

Package ‘ROct’

December 20, 2013

Type Package

Title Time-dependent ROC curve estimation and adaptation to the relative survival context

Version 0.8

Date 2013-12-20

Author Y. Foucher, K. Trebern-Launay and M. Lorent

Maintainer Y. Foucher <Yohann.Foucher@univ-nantes.fr>

Description This package contains functions in order to estimate time-dependent ROC with right-censored data. More precisely, two functions concern the estimation of ROC curves defined by Heagerty, Lumley and Pepe (Biometrics, 2000). The two other functions concern their adaptation for studying the capacity of a marker to predict the excess of mortality of a specific population compared to the general population (same age, gender and calendar year). This last part is based on additive relative survival models and the work of Pohar-Perme (Biometrics, 2011).

License GPL (>=2)

LazyLoad yes

Depends splines, date, survival, relsurv

URL www.r-project.org, www.divat.fr

R topics documented:

ROct-package	2
allcause.ROct	3
dataDIVAT	5
fr.ratetable	6
net.ROct	7
rein.ratetable	9

Index	11
--------------	-----------

ROct-package

*Time-dependent ROC curves estimation***Description**

Compute time-dependent ROC curve using Kaplan-Meier (KM) estimator or the k-nearest neighbor's (KNN) adaptation. Both approaches are developed for traditional survival analysis (all-cause analysis) and for the the additive relative survival analysis.

Details

Package:	ROct
Type:	Package
Version:	0.8
Date:	2013-12-20
License:	GPL (>=2)
LazyLoad:	yes

Compute time-dependent ROC curve using Kaplan-Meier (KM) or the k-nearest neighbor's (KNN) adaptation. Both approaches are developed for traditional survival analysis (all-cause analysis) and for the additive relative survival analysis:

allcause.ROct This function performs the characteristics of a time-dependent ROC curve.
 net.ROct This function performs the characteristics of a net time-dependent ROC curve .

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

- Heagerty PJ., Lumley T., Pepe MS. (2000) Time-dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics*, 56, 337-344.
- Pohar M., Stare J. (2006) Relative survival analysis in R. *Computer methods and programs in biomedicine*, 81, 272-278.
- Pohar M., Stare J., Esteve J. (2012) On Estimation in Relative Survival. *Biometrics*, 68, 113-120.
- Akritas MG. (1994) Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, 22, 1299-1327.
- Lorent M., Giral M., Foucher Y. (2013) Net time-dependent ROC curves: a solution for evaluating the accuracy of a marker to predict disease-related mortality. *Statistics in Medicine*. In press.

See Also

URL: <http://www.divat.fr>

allcause.ROct *Time-dependent ROC curves with right censored data*

Description

This function performs the characteristics of a time-dependent ROC curve based on k-nearest neighbor's (knn) estimator or only based on the Kaplan and Meier estimator.

Usage

```
allcause.ROct(times, failures, variable, pro.time, cut.off,
              knn=FALSE, lambda=NULL)
```

Arguments

times	A numeric vector with the follow up times.
failures	A numeric vector with the event indicator (0=right censored, 1=event).
variable	A numeric vector with the prognostic variable. This variable is collected at the baseline.
pro.time	The value of prognostic time represents the maximum delay for which the capacity of the variable is evaluated. The same unit than the one used in the argument time.
cut.off	The cut-off value(s) of the variable used to define the binary test(s).
knn	A logical value indicating whether k-nearest neighbor's estimator should be used.
lambda	This is the proportion of the nearest neighbors. The estimation will be based on $2 \times \lambda$ (1 lambda on the left and 1 lambda on the right) of the total sample size. This parameter will only be used in the k-nearest neighbor's estimation (knn=TRUE).

Details

This function computes time-dependent ROC curve with right-censoring data. It can use Akritas approach (nearest neighbor's estimation) for ensuring monotone and increasing ROC curve. This Akritas approach may be avoid if the sample size is large because of computing time. The theory was defined by Heagerty, Lumley and Pepe (Biometrics, 2000).

Value

table	This data frame presents the sensitivities and specificities associated with the cut-off values. One can observe NA if the value cannot be computed.
auc	The area under the time-dependent ROC curve for a prognostic up to prognostic time.
missing	Number of deleted observations due to missing data.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

Heagerty PJ., Lumley T., Pepe MS. (2000) Time-dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics*, 56, 337-344.

Akritas MG. (1994) Nearest neighbor estimation of a bivariate distribution under random censoring. *Annals of Statistics*, 22, 1299-1327.

Examples

```
# import and attach the data example

data(dataDIVAT)

# A subgroup analysis to reduce the time needed for this example

dataDIVAT <- dataDIVAT[1:400,]

# cut-off values definition (choose more values in practice)

age.cut <- quantile(dataDIVAT$ageR, probs=seq(0.0, 1.0, by=0.2))

# the ROC curve (with the knn correction) to predict the all-cause
# mortality up to the 3000 days

roc1 <- allcause.ROct(times=dataDIVAT$death.time,
  failures=dataDIVAT$death, variable=dataDIVAT$ageR,
  pro.time=3000, cut.off=age.cut, lambda=0.05,
  knn=TRUE)

# the sensibilities and specificities associated with the cut off values

roc1$table

# the ROC curve (without the knn correction) to predict the all-cause
# mortality up to the 3000 days

roc2 <- allcause.ROct(times=dataDIVAT$death.time,
  failures=dataDIVAT$death, variable=dataDIVAT$ageR,
  pro.time=3000, cut.off=age.cut, lambda=NULL,
  knn=FALSE)

# the traditional ROC graph

plot(c(1,1-roc1$table$sp,0), c(1, roc1$table$se, 0), ylim=c(0,1),
  xlim=c(0,1), ylab="sensitivity", xlab="1-specificity",
  type="l", col=2, lty=1, lwd=2)

lines(c(1,1-roc2$table$sp,0), c(1,roc2$table$se,0),
  type="l", col=4, lty=2, lwd=2)

abline(c(0,0), c(1,1), lty=2)

legend("bottomright", c(
  paste("With knn, (AUC=", round(roc1$auc, 2), ")"), sep=""),
  paste("Without knn, (AUC=", round(roc2$auc, 2), ")"), sep=""),
  lty=1:2, lwd=c(2,2), col=c(2,4))
```

`dataDIVAT`*A sample of the DIVAT cohort*

Description

A data frame with 4267 French kidney transplant recipients. The time-to-event is the time between the transplantation and the recipient death. This time can be right-censored. A vector of covariates is also collected at the transplantation.

Usage

```
data(dataDIVAT)
```

Format

A data frame with 4267 observations (rows) with the 8 following variables (columns).

`ageR` This numeric vector represents the age of the recipient (in years)

`sexeR` This numeric vector represents the gender of the recipient (1=men, 0=female)

`year.tx` This numeric vector represents the year of the transplantation

`ante.diab` This numeric vector represents the diabetes statute (1=yes, 0=no)

`pra` This numeric vector represents the pre-graft immunization using the panel reactive antibody (1=detectable, 0=undetectable)

`ageD` This numeric vector represents the age of the donor (in years)

`death.time` This numeric vector represents the follow up time (until death or censoring)

`death` This numeric vector represents the death indicator at the follow-up end (1=death, 0=alive)

Details

The immunology and nephrology department of the Nantes University hospital constituted a data bank with the monitoring of medical records for kidney and/or pancreas transplant recipients. The associated software is called DIVAT. Here is a sample of 4267 patients. The time-to-event is the time between the transplantation and the death of the recipient. This time can be right-censored. A vector of covariates, all measured at the transplantation, is also collected for each patient.

Source

URL: <http://www.divat.fr>.

Examples

```
data(dataDIVAT)
```

```
### a short summary of the recipient age at the transplantation
summary(dataDIVAT$ageR)
```

```
### Kaplan and Meier estimation of the recipient survival
plot(survfit(Surv(death.time/365.25, death) ~ 1, data = dataDIVAT),
     xlab="Post transplantation time (in years)", ylab="Patient survival",
     mark.time=FALSE)
```

`fr.ratetable`*Expected mortality rates of the general French population.*

Description

An object of class `ratetable` for the expected mortality of the French population. It is a array with three dimensions: age, sex and year.

Usage

```
data(fr.ratetable)
```

Format

The format is "ratetable". The attributes are:

`dim` A numeric vector with the length of each dimension.

`dimnames` A list of vectors with the names of each variable of the three dimensions.

`dimid` A character vector with the identification of the dimensions: age, year and sex.

`factor` A vector of indicators=1 if the corresponding dimension do not vary according to the time. Only the sex is associated to 1.

`cutpoints` A list of the thresholds to identify the mortality rates according to the time-varying dimensions (NULL for sex).

`class` The class of the object: `ratetable`.

Details

The organization of a `ratetable` object is described in details by Therneau (1999) and Pohar (2006).

Source

The original data and updates can be downloaded from the Human Life-Table Database (HMD, The Human Mortality Database). URL: <http://www.mortality.org/>

References

T. Therneau, J. Offord. Expected Survival Based on Hazard Rates (Update), Technical Report, Section of Biostatistics, Mayo Clinic 63, 1999.

M. Pohara, J.Stare. Relative survival analysis in R. Computer methods and programs in biomedicine, 81: 272-278, 2006.

Examples

```
data(fr.ratetable)
```

```
is.ratetable(fr.ratetable)
```

net.ROct

Net time-dependent ROC curves with right censored data

Description

This function performs the characteristics of a net time-dependent ROC curve based on k-nearest neighbor's (knn) estimator or only based on the Pohar-Perme estimator (Pohar, 2012).

Usage

```
net.ROct(times, failures, variable, p.age, p.sex, p.year,
         rate.table, pro.time, cut.off, knn=FALSE,
         lambda=NULL)
```

Arguments

times	A numeric vector with the follow up times.
failures	A numeric vector with the event indicator (0=right censored, 1=event).
variable	A numeric vector with the prognostic variable. This variable is collected at the baseline.
p.age	A numeric vector with the age of the individuals at the baseline (in days).
p.sex	A character vector with the gender the individuals ('male' or 'female').
p.year	A numeric vector with the calendar year at the baseline (number of days from the January 1, 1960).
rate.table	A ratetable object with the expected mortality rates by age, sex, and cohort year. The same units used in p.age, p.sex, p.year
pro.time	The value of prognostic time represents the maximum delay for which the capacity of the variable is evaluated. The same unit than the one used in the argument time.
cut.off	The cut-off value(s) of the variable used to define the binary test(s).
knn	A logical value indicating whether k-nearest neighbor's estimator should be used.
lambda	This is the proportion of the nearest neighbors. The estimation will be based on $2 \cdot \lambda$ (1 lambda on the left and 1 lambda on the right) of the total sample size.

Details

This function computes net time-dependent ROC curve with right-censoring data using estimator defined by Pohar-Perme et al. (2011) and the k-nearest neighbor's (knn) estimator. The aim is to evaluate the capacity of a variable (measured at the baseline) to predict the excess of mortality of a studied population compared to the general population mortality. Using the knn estimator ensures a monotone and increasing ROC curve, but the computation time may be long. This approach may thus be avoided if the sample size is large because of computing time.

Value

table	This data frame presents the sensitivities and specificities associated with the cut-off values. One can observe NA if the value cannot be computed.
auc	The area under the time-dependent ROC curve for a prognostic up to prognostic time.
missing	Number of deleted observations due to missing data.
warning	This message indicates possible difficulties in the computation of the net ROC curve, for instance if the net survival was not lower or equal 1 for particular cut-off values or times.

Author(s)

Y. Foucher <Yohann.Foucher@univ-nantes.fr>

References

- Pohar M., Stare J., Esteve J. (2012) On Estimation in Relative Survival. *Biometrics*, 68, 113-120.
- Lorent M., Giral M., Foucher Y. (2013) Net time-dependent ROC curves: a solution for evaluating the accuracy of a marker to predict disease-related mortality. *Statistics in Medicine*. In press.

Examples

```
# import the observed data

data(dataDIVAT)

# A subgroup analysis to reduce the time needed for this exemple

dataDIVAT <- dataDIVAT[1:400,]

# import the expected mortality rates

data(fr.ratetable)

# the values of recipient age used for computing
# the sensibilities and specificities
# (choose more values in practice)

age.cut <- quantile(dataDIVAT$ageR, probs=seq(0.0, 1.0, by=0.2))

# recoding of the variables for matching with the ratetable

dataDIVAT$sex <- "male"
dataDIVAT$sex[dataDIVAT$sexeR==0] <- "female"
dataDIVAT$year <- mdy.date(month=01, day=01, year=dataDIVAT$year.tx,
  nineteen = TRUE, fillday = FALSE, fillmonth = FALSE)
dataDIVAT$age <- dataDIVAT$ageR*365

# the ROC curve (with the knn correction) to predict the
# mortality excess up to the 3000 days.

roc1 <- net.ROct(times=dataDIVAT$death.time,
  failures=dataDIVAT$death, variable=dataDIVAT$ageR,
```



```

p.age=dataDIVAT$age, p.sex=dataDIVAT$sex, p.year=dataDIVAT$year,
rate.table=fr.ratetable, pro.time=3000, cut.off=age.cut, knn=TRUE,
lambda=0.05)

# the sensibilities and specificities associated with the cut off values

roc1$table

# the ROC curve (without the knn correction) to predict the
# mortality excess up to the 3000 days.

roc2 <- net.ROCt(times=dataDIVAT$death.time,
failures=dataDIVAT$death, variable=dataDIVAT$ageR,
p.age=dataDIVAT$age, p.sex=dataDIVAT$sex, p.year=dataDIVAT$year,
rate.table=fr.ratetable, pro.time=3000, cut.off=age.cut, knn=FALSE,
lambda=NULL)

# the traditional ROC graph

plot(c(1,1-roc1$table$sp,0), c(1,roc1$table$se,0), ylim=c(0,1),
xlim=c(0,1), ylab="sensitivity", xlab="1-specificity", type="l",
col=2, lwd=2, lty=1)
lines(c(1,1-roc2$table$sp,0), c(1,roc2$table$se,0), type="l",
col=4, lwd=2, lty=2)

abline(c(0,0), c(1,1), lty=2)

legend("bottomright", c(
paste("With knn, (AUC=", round(roc1$auc, 2), ")"), sep=""),
paste("Without knn, (AUC=", round(roc2$auc, 2), ")"), sep=""),
lty=1:2, lwd=c(2,2), col=c(2,4))

```

rein.ratetable	<i>Expected mortality of French patients with End Stage Kidney Disease (ESKD) in dialysis and registered previously on waiting list for renal transplantation</i>
----------------	---

Description

An object of the class `ratetable` which contains the expected mortality of French patients with End Stage Kidney Disease (ESKD) in dialysis and registered previously on waiting list for renal transplantation. It is an array with two dimensions: the time since the registry on waiting list and the patient profile at this baseline.

Usage

```
data(rein.ratetable)
```

Format

The format is "ratetable". The attributes are:

`dim` A numeric vector with the length of each dimension.

`dimnames` A list of vectors with the names of each variable of both dimensions (see details for modalities of profile).

`dimid` A character vector with the identification of the dimensions: `time`, `profile`.

`factor` A vector of indicators=1 if the corresponding dimension do not vary according to the time. The profile is associated to 1.

`cutpoints` A list of the thresholds to identify the mortality rates according to the time-varying dimensions (NULL for profile).

`class` The class of the object: `ratetable`.

Details

The organization of a `ratetable` object is described in details by Therneau (1999) and Pohar (2006). We have estimated these data by using the renal epidemiology and information network (REIN). A competing risk model was performed in order to deal with the loss of follow-up to the transplantation. The profile represents the concatenation of the significant binary variables associated with the mortality on waiting list. In order to obtain the same variable in your database, we have to concatenate the following values in the same order. These values are collected at the registration date:

- a. 1 if the patient is male and 0 if female.
- b. 1 if the time between the first dialysis and the registration is higher than 1 year and 0 otherwise.
- c. 1 if hemodialysis and 0 if peritoneal dialysis.
- d. 1 if the patient is diabetic and 0 otherwise.
- e. 1 if the patient had history of cardiac failure diabetic and 0 otherwise.
- f. 1 if the patient had history of vascular failure diabetic and 0 otherwise.
- g. 1 if the patient had history of cancer and 0 otherwise.
- h. 1 if the age of the patients is in between 40 and 60 years and 0 otherwise.
- i. 1 if the age of the patients is higher than 60 years and 0 otherwise.
- j. 1 if the calendar year at registration is strictly higher than 2007 and 0

Source

URL: <http://www.soc-nephrologie.org/REIN/>

References

Pohar M., Stare J. (2006) Relative survival analysis in R. Computer methods and programs in biomedicine, 81, 272-278.

Examples

```
data(rein.ratetable)
```

```
is.ratetable(rein.ratetable)
```

Index

- *Topic **Akritas**
 - allcause.ROct, 3
 - net.ROct, 7
 - *Topic **Divat**
 - dataDIVAT, 5
 - *Topic **France**
 - fr.ratetable, 6
 - *Topic **Net survival**
 - net.ROct, 7
 - *Topic **ROC**
 - allcause.ROct, 3
 - net.ROct, 7
 - ROct-package, 2
 - *Topic **Time-dependent**
 - allcause.ROct, 3
 - *Topic **datasets**
 - dataDIVAT, 5
 - fr.ratetable, 6
 - rein.ratetable, 9
 - *Topic **ratetable**
 - fr.ratetable, 6
 - rein.ratetable, 9
- allcause.ROct, 3
- dataDIVAT, 5
- fr.ratetable, 6
- net.ROct, 7
- rein.ratetable, 9
- ROct-package, 2