Package: frailtyROC (via r-universe)

October 10, 2026

Type Package

Title Time-Dependent ROC Curve Estimation for Correlated Right-Censored Survival Data

Version 1.0.0

Description This contains functions that can be used 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 for correlated right-censored time-to-event data. See Beyene and Chen (2024) <doi:10.1177/09622802231220496>.

License GPL (>= 2)

Depends R(>= 4.2)

Imports frailtypack, stats, survival, xpectr

Encoding UTF-8

LazyData true

RoxygenNote 7.3.3

NeedsCompilation no

Author Kassu Mehari Beyene [aut, cre] (ORCID: <https://orcid.org/0000-0002-2067-6054>), Ding-Geng Chen [ctb]

Maintainer Kassu Mehari Beyene <m2kassu@gmail.com>

Config/pak/sysreqs cmake make libicu-dev libuv1-dev libx11-dev zlib1g-dev

Repository https://kmbeyene.r-universe.dev

Date/Publication 2025-10-14 18:20:09 UTC

RemoteUrl https://github.com/cran/frailtyROC

RemoteRef HEAD

RemoteSha 3f37f0fefe10addf63bd88bc547b492787b36540

Contents

frailtyROC-package 2
frailtyROC 3
kidney 6
LungCancer 7
wbw 8

Index **10**

`frailtyROC-package` *frailtyROC*

Description

frailtyROC is a tool for estimating and visualizing time-dependent receiver operating characteristic (ROC) curves, as well as the corresponding time-dependent area under the curve (AUC), in the context of correlated right-censored time-to-event data. Confidence bands for the ROC curve and confidence intervals for the AUC can be constructed using bootstrap-derived standard errors.

Details

frailtyROC is a comprehensive tool for estimating and visualizing time-dependent receiver operating characteristic (ROC) curves and their corresponding area under the curve (AUC) in the context of correlated right-censored time-to-event data. The ROC curves can be estimated either empirically (non-smoothed) or smoothed with or without boundary correction. For the latter case, the data-driven smoothing parameter selection methods introduced by Beyene and El Ghouch (2020) for smoothed ROC curves are implemented, offering an automatic approach to bandwidth selection during the smoothing process.

The package enables the estimation of time-dependent ROC curves at specific time points to evaluate the discriminatory ability of prognostic models or biomarkers. The time-dependent AUC serves as a summary measure of predictive accuracy over time. Confidence bands for the ROC curves and confidence intervals for AUC estimates are obtained via non-parametric bootstrap sampling based on the percentile method. This approach yields standard error estimates, which serve as the basis for bootstrap-based hypothesis testing of AUC at a given time point.

An essential component of the methodology involves estimating the survival function conditional on the observed data. This is accomplished through the use of shared frailty models specifically developed for correlated censored data. To this end, both semi-parametric and parametric frailty models are estimated using a penalized likelihood estimation framework that implemented in Rondeau et al. (2012). Users may specify either a gamma or log-normal frailty distribution.

Abbreviations

In this package, the following abbreviations are commonly used:

ROC Receiver Operating Characteristic curve.

AUC Area under the ROC curve at a given time horizon `t`.

Dataset

This package comes with a correlated right-censored marker data sets. For details see `kidney` and `LungCancer`

Installing and using

Ensure that your system has an active internet connection, then execute the following command in the R console to install the package:

```
install.packages("frailtyROC")
```

To load the package after installation, use the following command:

```
library(frailtyROC)
```

Author(s)

Kassu Mehari Beyene and Ding-Geng Chen

Maintainer: Kassu Mehari Beyene <m2kassu@gmail.com>

References

Beyene, K. M., and Chen, D. G. (2024). Time-dependent receiver operating characteristic curve estimator for correlated right-censored time-to-event data. *Statistical Methods in Medical Research*, 33(1), 162-181.

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.

Rondeau, V., Marzroui, Y., & Gonzalez, J. R. (2012). frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47, 1-28.

`frailtyROC` *Estimation of the time-dependent ROC curve for correlated right-censored survival data*

Description

This function computes a time-dependent ROC curve for correlated 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

```
frailtyROC(
  Y,
  M,
  censor,
  group = NULL,
  w,
  t = 1e-04,
  U = NULL,
  bw = "NR",
  len = 151,
  method = "tra",
  method1 = "marg",
  ktype = "normal",
  knots = 10,
  kappa = 10000,
  RandDist = "Gamma",
  hazard = "Splines",
  maxit = 300,
  B = 0,
  alpha = 0.05,
  plot = "TRUE"
)
```

Arguments

Y	a numeric vector of event-times or observed times.
M	a numeric vector of (bio)marker or risk score values.
censor	a vector of censoring indicator, 1 if event, 0 otherwise.
group	a categorical vector of group/cluster.
w	a scalar window for prediction.
t	a scalar time for prediction. The default value is 0.0001.
U	a vector of grid points where the ROC curve is estimated. The default is a sequence of 151 numbers between 0 and 1.
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 give a numerical value.
len	a scalar value specifying the length of vector U. The default value is 151.
method	a character string specifying the method of ROC curve estimation. The possible options are "emp" empirical method; "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.
method1	a character string specifying prediction method applied on model. The possible options are "cox" for the classical Cox; "marg" for marginal and "cond" conditional prediction methods on shared models. The default is "cond".

ktype	a character string specifying the type kernel distribution to be used for smoothing the ROC curve: "normal", "epanechnikov", "biweight", or "triweight". By default, the "normal" kernel is used.
knots	a scalar for specifying the number of knots to use. Value required in the penalized likelihood estimation. It corresponds to the (knots+2) splines functions for the approximation of the hazard or the survival functions. Rondeau, et al. (2012) suggested that the number of knots must be between 4 and 20. The default is 10.
kappa	a positive smoothing parameter value for the penalized likelihood estimation. The defaults is "10000".
RandDist	a character string to state the distribution of random effect: "Gamma" for a gamma distribution, "LogN" for a log-normal distribution. Default is "Gamma".
hazard	types of hazard functions: "Splines" represents a semi-parametric hazard function using equidistant intervals and is estimated via penalized likelihood, "Splines-per" uses percentiles instead of equidistant intervals, "Piecewise-per" is a piecewise constant hazard function based on percentiles, "Piecewise-equi" is a piecewise constant hazard using equidistant intervals, and "Weibull" is a fully parametric hazard function based on the Weibull distribution. "Splines" is used as the default setting.
maxit	maximum number of iterations. The default is 300.
B	a 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	a logical parameter to see the ROC curve plot. The default is TRUE.

Details

This function takes correlated right-censored survival data and returns an empirical (non-smoothed) ROC estimate and the smoothed time-dependent ROC estimate with/without boundary correction and the corresponding time-dependent AUC estimates. For the smoothing parameter estimation, three data-driven methods: the normal reference "NR", the plug-in "PI" and the cross-validation "CV" introduced in Beyene and El Ghouch (2020) were implemented. See Beyene and El Ghouch (2020) for details. The conditional survival function estimation can done by using semi-parametric or fully parametric shared frailty models.

Value

Returns the following items:

ROC	vector of estimated ROC values. These will be numeric values between zero and one.
AUC	data frame of dimension 1×4. The columns are: AUC, standard error of AUC, the lower and upper limits of bootstrap CI.
U	vector of grid points used.
bw	computed value of bandwidth parameter. For the empirical method this is always NA.
Dt	vector of estimated event status.
M	vector of (bio)marker vlaues.

References

Beyene, K. M., and Chen, D. G. (2024). Time-dependent receiver operating characteristic curve estimator for correlated right-censored time-to-event data. *Statistical Methods in Medical Research*, 33(1), 162-181.

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.

Rondeau, V., Marzroui, Y., & Gonzalez, J. R. (2012). frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *Journal of Statistical Software*, 47, 1-28.

Examples

```
 library(frailtyROC)

data(kidney)

out1 <- frailtyROC(Y=kidney$time, M=kidney$Marker, censor=kidney$status, group = kidney$id, w=120,
       method = "emp", method1 = "cond", hazard = "Weibull")

out1$AUC
```

`kidney` *Kidney Marker Data*

Description

This dataset contains four columns: `id`, patient code; `marker`, the risk score; `time`, the observed follow-up time; and `status`, the event indicator for subjects in the kidney marker dataset.

Usage

```
data(kidney)
```

Format

This is a data frame with 76 observations and the following 4 variables.

id patient code

time time to recurrence of infection

status censoring indicator; 0=censored, 1=recurrence

marker risk score derived from observed data using frailty model

Details

This dataset pertains to the recurrence of infection in patients with kidney disease who use portable dialysis equipment. Recurrent infection is the primary complication in these patients, typically occurring at the catheter insertion site. When an infection occurs, the catheter is removed and reinserted after successful treatment. In some cases, the catheter is removed for reasons unrelated to infection; these observations are treated as censored. A total of 38 patients were followed to assess the time to recurrence of infection. Each patient contributes exactly two observations.

The dataset includes the following covariates: `sex` (1 = Male, 2 = Female), `age`, and `disease` (Disease type; a factor with four levels: `"GN"`, `"AN"`, `"PKD"`, and `"Other"`). The marker is derived using a gamma frailty model as follows:

$$\hat{\nu}\exp(\hat{\beta}_1\,\mathrm{sex} + \hat{\beta}_2\,\mathrm{age} + \hat{\beta}_3\,\mathrm{GN} + \hat{\beta}_4\,\mathrm{AN} + \hat{\beta}_5\,\mathrm{PKD})$$

where $\hat{\nu}$ is the estimated frailty term, and $\hat{\beta}_i$(for `i`=1,2,3) are the estimated regression coefficients from the frailty model.

References

Beyene, K. M., and Chen, D. G. (2024). Time-dependent receiver operating characteristic curve estimator for correlated right-censored time-to-event data. *Statistical Methods in Medical Research*, 33(1), 162-181.

`LungCancer` *NCCTG Lung Cancer Marker Data*

Description

This dataset contains four columns: `id`, the identifier for health institutions (clusters); `marker`, the risk score; `time`, the observed follow-up time; and `status`, the event indicator for subjects in the NCCTG lung cancer marker dataset.

Usage

```
data(LungCancer)
```

Format

This is a data frame with 238 observations and the following 4 variables.

id health institutions code

time time to death in days

status censoring indicator; 1=censored, 2=dead

marker risk score derived from the observed data using frailty model

Details

The NCCTG lung cancer dataset was collected from 228 patients across 18 different healthcare institutions. The number of subjects per institution ranged from 2 to 36. For the final analysis, only 226 patients with complete records were included. The dataset contains survival times along with several important predictor variables, including: sex (coded as Male = 1, Female = 2), age (in years), ph.ecog (Eastern Cooperative Oncology Group performance status, assessed by a physician on a scale from 0 [asymptomatic] to 5 [dead]), and pat.karno (Karnofsky performance status, assessed by the patient).

The marker (risk score) was derived from three predictor variables: sex, age, and ph.ecog. To this end, a frailty model with gamma-distributed frailty was fitted. As in the previous example, the prognostic marker is defined as:

$$\hat{\nu} \exp(\hat{\beta}_1 \cdot \text{sex} + \hat{\beta}_2 \cdot \text{age} + \hat{\beta}_3 \cdot \text{ph.ecog})$$

where $\hat{\nu}$ is the estimated frailty term, and $\hat{\beta}_i$(for i=1,2,3) are the estimated regression coefficients from the frailty model.

References

Beyene, K. M., and Chen, D. G. (2024). Time-dependent receiver operating characteristic curve estimator for correlated right-censored time-to-event data. *Statistical Methods in Medical Research*, 33(1), 162-181.

---

wbw *Function to select the bandwidth parameter needed for smoothing the time-dependent ROC curve.*

Description

This function computes a data-driven bandwidth for smoothing the ROC curve, supporting three methods: the normal reference method, the plug-in method, and the cross-validation method introduced in Beyene and El Ghouch (2020). It is particularly important for estimating the bandwidth in the presence of weighted data.

Usage

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

Arguments

X	numeric data vector.
wt	non-negative weight vector.
bw	a character string specifying the bandwidth selection method. The possible options are "NR" for the normal reference, the plug-in "PI" and cross-validation "CV".
ktype	a character string indicating the type of kernel function: "normal", "epanechnikov", "biweight", or "triweight". Default is "normal" kernel.

Value

Returns the estimated value for the bandwidth parameter.

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(frailtyROC)

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

# Normal reference bandwidth selection

wbw(X = X, wt = wt)$bw
```

Index

∗ **datasets**
kidney, 6
LungCancer, 7
∗ **package**
frailtyROC-package, 2

frailtyROC, 3
frailtyROC-package, 2

kidney, *3*, 6

LungCancer, *3*, 7

wbw, 8