# Package 'ROCt'

September 17, 2014

**Type** Package

**Title** Time-dependent ROC curve estimators and expected utility

**Version** 0.9

**Date** 2014-09-17

**Author** Y. Foucher, E. Dantan, P. Tessier 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). In addition, we propose a function for optimal cut-off estimation for medical decision making by maximizing time-dependent expected utility function.

**License** GPL (>=2)

**LazyLoad** yes

**Depends** splines, date, survival, relsurv

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

## R topics documented:

ROCt-package *Time-dependent ROC curves estimation*

## Description

Compute time-dependent ROC curve and expected utility for medical decision making.

## Details

| | |
|---|---|
| Package: | ROCt |
| Type: | Package |
| Version: | 0.9 |
| Date: | 2014-09-17 |
| License: | GPL (>=2) |
| LazyLoad: | yes |

Several functions are available:

| | |
|---|---|
| allcause.ROCt | This function allows the estimation of usual time-dependent ROC curve. |
| net.ROCt | This function allows the estimation of net time-dependent ROC curve. |
| e.utility | The expected utility theory allows the estimation of optimal cut-of for medical decision making. |

## 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., 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, 33, 2379-89.

Tessier P et al. Manuscript submitted.

## 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, estimator, lambda)
```

## 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 values of the variable used to define the possible binary tests. |
| `estimator` | Three possible estimators can be used: 'kaplan-meier', 'akritas' or 'naive'. The naive estimator is selected by default. |
| `lambda` | This is the proportion of the nearest neighbors. The estimation will be based on 2*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 (`estimator='akritas'`). |

## Details

This function computes time-dependent ROC curve with right-censoring data. It can use Akritas approach (nearest neighbor's estimation) for ensuring monotone increasing ROC curve, instead of the simple Kaplan-Meier estimator. This Akritas approach may be avoid if the sample size is large because of computing time. Both estimators were defined by Heagerty, Lumley and Pepe (Biometrics, 2000). A third alternative is the use of the naive estimator as explained by Blanche, Dartigues and Jacqmin-Gadda (2013). This estimator is less time-consuming compared to the Akritas approach.

## 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.

Uno H, Cai T, Tian L, JJ W. (2007) Evaluating prediction rules for t-year survivors with censored regression models. Journal of the American Statistical Association, 102, 527-37.

Blanche P, Dartigues J, Jacqmin-Gadda H. (2013) Review and comparison of roc curve estimators for a time-dependent outcome with marker-dependent censoring. Biometrical Journal, 55, 687-704.

## Examples

```
# import and attach the data example

data(dataDIVAT)

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

dataDIVAT <- dataDIVAT[1:400,]

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

age.cut <- quantile(dataDIVAT$ageR, probs=seq(0.1, 0.9, by=0.1))

# the ROC curve (with the naive estimator) 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, estimator="naive")

# the sensibilities and specificities associated with the cut off values

roc1$table

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

# the 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", lty=1, col=2, lwd=2)

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

legend("bottomright", paste("Naive, (AUC=", round(roc1$auc, 2), ")", sep=""),
lty=1, lwd=2, col=2)
```

dataDIVAT *A sample of the DIVAT cohort in orter to study the mortality of kidney transplant recipients.*

## 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 (raws) 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 in days (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)
```

dataKTFS *A sample of the DIVAT cohort in orter to study the prognostic capacities of the Kidney Transplant Failure Score (KTFS).*

## Description

A data frame with 2169 French kidney transplant recipients. The time-to-event is the time between the transplantation and the return in dialysis. This time can be right-censored, especially at the time of the patient death with a functionning graft. The KTFS is a score proposed by Foucher et al. (2010) to stratify the recipients according to their risk of return in dialysis.

## Usage

```
data(dataKTFS)
```

## Format

A data frame with 2169 observations (raws) with the 3 following variables (columns).

`time` This numeric vector represents the follow up time in years (until return in dialyis or censoring)

`failure` This numeric vector represents the graft failure indicator at the follow-up end (1=return, 0=censoring)

`score` This numeric vector represents the KTFS values.

## 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 the training sample of 2169 patients used for the construction of the KTFS, a scoring system proposed by Foucher et al. (2010) in order to stratify patients according to their risk of return in dialysis. The KTFS is based on 8 parameters collected during the first year post transplantation. Patients with a KTFS value higher than 4.17 were considered at high-risk.

## Source

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

## References

Foucher Y., Daguin P, Akl A et al. (2010) A clinical scoring system highly predictive of long-term kidney graft survival. Kidney International, 78, 1288-94.

## Examples

```
data(dataKTFS)

### a short summary of the recipient age at the transplantation
summary(dataKTFS$score)

### Kaplan and Meier estimation of the recipient survival
plot(survfit(Surv(time, failure) ~ I(score>4.17), data = dataKTFS),
```

```
 xlab="Post transplantation time (in years)", ylab="Patient survival",
 mark.time=FALSE, col=c(2,1), lty=c(2,1))
legend("bottomleft", c("Recipients in the high-risk group",
 "Recipients in the low-risk group"), col=1:2, lty=1:2)
```

---

EUt *Optimal cut-off estimation of a prognostic variable*

---

## Description

This function allows the estimation of an optimal cut-off estimation for medical decison making from a prognostic marker by maximizing the expected utility.

## Usage

```
EUt(times, failures, variable, pro.time, cut.off, estimator, lambda = NULL,
 u.tn, u.fp, u.tp0, u.tp1, u.fn0, u.fn1, full.efficacy = FALSE, gain = 0)
```

## 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 values of the variable used to define the possible binary tests. |
| `estimator` | Three possible estimators can be used: 'kaplan-meier', 'akritas' or 'naive'. The naive estimator is selected by default. |
| `lambda` | This is the proportion of the nearest neighbours. The estimation will be based on 2*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 neighbour's estimation (`estimator='akritas'`). |
| `u.tn` | The value of the utility related to a patient with a true negative test. This value should respect the 0-1 scale death-perfect health. |
| `u.fp` | The value of the utility related to a patient with a false positive test. This value should respect the 0-1 scale death-perfect health. |
| `u.tp0` | The value of the utility related to a patient with a true positive test before the failure occurs. This value should respect the 0-1 scale death-perfect health. |
| `u.tp1` | The value of the utility related to a patient with a true positive test after the failure occurs. This value should respect the 0-1 scale death-perfect health. |
| `u.fn0` | The value of the utility related to a patient with a false negative test before the failure occurs. This value should respect the 0-1 scale death-perfect health. |
| `u.fn1` | The value of the utility related to a patient with a false negative test after the failure occurs. This value should respect the 0-1 scale death-perfect health. |

`full.efficacy` A logical value indicating whether the full efficacy of the novel therapy should be consider. The full efficacy represents the situation for which all the failures will be prevent before the considered prognostic time by using the novel therapy. The default value is FALSE.

`gain` The expected percentage in the increase of the mean time-to-failure due to the novel therapy. This value will be consider only if `full.efficacy=FALSE`. The default value is 0. This value is in-between 0 and 100.

## Details

This function computes the expected utility function for the cut-off defined by the user. The individual data may be right-censored. It can use Akritas approach (nearest neighbour's estimation) for ensuring monotone property of the corresponding function, instead of the simple Kaplan-Meier estimator. This Akritas approach may be avoid if the sample size is large because of computing time. Both estimators were defined by Heagerty, Lumley and Pepe (Biometrics, 2000). A third alternative is the use of the naive estimator as explained by Blanche, Dartigues and Jacqmin-Gadda (2013). This estimator is less time-consuming compared to the Akritas approach. The user have to define the utilities/qualities of life related to the possible health-states of the patients after the prognostic marker-based decision. A positive test is defined by a marker value higher than the cut-off: the patient is considered at risk of the failure before the prognostic time. The situation consists in the prescription of a novel therapy to avoid/delay the failure in high-risk patients, but with a decrease in the quality of life versus non-treated patients. For true positive and false negative test, we distinguish the quality of life before and after the failure, the quality of life being higher before than after.

## Value

`missing` Number of deleted observations due to missing data.

`proba.table` This data frame is composed by 4 columns representing respectively the probabilities of true positive test (tp), false positive test (fp), true negative test (tn) and false negative test (fp). Each line corresponds to the cut-off values entered by the user.

`qaly.table` This data frame is composed by 4 columns representing respectively the QALYs of true positive test (tp), false positive test (fp), true negative test (tn) and false negative test (fp). Each line corresponds to the cut-off values entered by the user.

`eu.table` This data frame returns the expected utilities associated with the cut-off values entered by the user.

`eu.all` The expected utility if all patients are treated (the cut-off is the minimum of the prognostic variable).

`eu.no` The expected utility if no patient is treated (the cut-off is the maximum of the prognostic variable).

`eu.maximum` The maximum expected utility associated with the cut-off values entered by the user or the extreme choice when all patients are treated (minimal cut-off) or no patient is treated (maximal cut-off).

`estimation` The cut-off that maximizes the expected utility of the medical decision. If this value corresponds to the minimum or the maximum of the prognostic variable, all the patients or no patient should respectively be treated.

## Author(s)

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

## References

Foucher et al. (2010) A clinical scoring system highly predictive of long-term kidney graft survival. Kidney International, 78:1288-94.

Tessier P et al. Manuscript submitted.

## Examples

```
# import and attach the data example

data(dataKTFS)

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

ktfs.cut <- quantile(dataKTFS$score, probs=seq(0.1, 0.9, by=0.1), na.rm=TRUE)

# the corresponding expected utility function for a prognostic
# up to 8 years post-transplantation

estim.naive <- EUt(dataKTFS$time, dataKTFS$failure, dataKTFS$score,
 pro.time=8, cut.off=ktfs.cut, estimator="naive", lambda=NULL, u.tn=0.81,
 u.fp=0.81*0.99, u.tp0=0.81*0.99, u.tp1=0.53, u.fn0=0.81, u.fn1=0.53,
 full.efficacy=FALSE, gain=0.5)

plot(estim.naive$eu.table$cut.off, estim.naive$eu.table$utility, type="l",
 xlab="Cut-off values", ylab="Expected utility", col=1, lty=1)
```

---

# 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 values of the variable used to define the possible binary tests. |
| 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*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-Perm 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 equal1 to 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, 33, 2379-89.

## 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(0.1, 0.9, by=0.1))

# 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 (without correction by the knn estimator) to
# reduce the time for comuting this exemple. In prectice, the
# correction should by used in case of non-montone results.

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=FALSE,
 lambda=NULL)

# the sensibilities and specificities associated with the cut off values

roc1$table

# 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",
 lty=1, col=2, lwd=2)

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

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

---

`rein.ratetable` *Expected mortality of French patients with End Stage Kidney Disease (ESKD) in dialysis and registrited previously on waiting list for renal transplantation*

---

## 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