervals of interest. For each prediction time intervals, the output includes the estimated event rate and a point estimate of AUC with a confidence interval.
PPV	Available object, time-dependent positive predictive values at the unique risk score in the data.
TPF	Available object, time-dependent true positive fractions at the unique risk score in the data.

References

Yuan, Y., Zhou, Q. M., Li, B., Cai, H., Chow, E. J., Armstrong, G. T. (2018). A threshold-free summary index of prediction accuracy for censored time to event data. Statistics in medicine, 37(10), 1671-1681.

Bingying Li (2015) Threshold-free Measure for Assessing the Performance of Risk Prediction with Censored Data, MSc. thesis, Simon Fraser University, Canada

Examples

```
library(APtools)
data(mayo)
t0.list=seq(from=min(mayo[,1]),to=max(mayo[,1]),length.out=5)[-c(1,5)]
cut.values=seq(min(mayo[,3]),max(mayo[,3]),length.out=10)[-10]
out <- APSurv(stime=mayo[,1],status=mayo[,2],marker=mayo[,3],
t0.list=t0.list,cut.values=cut.values,method='bootstrap',
alpha=0.90,B=500,weight=rep(1,nrow(mayo)),Plot=FALSE)
out
```

CompareAP	*Comparison of two risk scores based on the differences and ratio of their APs.*

Description

This function estimates the difference between and the ratio of two APs in order to compare two markers for censored time to event data or binary data. The corresponding confidence intervals are provided.

Usage

```
CompareAP(status, marker1, marker2, stime = NULL,
    t0.list = NULL, method = "none", alpha = 0.95,
    B = 1000, weight = NULL, Plot = TRUE)
```

Arguments

status	Binary indicator. For binary data, 1 indicates case and 0 otherwise. For survival data, 1 indicates event and 0 otherwise.
marker1	Risk score 1 (to be compared to risk score 2). Its length is required to be the same as the length of status.
marker2	Risk score 2 (to be compared to risk score 1). Its length is required to be the same as the length of status.
stime	Censored event time. If dealing with binary outcome, skip this argument which is set to be NULL.
t0.list	Prediction time intervals of interest for event time outcome. It could be one numerical value or a vector of numerical values, which must be in the range of stime. It is set to be NULL if stime is NULL.
method	Method to obtain confidence intervals. The default is method = "none", in which case only point estimates will be given without confidence intervals. If method= "perturbation", then perturbation based CI will be calculated. If method = "bootstrap", then nonparametric bootstrap based CI will be calculated.
alpha	Confidence level. The default level is 0.95.
B	Number of resampling for obtaining a confidence interval. The default value is 1000.
weight	Optional argument for event time data, i.e. stime is not NULL. Its default value is NULL, in which the observations are weighted by the inverse of the probability that their respective time-dependent event status (whether the event occurs within a specified time period) is observed. In estimating the probability, the survival function of the censoring time is estimated by a Kaplan-Meier estimator under the assumption that the censoring time is independent of both the event time and risks score. Users can use their own weights, in which case the t0.list should be a scalar and the length of weight is required to be the same as the length of status.

Plot   Optional argument for event time data, i.e. stime is not NULL. For binary data,
                    it is set to FALSE. For event time data, its default value is TRUE and three
                    plots are generated: 1) the time-dependent AUC of two markers; 2) the time-
                    dependent AP of two markers; and 3) the time-dependent ratio of APs, all versus
                    the prediction time intervals. The quantities in 1)-3) are evaluated at the time
                    points which partition the range of the event times of the data to 100 intervals.

Value

dap_summary   Summary of the APs of two markers and the differences (AP1-AP2) and their
                    ratio (AP1/AP2). For event time data, these quantities are estimated at the
                    specified prediction time intervals. The output includes the estimated event
                    rate/proportion of cases, point estimates of the APs of the two markers, point
                    estimates of the difference between and ratio of the two APs as well as their
                    respective confidence intervals.

References

Yuan, Y., Zhou, Q. M., Li, B., Cai, H., Chow, E. J., Armstrong, G. T. (2018). A threshold-free
summary index of prediction accuracy for censored time to event data. Statistics in medicine,
37(10), 1671-1681.

Yuan, Y., Su, W., and Zhu, M. (2015). Threshold-free measures for assessing the performance of
medical screening tests. Frontiers in Public Health, 3.57.

Bingying Li (2015) Threshold-free Measure for Assessing the Performance of Risk Prediction with
Censored Data, MSc. thesis, Simon Fraser University, Canada

Examples

```
library(APtools)
status=c(rep(1,10),rep(0,1),rep(1,18),rep(0,11),rep(1,25),
rep(0,44),rep(1,85),rep(0,176))
marker1=c(rep(7,11),rep(6,29),rep(5,69),rep(4,261))
marker2=c(rep(7,17),rep(6,29),rep(5,70),rep(4,254))
out_binary <- CompareAP(status,marker1,marker2)
out_binary
data(mayo)
t0.list=seq(from=min(mayo[,1]),to=max(mayo[,1]),length.out=5)[-c(1,5)]
out_survival <- CompareAP(status=mayo[,2],marker1=mayo[,3],
marker2=mayo[,4],stime=mayo[,1],t0.list=t0.list,
method='bootstrap',alpha=0.90,B=500,
weight=rep(1,nrow(mayo)),Plot=FALSE)
out_survival
```

---

mayo   *Mayo Marker data*

Description

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

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.

Source

T Therneau, P Grambsch (2000) Modeling Survival Data: Extending the Cox Model *Springer-Verlag*, New York, ISBN: 0-387-98784-3.

References

Fleming T, Harrington D. (1991) Counting Processes and Survival Analysis *Wiley*, New York.

Heagerty, P.J., Zheng, Y. (2005) Survival Model Predictive Accuracy and ROC Curves *Biometrics*, **61**, 92 – 105

Index