

Package ‘Multistate’

March 3, 2015

Type Package

Title Multistate models (Markov and semi-Markov) for modelling
evolution of chronic diseases with extensions such as the additive relative survival approach.

Version 0.1

Date 2015-03-03

Author Yohann Foucher, Florence Gillaizeau

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

Description Medical researchers are often interested in investigating the relationship between explicative variables and times-to-events such as disease progression or death. Multistate models allows multiple times-to-event to be studied simultaneously. Time-inhomogeneous Markov models consist of modelling the probabilities of transitions according to the chronological times (times since the baseline of the study). Semi-Markov (SM) models consist of modelling the probabilities of transitions according to the times spent in states. These SM models are becoming increasingly popular to deal with the complex evolution of chronic diseases. In this package, we propose functions implementing usual 3-state and 4-state multistate models (SM models and time-inhomogeneous Markov models). We also propose to take into account the mortality of the general population (relative survival approach). The methods proposed are limited to parametric multistate models.

License GPL (>=2)

LazyLoad yes

Depends survival, statmod, date

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

R topics documented:

Multistate-package	2
dataDIVAT	3
fr.ratetable	4
m3	5
m4	9
mrs3	13
mrs4	16
sm3	21

sm4	24
smrs3	28
smrs4	32
usa.ratetable	37

Index	39
--------------	-----------

Multistate-package	<i>Multistate models</i>
--------------------	--------------------------

Description

Medical researchers are often interested in investigating the relationship between explicative variables and times-to-events such as disease progression or death. Multistate models allows multiple times-to-event to be studied simultaneously. Time-inhomogeneous Markov models consist of modelling the probabilities of transitions according to the chronological times (times since the baseline of the study). Semi-Markov (SM) models consist of modelling the probabilities of transitions according to the times spent in states. These SM models are becoming increasingly popular to deal with the complex evolution of chronic diseases. In this package, we propose functions implementing usual 3-state and 4-state multistate models (SM models and time-inhomogeneous Markov models). We also propose to take into account the mortality of the general population (relative survival approach).

Additive relative survival The excess mortality related to a disease is estimated by subtracting the expected mortality rates from general population life tables to the observed mortality rates.

The methods proposed are limited to parametric models. The 3-state (illness-death) models model include one initial state (X=1), one transient state (X=2), and an absorbing state (X=3). The possible transitions are: 1->2, 1->3 and 2->3. The 4-state models include one initial state (X=1), one transient state (X=2), and two competing and absorbing states (X=3 and X=4). The possible transitions are: 1->2, 1->3, 1->4, 2->3 and 2->4.

Details

Package: Multistate
Type: Package
Version: 0.2
Date: 2015-03-03
License: GPL (>=2)
LazyLoad: yes

- m3 computes a 3-state time-inhomogeneous Markov model
- m4 computes a 4-state time-inhomogeneous Markov model
- sm3 computes a 3-state semi-Markov model
- sm4 computes a 4-state semi-Markov model
- mrs3 computes a 3-state additive relative survival time-inhomogeneous Markov model
- mrs4 computes a 4-state additive relative survival time-inhomogeneous Markov model

smrs3 computes a 3-state additive relative survival semi-Markov model
 smrs4 computes a 4-state additive relative survival semi-Markov model

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr>
 and Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. Manuscript submitted. 2015.

See Also

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

dataDIVAT

A sample of the DIVAT cohort.

Description

A data frame with 5943 French kidney transplant recipients (KTR) from the DIVAT cohort. As described in `eSemiMarkov` package, the KTR evolution can be described according to the 4-state structure. X=2 represents the acute rejection episode, X=3 represents the definitive return to dialysis and X=4 represents the death. These times can be right-censored. A vector of covariates is also collected at the transplantation.

Usage

```
data(dataDIVAT)
```

Format

A data frame with 5943 observations (rows) with the 7 following variables (columns):

trajectory	this numeric vector represents the sequences of states of the recipient.
time1	this numeric vector represents the time-to-first event in days, i.e. the time between the transplantation and the first clinical event (acute rejection episode, return to dialysis, or recipient death with a functioning graft), or the time-to-censoring if applicable (X=1 at the last follow-up).
time2	this numeric vector represents the time-to-second event in days, i.e. the time between the transplantation and the second clinical event (return to dialysis or recipient death with a functioning graft), or the the time-to-censoring if applicable (X=2 at the last follow-up).
ageR	this numeric vector represents the age of the recipient at the transplantation (in years).
sexR	this character vector represents the gender of the recipient (male or female).
year.tx	this numeric vector represents the calendar year of the transplantation.

- z this numeric vector represents the explicative variable under interest, i.e. the delayed graft function (1=yes, 0=no).

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. Here is a sample of 5943 patients from this DIVAT cohort. A vector of covariates, all measured at the transplantation, is collected for each patient.

Source

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

Examples

```
data(dataDIVAT)

### a description of transitions
table(dataDIVAT$trajectory)

### patient-graft survival (first event between the return to dialysis and the patient
### death with a functioning graft)

dataDIVAT$failure<-1*(dataDIVAT$trajectory!=1 & dataDIVAT$trajectory!=12)

dataDIVAT$time_to_failure<-NA
dataDIVAT$time_to_failure<-ifelse(dataDIVAT$trajectory %in% c(1,12,13,14),
dataDIVAT$time1,dataDIVAT$time1+dataDIVAT$time2)

plot(survfit(Surv(time_to_failure/365.24, failure) ~ 1 , data=dataDIVAT), mark.time=FALSE,
xlim=c(0,12), ylim=c(0,1), cex=1.5, col=1, lwd=2, lty=1,
xlab="Times after the transplantation (years)",
ylab="Patient-graft survival")
```

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 an 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 vector 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 equals to 1 if the corresponding dimension does not vary according to the time. Only the dimension related to sex is associated to 1.
cutpoints	a list of the thresholds to identify the changes in the mortality rates according to the time-dependent 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. Pohar, 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)
```

m3

Computes a 3-state parametric Markov model by likelihood maximization

Description

The 3-state Markov model includes an initial state (X=1), a transient state (X=2) and an absorbing state (X=3). Usually, X=1 corresponds to disease-free or remission, X=2 to relapse, and X=3 to death. In this illness-death model, the possible transitions are: 1->2, 1->3 and 2->3.

Usage

```
m3(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
ini.base.12=NULL, ini.base.13=NULL, ini.base.23=NULL, cov.12=NULL,
init.cov.12=NULL, names.12=NULL, cov.13=NULL, init.cov.13=NULL,
names.13=NULL, cov.23=NULL, init.cov.23=NULL, names.23=NULL,
conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)
```

Arguments

t1	a numeric vector with the observed times in days from baseline to the first transition (X=2 or X=3) or to the right-censoring in X=1.
t2	a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in X=2. NA for individuals right-censored in X=1 or individuals who directly transited from X=1 to X=3.
sequence	a numeric vector with the sequence of observed states. Four possible values are allowed: 1 (individual right-censored in X=1), 12 (individual right-censored in X=2), 13 (individual who directly transited from X=1 to X=3), 123 (individual who transited from X=1 to X=3 through X=2).
dist	a character vector with three arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution.
cuts.12	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=2. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.13	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.23	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
ini.base.12	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=2. Default initial value is 1 for all parameters. See details for more details.
ini.base.13	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=3. Default initial value is 1 for all parameters. See details for more details.
ini.base.23	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=3. Default initial value is 1 for all parameters. See details for more details.
cov.12	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=2.
init.cov.12	a numeric vector of initial values for regression coefficients associated to cov.12. Default initial value is 0 for all parameters.
names.12	a character vector with name of explicative variables associated to cov.12 (optional).
cov.13	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
init.cov.13	a numeric vector of initial values for regression coefficients associated to cov.13. Default initial value is 0 for all parameters.

<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.23</code>	a numeric matrix (or <code>data.frame</code>) with the explicative time-fixed variable(s) related to the time from $X=2$ to $X=3$.
<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.
<code>names.23</code>	a character vector with name of explicative variables associated to <code>cov.23</code> (optional).
<code>conf.int</code>	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is TRUE.
<code>silent</code>	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is TRUE, which corresponds to silent mode (no display).
<code>precision</code>	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .
<code>maxiter</code>	a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

$$\begin{array}{ll}
 \text{Exponential distribution} & \lambda(t) = 1/\sigma \\
 \text{Weibull distribution} & \lambda(t) = \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \\
 \text{Generalized Weibull distribution} & \lambda(t) = \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1}
 \end{array}$$

with σ, ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in `ini.base.12`, `ini.base.13` and `ini.base.23`. Moreover, the estimated values also correspond to the logarithms. Maximization of the log-likelihood function is performed using the algorithms of Nelder and Mead (`optim` function).

Value

<code>Table</code>	a <code>data.frame</code> containing the parameter estimations of the model (<code>Estimate</code>). When the option <code>conf.int=TRUE</code> is specified, this <code>data.frame</code> includes three additional columns: the Standard Errors of parameters (<code>Std.Error</code>), the value of the Wald statistic (<code>t.value</code>), and the p-value for the Wald test (<code>Pr(> t)</code>).
<code>CovarianceMatrix</code>	a <code>data.frame</code> corresponding to variance-covariance matrix of the estimated model.
<code>LogLik</code>	a numeric value corresponding to the log-likelihood of the estimated model.
<code>AIC</code>	a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
 Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Huszti E, Abrahamowicz M, Alioum A, Binquet C, Quantin C. Relative survival multistate Markov model. *Stat Med.* 2012 Feb 10;31(3):269-86.

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
dataDIVATsample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
replace = FALSE),]

# To illustrate the use of a 3-state model, individuals with trajectory 13 and 123 are
# censored at the time of transition into state X=3

dataDIVAT_3state<-dataDIVATsample
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==13]<-1
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==123]<-12
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==14]<-13
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==124]<-123

# 3-state parametric Markov model including one explicative variable ('z')
# on all transitions

m3(t1=dataDIVAT_3state$time1,
    t2=dataDIVAT_3state$time2,
    sequence=dataDIVAT_3state$trajectory,
    dist=c("W", "W", "W"),
    cuts.12=NULL,
    cuts.13=NULL,
    cuts.23=NULL,
    ini.base.12=c(9.93,-1.00),
    ini.base.13=c(11.54,-0.20),
    ini.base.23=c(10.21,0.21),
    cov.12=dataDIVAT_3state$z,
    init.cov.12=c(-0.13),
    names.12=c("beta12_z") ,
    cov.13=dataDIVAT_3state$z,
    init.cov.13=c(1.61),
    names.13=c("beta13_z") ,
    cov.23=dataDIVAT_3state$z,
    init.cov.23=c(1.39),
    names.23=c("beta23_z"),
    conf.int=TRUE,
    silent=FALSE,
    precision=0.001,
    maxiter=1000)
```

m4	<i>Computes a 4-state parametric Markov model by likelihood maximization</i>
----	--

Description

The 4-state Markov model includes an initial state (X=1), a transient state (X=2) and two absorbing states (X=3 and X=4). Usually, X=1 corresponds to disease-free or remission and X=4 to death. The possible transitions are: 1->2, 1->3, 1->4, 2->3 and 2->4.

Usage

```
m4(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.14=NULL, cuts.23=NULL,
cuts.24=NULL, ini.base.12=NULL, ini.base.13=NULL, ini.base.14=NULL, ini.base.23=NULL,
ini.base.24=NULL, cov.12=NULL, init.cov.12=NULL, names.12=NULL, cov.13=NULL,
init.cov.13=NULL, names.13=NULL, cov.14=NULL, init.cov.14=NULL, names.14=NULL,
cov.23=NULL, init.cov.23=NULL, names.23=NULL, cov.24=NULL, init.cov.24=NULL,
names.24=NULL, conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)
```

Arguments

t1	a numeric vector with the observed times in days from baseline to the first transition (X=2 or direct transition to X=3 or X=4) or to the right-censoring in X=1.
t2	a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in X=2. NA for individuals right-censored in X=1 or individuals who directly transitioned from X=1 to X=3 or to X=1 to X=4.
sequence	a numeric vector with the sequence of observed states. Six possible values are allowed: 1 (individual right-censored in X=1), 12 (individual right-censored in X=2), 13 (individual who directly transitioned from X=1 to X=3), 14 (individual who directly transitioned from X=1 to X=4), 123 (individual who transitioned from X=1 to X=3 through X=2), 124 (individual who transitioned from X=1 to X=4 through X=2).
dist	a character vector with five arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3 and 2->4. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution.
cuts.12	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=2. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.13	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.14	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=4. Only internal timepoints are

	allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>cuts.23</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>cuts.24</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=4. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>ini.base.12</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=2. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.13</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.14</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=4. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.23</code>	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.24</code>	a numeric vector of initial values for parameters of distribution from X=2 to X=4. Default initial value is 1 for all parameters. See details for more details.
<code>cov.12</code>	a numeric vector (matrix or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=2.
<code>init.cov.12</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.12</code> . Default initial value is 0 for all parameters.
<code>names.12</code>	a character vector with name of explicative variables associated to <code>cov.12</code> (optional).
<code>cov.13</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.13</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.13</code> . Default initial value is 0 for all parameters.
<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.14</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.14</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.14</code> . Default initial value is 0 for all parameters.
<code>names.14</code>	a character vector with name of explicative variables associated to <code>cov.14</code> (optional).
<code>cov.23</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=3.

<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.
<code>names.23</code>	a character vector with name of explicative variables associated to <code>cov.23</code> (optional).
<code>cov.24</code>	a numeric matrix (or <code>data.frame</code>) with the explicative time-fixed variable(s) related to the time from <code>X=2</code> to <code>X=3</code> .
<code>init.cov.24</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.24</code> . Default initial value is 0 for all parameters.
<code>names.24</code>	a character vector with name of explicative variables associated to <code>cov.24</code> (optional).
<code>conf.int</code>	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is <code>TRUE</code> .
<code>silent</code>	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is <code>TRUE</code> , which corresponds to silent mode (no display).
<code>precision</code>	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .
<code>maxiter</code>	a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

$$\begin{aligned}
 \text{Exponential distribution} \quad & \lambda(t) = 1/\sigma \\
 \text{Weibull distribution} \quad & \lambda(t) = \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \\
 \text{Generalized Weibull distribution} \quad & \lambda(t) = \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1}
 \end{aligned}$$

with σ, ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in `ini.base.12`, `ini.base.13`, `ini.base.14`, `ini.base.23` and `ini.base.24`. Moreover, the estimated values also correspond to the logarithms. Maximization of the log-likelihood function is performed using the algorithms of Nelder and Mead (`optim` function).

Value

<code>Table</code>	a <code>data.frame</code> containing the parameter estimations of the model (<code>Estimate</code>). When the option <code>conf.int=TRUE</code> is specified, this <code>data.frame</code> includes three additional columns: the Standard Errors of parameters (<code>Std.Error</code>), the value of the Wald statistic (<code>t.value</code>), and the p-value for the Wald test (<code>Pr(> t)</code>).
<code>CovarianceMatrix</code>	a <code>data.frame</code> corresponding to variance-covariance matrix of the estimated model.
<code>LogLik</code>	a numeric value corresponding to the log-likelihood of the estimated model.
<code>AIC</code>	a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
 Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Husztı E, Abrahamowicz M, Alioum A, Biquet C, Quantin C. Relative survival multistate Markov model. Stat Med. 2012 Feb 10;31(3):269-86.

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
dataDIVATsample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
  replace = FALSE),]

# 4-state parametric Markov model including one explicative variable ('z')
# on all transitions

m4(t1=dataDIVATsample$time1,
  t2=dataDIVATsample$time2,
  sequence=dataDIVATsample$trajectory,
  dist=c("E", "E", "E", "E", "E"),
  cuts.12=NULL,
  cuts.13=NULL,
  cuts.14=NULL,
  cuts.23=NULL,
  cuts.24=NULL,
  ini.base.12=c(8.31),
  ini.base.13=c(10.46),
  ini.base.14=c(10.83),
  ini.base.23=c(9.01),
  ini.base.24=c(10.81),
  cov.12=dataDIVATsample$z,
  init.cov.12=c(-0.02),
  names.12=c("beta12_z") ,
  cov.13=dataDIVATsample$z,
  init.cov.13=c(1.42),
  names.13=c("beta13_z") ,
  cov.14=dataDIVATsample$z,
  init.cov.14=c(1.64),
  names.14=c("beta14_z") ,
  cov.23=dataDIVATsample$z,
  init.cov.23=c(0.30),
  names.23=c("beta23_z") ,
  cov.24=dataDIVATsample$z,
  init.cov.24=c(1.40),
  names.24=c("beta24_z"),
  conf.int=TRUE,
  silent=FALSE,
  precision=0.001,
  maxiter=1000)
```

mrs3

Computes a 3-state parametric additive relative survival Markov model by likelihood maximization

Description

The 3-state Markov Relative Survival model includes an initial state ($X=1$), a transient state ($X=2$), and the death ($X = 3$ for observed death regardless of cause). Usually, $X=1$ corresponds to disease-free or remission and $X=2$ to relapse. The possible transitions are: $1 \rightarrow 2$, $1 \rightarrow 3$ and $2 \rightarrow 3$. In additive relative survival analysis, the observed mortality hazard ($X=3$) is assumed to be the sum of two components: the expected population mortality ($X=P$) and the excess mortality related to the disease under study ($X=E$). The expected population mortality hazard ($X=P$) can be obtained from the death rates provided by life tables of statistical national institutes. These tables indicate the proportion of people dead in a calendar year stratified by birthdate and gender. The instantaneous hazard functions for times to death related to the disease ($X=E$) are modelled here using parametric distributions.

Usage

```
mrs3(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
ini.base.12=NULL, ini.base.13=NULL, ini.base.23=NULL, cov.12=NULL,
init.cov.12=NULL, names.12=NULL, cov.13=NULL, init.cov.13=NULL, names.13=NULL,
cov.23=NULL, init.cov.23=NULL, names.23=NULL, p.age, p.sex, p.year, p.rate.table,
conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)
```

Arguments

t1	a numeric vector with the observed times in days from baseline to the first transition ($X=2$ or $X=3$) or to the right-censoring in $X=1$.
t2	a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in $X=2$. NA for individuals right-censored in $X=1$ or individuals who directly transitioned from $X=1$ to $X=3$.
sequence	a numeric vector with the sequence of observed states. Four possible values are allowed: 1 (individual right-censored in $X=1$), 12 (individual right-censored in $X=2$), 13 (individual who directly transitioned from $X=1$ to $X=3$), 123 (individual who transitioned from $X=1$ to $X=3$ through $X=2$).
dist	a character vector with three arguments describing respectively the distributions of duration time for transitions $1 \rightarrow 2$, $1 \rightarrow 3$ and $2 \rightarrow 3$. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution.
cuts.12	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=1$ to $X=2$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.13	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=1$ to $X=3$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.

<code>cuts.23</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=2$ to $X=3$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>ini.base.12</code>	a numeric vector of initial logarithm values for parameters of distribution from $X=1$ to $X=2$. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.13</code>	a numeric vector of initial logarithm values for parameters of distribution from $X=1$ to $X=3$. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.23</code>	a numeric vector of initial logarithm values for parameters of distribution from $X=2$ to $X=3$. Default initial value is 1 for all parameters. See details for more details.
<code>cov.12</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from $X=1$ to $X=2$.
<code>init.cov.12</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.12</code> . Default initial value is 0 for all parameters.
<code>names.12</code>	a character vector with name of explicative variables associated to <code>cov.12</code> (optional).
<code>cov.13</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from $X=1$ to $X=3$.
<code>init.cov.13</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.13</code> . Default initial value is 0 for all parameters.
<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.23</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from $X=2$ to $X=3$.
<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.
<code>names.23</code>	a character vector with name of explicative variables associated to <code>cov.23</code> (optional).
<code>p.age</code>	a numeric vector with the ages in days at baseline ($X=1$).
<code>p.sex</code>	a character vector with the gender: male or female.
<code>p.year</code>	a numeric vector with entry date in the study in the date format, i.e. in number of days since 01.01.1960.
<code>p.rate.table</code>	a table of event rates, organized as a <code>ratetable</code> object.
<code>conf.int</code>	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is TRUE.
<code>silent</code>	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is TRUE, which corresponds to silent mode (no display).
<code>precision</code>	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .
<code>maxiter</code>	a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

$$\begin{array}{ll}
 \text{Exponential distribution} & \lambda(t) = 1/\sigma \\
 \text{Weibull distribution} & \lambda(t) = \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \\
 \text{Generalized Weibull distribution} & \lambda(t) = \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1}
 \end{array}$$

with σ, ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in `ini.base.12`, `ini.base.13` and `ini.base.23`. Moreover, the estimated values also correspond to the logarithms. Maximization of the log-likelihood function is performed using the algorithms of Nelder and Mead (`optim` function). Works only with ratetables that are split by age, sex and year.

Value

Table	a data.frame containing the parameter estimations of the model (Estimate). When the option <code>conf.int=TRUE</code> is specified, this data.frame includes three additional columns: the Standard Errors of parameters (<code>Std.Error</code>), the value of the Wald statistic (<code>t.value</code>), and the p-value for the Wald test (<code>Pr(> t)</code>).
CovarianceMatrix	a data.frame corresponding to variance-covariance matrix of the estimated model.
LogLik	a numeric value corresponding to the log-likelihood of the estimated model.
AIC	a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
 Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Husztai E, Abrahamowicz M, Alioum A, Binquet C, Quantin C. Relative survival multistate Markov model. *Stat Med*. 2012 Feb 10;31(3):269-86.
 Pohar M., Stare J. (2006) "Relative survival analysis in R." *Computer Methods and Programs in Biomedicine*, 81: 272-278

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
dataDIVAT$sample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
  replace = FALSE),]

# To use a 3-state model, individuals with trajectory 13 and 123 are censored at the time
```

```

# of transition into state X=3

dataDIVAT_3state<-dataDIVATsample
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==13]<-1
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==123]<-12
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==14]<-13
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==124]<-123

# import the expected mortality rates

data(fr.ratetable)

# 3-state parametric additive relative survival Markov model including one
# explicative variable ('z') on all transitions

mrs3(t1=dataDIVAT_3state$time1,
     t2=dataDIVAT_3state$time2,
     sequence=dataDIVAT_3state$trajectory,
     dist=c("E", "E", "E"),
     cuts.12=NULL,
     cuts.13=NULL,
     cuts.23=NULL,
     ini.base.12=c(8.34),
     ini.base.13=c(10.70),
     ini.base.23=c(11.10),
     cov.12=dataDIVAT_3state$z,
     init.cov.12=c(0.04),
     names.12=c("beta12_z") ,
     cov.13=dataDIVAT_3state$z,
     init.cov.13=c(1.04),
     names.13=c("beta1E_z") ,
     cov.23=dataDIVAT_3state$z,
     init.cov.23=c(0.29),
     names.23=c("beta2E_z"),
     p.age=dataDIVAT_3state$ageR*365.24,
     p.sex=dataDIVAT_3state$sexR,
     p.year=as.date(paste("01", "01", dataDIVAT_3state$year.tx), order = "mdy"),
     p.rate.table=fr.ratetable,
     conf.int=TRUE,
     silent=FALSE,
     precision=0.001,
     maxiter=1000)

```

mrs4

Computes a 4-state parametric additive relative survival Markov model by likelihood maximization

Description

The 4-state Markov Relative Survival model includes an initial state ($X=1$), a transient state ($X=2$) and two absorbing states including death ($X=3$, and $X=4$ for observed death regardless of cause). The possible transitions are: 1->2, 1->3, 1->4, 2->3 and 2->4. In additive relative survival analysis,

the observed mortality hazard ($X=4$) is assumed to be the sum of two components: the expected population mortality ($X=P$) and the excess mortality related to the disease under study ($X=E$). The expected population mortality hazard ($X=P$) can be obtained from the death rates provided by life tables of statistical national institutes. These tables indicate the proportion of people dead in a calendar year stratified by birthdate and gender. The instantaneous hazard functions for times to death related to the disease ($X=E$) are modelled here using parametric distributions.

Usage

```
mrs4(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.14=NULL, cuts.23=NULL,
cuts.24=NULL, ini.base.12=NULL, ini.base.13=NULL, ini.base.14=NULL, ini.base.23=NULL,
ini.base.24=NULL, cov.12=NULL, init.cov.12=NULL, names.12=NULL, cov.13=NULL,
init.cov.13=NULL, names.13=NULL, cov.14=NULL, init.cov.14=NULL, names.14=NULL,
cov.23=NULL, init.cov.23=NULL, names.23=NULL, cov.24=NULL, init.cov.24=NULL,
names.24=NULL, p.age, p.sex, p.year, p.rate.table,
conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)
```

Arguments

t1	a numeric vector with the observed times in days from baseline to the first transition ($X=2$ or direct transition to $X=3$ or $X=4$) or to the right-censoring in $X=1$.
t2	a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in $X=2$. NA for individuals right-censored in $X=1$ or individuals who directly transited from $X=1$ to $X=3$ or to $X=1$ to $X=4$.
sequence	a numeric vector with the sequence of observed states. Six possible values are allowed: 1 (individual right-censored in $X=1$), 12 (individual right-censored in $X=2$), 13 (individual who directly transited from $X=1$ to $X=3$), 14 (individual who directly transited from $X=1$ to $X=4$), 123 (individual who transited from $X=1$ to $X=3$ through $X=2$), 124 (individual who transited from $X=1$ to $X=4$ through $X=2$).
dist	a character vector with five arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3 and 2->4. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution.
cuts.12	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=1$ to $X=2$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.13	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=1$ to $X=3$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.14	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=1$ to $X=4$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.

<code>cuts.23</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>cuts.24</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=4. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>ini.base.12</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=2. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.13</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.14</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=4. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.23</code>	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.24</code>	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=4. Default initial value is 1 for all parameters. See details for more details.
<code>cov.12</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=2.
<code>init.cov.12</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.12</code> . Default initial value is 0 for all parameters.
<code>names.12</code>	a character vector with name of explicative variables associated to <code>cov.12</code> (optional).
<code>cov.13</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.13</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.13</code> . Default initial value is 0 for all parameters.
<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.14</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.14</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.14</code> . Default initial value is 0 for all parameters.
<code>names.14</code>	a character vector with name of explicative variables associated to <code>cov.14</code> (optional).
<code>cov.23</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=3.
<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.

names.23	a character vector with name of explicative variables associated to cov.23 (optional).
cov.24	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=3.
init.cov.24	a numeric vector of initial values for regression coefficients associated to cov.24. Default initial value is 0 for all parameters.
names.24	a character vector with name of explicative variables associated to cov.24 (optional).
p.age	a numeric vector with the ages in days at baseline (X=1).
p.sex	a character vector with the gender: male or female.
p.year	a numeric vector with entry date in the study in the date format, i.e. in number of days since 01.01.1960.
p.rate.table	a table of event rates, organized as a ratetable object.
conf.int	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is TRUE.
silent	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is TRUE, which corresponds to silent mode (no display).
precision	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .
maxiter	a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

$$\begin{aligned}
 \text{Exponential distribution} \quad & \lambda(t) = 1/\sigma \\
 \text{Weibull distribution} \quad & \lambda(t) = \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \\
 \text{Generalized Weibull distribution} \quad & \lambda(t) = \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1}
 \end{aligned}$$

with σ, ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in ini.base.12, ini.base.13, ini.base.14, ini.base.23 and ini.base.24. Moreover, the estimated values also correspond to the logarithms. Maximization of the log-likelihood function is performed using the algorithms of Nelder and Mead (optim function). Works only with ratetables that are split by age, sex and year.

Value

Table	a data.frame containing the parameter estimations of the model (Estimate). When the option conf.int=TRUE is specified, this data.frame includes three additional columns: the Standard Errors of parameters (Std.Error), the value of the Wald statistic (t.value), and the p-value for the Wald test (Pr(> t)).
CovarianceMatrix	a data.frame corresponding to variance-covariance matrix of the estimated model.
LogLik	a numeric value corresponding to the log-likelihood of the estimated model.
AIC	a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
 Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Husztai E, Abrahamowicz M, Alioum A, Binquet C, Quantin C. Relative survival multistate Markov model. Stat Med. 2012 Feb 10;31(3):269-86.

Pohar M., Stare J. (2006) "Relative survival analysis in R." Computer Methods and Programs in Biomedicine, 81: 272-278

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
dataDIVAT$sample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
  replace = FALSE),]

# import the expected mortality rates

data(fr.ratetable)

# 4-state parametric additive relative survival Markov model including one
# explicative variable ('z') on all transitions

mrs4(t1=dataDIVAT$sample$time1,
     t2=dataDIVAT$sample$time2,
     sequence=dataDIVAT$sample$trajectory,
     dist=c("E", "E", "E", "E", "E"),
     cuts.12=NULL,
     cuts.13=NULL,
     cuts.14=NULL,
     cuts.23=NULL,
     cuts.24=NULL,
     ini.base.12=c(8.34),
     ini.base.13=c(10.44),
     ini.base.14=c(10.70),
     ini.base.23=c(9.43),
     ini.base.24=c(11.11),
     cov.12=dataDIVAT$sample$z,
     init.cov.12=c(0.04),
     names.12=c("beta12_z") ,
     cov.13=dataDIVAT$sample$z,
     init.cov.13=c(1.02),
     names.13=c("beta13_z") ,
     cov.14=dataDIVAT$sample$z,
```

```

init.cov.14=c(1.04),
names.14=c("beta1E_z") ,
cov.23=dataDIVATsample$z,
init.cov.23=c(0.94),
names.23=c("beta23_z") ,
cov.24=dataDIVATsample$z,
init.cov.24=c(0.30),
names.24=c("beta2E_z"),
p.age=dataDIVATsample$ageR*365.24,
p.sex=dataDIVATsample$sexR,
p.year=as.date(paste("01", "01", dataDIVATsample$year.tx), order = "mdy"),
p.rate.table=fr.ratetable,
conf.int=TRUE,
silent=FALSE,
precision=0.001,
maxiter=1000)

```

sm3

Computes a 3-state parametric semi-Markov model by likelihood maximization

Description

The 3-state SM model includes an initial state (X=1), a transient state (X=2) and an absorbing state (X=3). Usually, X=1 corresponds to disease-free or remission, X=2 to relapse, and X=3 to death. In this illness-death model, the possible transitions are: 1->2, 1->3 and 2->3.

Usage

```

sm3(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
ini.base.12=NULL, ini.base.13=NULL, ini.base.23=NULL, cov.12=NULL,
init.cov.12=NULL, names.12=NULL, cov.13=NULL, init.cov.13=NULL,
names.13=NULL, cov.23=NULL, init.cov.23=NULL, names.23=NULL,
conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)

```

Arguments

- | | |
|----------|---|
| t1 | a numeric vector with the observed times in days from baseline to the first transition (X=2 or X=3) or to the right-censoring in X=1. |
| t2 | a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in X=2. NA for individuals right-censored in X=1 or individuals who directly transitioned from X=1 to X=3. |
| sequence | a numeric vector with the sequence of observed states. Four possible values are allowed: 1 (individual right-censored in X=1), 12 (individual right-censored in X=2), 13 (individual who directly transitioned from X=1 to X=3), 123 (individual who transitioned from X=1 to X=3 through X=2). |
| dist | a character vector with three arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution. |

<code>cuts.12</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=2. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>cuts.13</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>cuts.23</code>	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>ini.base.12</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=2. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.13</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.23</code>	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>cov.12</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=2.
<code>init.cov.12</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.12</code> . Default initial value is 0 for all parameters.
<code>names.12</code>	a character vector with name of explicative variables associated to <code>cov.12</code> (optional).
<code>cov.13</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.13</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.13</code> . Default initial value is 0 for all parameters.
<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.23</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=3.
<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.
<code>names.23</code>	a character vector with name of explicative variables associated to <code>cov.23</code> (optional).
<code>conf.int</code>	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is TRUE.
<code>silent</code>	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is TRUE, which corresponds to silent mode (no display).
<code>precision</code>	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .

maxiter a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

$$\begin{aligned} \text{Exponential distribution} \quad \lambda(t) &= 1/\sigma \\ \text{Weibull distribution} \quad \lambda(t) &= \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \\ \text{Generalized Weibull distribution} \quad \lambda(t) &= \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \end{aligned}$$

with σ, ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in `ini.base.12`, `ini.base.13` and `ini.base.23`. Moreover, the estimated values also correspond to the logarithms. Maximization of the log-likelihood function is performed using the algorithms of Nelder and Mead (`optim` function).

Value

Table a data.frame containing the parameter estimations of the model (`Estimate`). When the option `conf.int=TRUE` is specified, this data.frame includes three additional columns: the Standard Errors of parameters (`Std.Error`), the value of the Wald statistic (`t.value`), and the p-value for the Wald test (`Pr(>|t|)`).

CovarianceMatrix a data.frame corresponding to variance-covariance matrix of the estimated model.

LogLik a numeric value corresponding to the log-likelihood of the estimated model.

AIC a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. Manuscript submitted. 2015.

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
dataDIVAT$sample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
```

```

replace = FALSE),]

# To illustrate the use of a 3-state model, individuals with trajectory 13 and 123 are
# censored at the time of transition into state X=3

dataDIVAT_3state<-dataDIVATsample
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==13]<-1
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==123]<-12
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==14]<-13
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==124]<-123

# 3-state parametric semi-Markov model including one explicative variable ('z')
# on all transitions

sm3(t1=dataDIVAT_3state$time1,
    t2=dataDIVAT_3state$time2,
    sequence=dataDIVAT_3state$trajectory,
    dist=c("W", "W", "W"),
    cuts.12=NULL,
    cuts.13=NULL,
    cuts.23=NULL,
    ini.base.12=c(9.93, -1.00),
    ini.base.13=c(11.54, -0.20),
    ini.base.23=c(10.21, 0.21),
    cov.12=dataDIVAT_3state$z,
    init.cov.12=c(-0.13),
    names.12=c("beta12_z") ,
    cov.13=dataDIVAT_3state$z,
    init.cov.13=c(1.61),
    names.13=c("beta13_z") ,
    cov.23=dataDIVAT_3state$z,
    init.cov.23=c(1.39),
    names.23=c("beta23_z"),
    conf.int=TRUE,
    silent=FALSE,
    precision=0.001,
    maxiter=1000)

```

sm4

Computes a 4-state parametric semi-Markov model by likelihood maximization

Description

The 4-state SM model includes an initial state (X=1), a transient state (X=2) and two absorbing states (X=3 and X=4). Usually, X=1 corresponds to disease-free or remission and X=4 to death. The possible transitions are: 1->2, 1->3, 1->4, 2->3 and 2->4.

Usage

```

sm4(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.14=NULL, cuts.23=NULL,
    cuts.24=NULL, ini.base.12=NULL, ini.base.13=NULL, ini.base.14=NULL, ini.base.23=NULL,
    ini.base.24=NULL, cov.12=NULL, init.cov.12=NULL, names.12=NULL, cov.13=NULL,

```

```
init.cov.13=NULL, names.13=NULL, cov.14=NULL, init.cov.14=NULL, names.14=NULL,
cov.23=NULL, init.cov.23=NULL, names.23=NULL, cov.24=NULL, init.cov.24=NULL,
names.24=NULL, conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)
```

Arguments

t1	a numeric vector with the observed times in days from baseline to the first transition (X=2 or direct transition to X=3 or X=4) or to the right-censoring in X=1.
t2	a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in X=2. NA for individuals right-censored in X=1 or individuals who directly transitioned from X=1 to X=3 or to X=1 to X=4.
sequence	a numeric vector with the sequence of observed states. Six possible values are allowed: 1 (individual right-censored in X=1), 12 (individual right-censored in X=2), 13 (individual who directly transitioned from X=1 to X=3), 14 (individual who directly transitioned from X=1 to X=4), 123 (individual who transitioned from X=1 to X=3 through X=2), 124 (individual who transitioned from X=1 to X=4 through X=2).
dist	a character vector with five arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3 and 2->4. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution.
cuts.12	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=2. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.13	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.14	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=4. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.23	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.24	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=4. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
ini.base.12	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=2. Default initial value is 1 for all parameters. See details for more details.

<code>ini.base.13</code>	a numeric vector of initial logarithm values for parameters of distribution from $X=1$ to $X=3$. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.14</code>	a numeric vector of initial logarithm values for parameters of distribution from $X=1$ to $X=4$. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.23</code>	a numeric vector of initial logarithm values for parameters of distribution from $X=2$ to $X=3$. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.24</code>	a numeric vector of initial values for parameters of distribution from $X=2$ to $X=4$. Default initial value is 1 for all parameters. See details for more details.
<code>cov.12</code>	a numeric vector (matrix or data.frame) with the explicative time-fixed variable(s) related to the time from $X=1$ to $X=2$.
<code>init.cov.12</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.12</code> . Default initial value is 0 for all parameters.
<code>names.12</code>	a character vector with name of explicative variables associated to <code>cov.12</code> (optional).
<code>cov.13</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from $X=1$ to $X=3$.
<code>init.cov.13</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.13</code> . Default initial value is 0 for all parameters.
<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.14</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from $X=1$ to $X=3$.
<code>init.cov.14</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.14</code> . Default initial value is 0 for all parameters.
<code>names.14</code>	a character vector with name of explicative variables associated to <code>cov.14</code> (optional).
<code>cov.23</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from $X=2$ to $X=3$.
<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.
<code>names.23</code>	a character vector with name of explicative variables associated to <code>cov.23</code> (optional).
<code>cov.24</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from $X=2$ to $X=3$.
<code>init.cov.24</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.24</code> . Default initial value is 0 for all parameters.
<code>names.24</code>	a character vector with name of explicative variables associated to <code>cov.24</code> (optional).
<code>conf.int</code>	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is TRUE.
<code>silent</code>	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is TRUE, which corresponds to silent mode (no display).
<code>precision</code>	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .

`maxiter` a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

$$\begin{aligned} \text{Exponential distribution} \quad \lambda(t) &= 1/\sigma \\ \text{Weibull distribution} \quad \lambda(t) &= \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \\ \text{Generalized Weibull distribution} \quad \lambda(t) &= \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \end{aligned}$$

with σ, ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in `ini.base.12`, `ini.base.13`, `ini.base.14`, `ini.base.23` and `ini.base.24`. Moreover, the estimated values also correspond to the logarithms. Maximization of the log-likelihood function is performed using the algorithms of Nelder and Mead (`optim` function).

Value

`Table` a data.frame containing the parameter estimations of the model (`Estimate`). When the option `conf.int=TRUE` is specified, this data.frame includes three additional columns: the Standard Errors of parameters (`Std.Error`), the value of the Wald statistic (`t.value`), and the p-value for the Wald test (`Pr(>|t|)`).

`CovarianceMatrix` a data.frame corresponding to variance-covariance matrix of the estimated model.

`LogLik` a numeric value corresponding to the log-likelihood of the estimated model.

`AIC` a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
 Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. Manuscript submitted. 2015.

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
```

```

dataDIVATsample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
  replace = FALSE),]

# 4-state parametric semi-Markov model including one explicative variable ('z')
# on all transitions

# Note: a semi-Markovian process with sojourn times exponentially distributed
# is a time-homogeneous Markov process

sm4(t1=dataDIVATsample$time1,
    t2=dataDIVATsample$time2,
    sequence=dataDIVATsample$trajectory,
    dist=c("E","E","E","E","E"),
    cuts.12=NULL,
    cuts.13=NULL,
    cuts.14=NULL,
    cuts.23=NULL,
    cuts.24=NULL,
    ini.base.12=c(8.31),
    ini.base.13=c(10.46),
    ini.base.14=c(10.83),
    ini.base.23=c(9.01),
    ini.base.24=c(10.81),
    cov.12=dataDIVATsample$z,
    init.cov.12=c(-0.02),
    names.12=c("beta12_z") ,
    cov.13=dataDIVATsample$z,
    init.cov.13=c(1.42),
    names.13=c("beta13_z") ,
    cov.14=dataDIVATsample$z,
    init.cov.14=c(1.64),
    names.14=c("beta14_z") ,
    cov.23=dataDIVATsample$z,
    init.cov.23=c(0.30),
    names.23=c("beta23_z") ,
    cov.24=dataDIVATsample$z,
    init.cov.24=c(1.40),
    names.24=c("beta24_z"),
    conf.int=TRUE,
    silent=FALSE,
    precision=0.001,
    maxiter=1000)

```

smrs3

Computes a 3-state parametric additive relative survival semi-Markov model by likelihood maximization

Description

The 3-state SMRS model includes an initial state ($X=1$), a transient state ($X=2$), and the death ($X=3$ for observed death regardless of cause). Usually, $X=1$ corresponds to disease-free or remission and $X=2$ to relapse. The possible transitions are: $1 \rightarrow 2$, $1 \rightarrow 3$ and $2 \rightarrow 3$. In additive relative survival analysis, the observed mortality hazard ($X=3$) is assumed to be the sum of two components: the

expected population mortality ($X=P$) and the excess mortality related to the disease under study ($X=E$). The expected population mortality hazard ($X=P$) can be obtained from the death rates provided by life tables of statistical national institutes. These tables indicate the proportion of people dead in a calendar year stratified by birthdate and gender. The instantaneous hazard functions for times to death related to the disease ($X=E$) are modelled here using parametric distributions.

Usage

```
smrs3(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
      ini.base.12=NULL, ini.base.13=NULL, ini.base.23=NULL, cov.12=NULL,
      init.cov.12=NULL, names.12=NULL, cov.13=NULL, init.cov.13=NULL,
      names.13=NULL, cov.23=NULL, init.cov.23=NULL, names.23=NULL,
      p.age, p.sex, p.year, p.rate.table,
      conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)
```

Arguments

t1	a numeric vector with the observed times in days from baseline to the first transition ($X=2$ or $X=3$) or to the right-censoring in $X=1$.
t2	a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in $X=2$. NA for individuals right-censored in $X=1$ or individuals who directly transited from $X=1$ to $X=3$.
sequence	a numeric vector with the sequence of observed states. Four possible values are allowed: 1 (individual right-censored in $X=1$), 12 (individual right-censored in $X=2$), 13 (individual who directly transited from $X=1$ to $X=3$), 123 (individual who transited from $X=1$ to $X=3$ through $X=2$).
dist	a character vector with three arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution.
cuts.12	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=1$ to $X=2$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.13	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=1$ to $X=3$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.23	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from $X=2$ to $X=3$. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
ini.base.12	a numeric vector of initial logarithm values for parameters of distribution from $X=1$ to $X=2$. Default initial value is 1 for all parameters. See details for more details.
ini.base.13	a numeric vector of initial logarithm values for parameters of distribution from $X=1$ to $X=3$. Default initial value is 1 for all parameters. See details for more details.

<code>ini.base.23</code>	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=3. Default initial value is 1 for all parameters. See <code>details</code> for more details.
<code>cov.12</code>	a numeric matrix (or <code>data.frame</code>) with the explicative time-fixed variable(s) related to the time from X=1 to X=2.
<code>init.cov.12</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.12</code> . Default initial value is 0 for all parameters.
<code>names.12</code>	a character vector with name of explicative variables associated to <code>cov.12</code> (optional).
<code>cov.13</code>	a numeric matrix (or <code>data.frame</code>) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.13</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.13</code> . Default initial value is 0 for all parameters.
<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.23</code>	a numeric matrix (or <code>data.frame</code>) with the explicative time-fixed variable(s) related to the time from X=2 to X=3.
<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.
<code>names.23</code>	a character vector with name of explicative variables associated to <code>cov.23</code> (optional).
<code>p.age</code>	a numeric vector with the ages in days at baseline (X=1).
<code>p.sex</code>	a character vector with the gender: male or female.
<code>p.year</code>	a numeric vector with entry date in the study in the date format, i.e. in number of days since 01.01.1960.
<code>p.rate.table</code>	a table of event rates, organized as a <code>ratetable</code> object.
<code>conf.int</code>	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is TRUE.
<code>silent</code>	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is TRUE, which corresponds to silent mode (no display).
<code>precision</code>	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .
<code>maxiter</code>	a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

$$\begin{aligned}
 \text{Exponential distribution} \quad & \lambda(t) = 1/\sigma \\
 \text{Weibull distribution} \quad & \lambda(t) = \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1} \\
 \text{Generalized Weibull distribution} \quad & \lambda(t) = \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1}
 \end{aligned}$$

with σ , ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in `ini.base.12`, `ini.base.13` and `ini.base.23`. Moreover, the estimated values also correspond to the logarithms. Maximization of

the log-likelihood function is performed using the algorithms of Nelder and Mead (`optim` function). Works only with ratetables that are split by age, sex and year.

Value

Table	a <code>data.frame</code> containing the parameter estimations of the model (<code>Estimate</code>). When the option <code>conf.int=TRUE</code> is specified, this <code>data.frame</code> includes three additional columns: the Standard Errors of parameters (<code>Std.Error</code>), the value of the Wald statistic (<code>t.value</code>), and the p-value for the Wald test (<code>Pr(> t)</code>).
CovarianceMatrix	a <code>data.frame</code> corresponding to variance-covariance matrix of the estimated model.
LogLik	a numeric value corresponding to the log-likelihood of the estimated model.
AIC	a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
 Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. Manuscript submitted. 2015.

Pohar M., Stare J. (2006) "Relative survival analysis in R." Computer Methods and Programs in Biomedicine, 81: 272-278

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
dataDIVAT$sample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
  replace = FALSE),]

# To use a 3-state model, individuals with trajectory 13 and 123 are censored at the time
# of transition into state X=3

dataDIVAT_3state<-dataDIVAT$sample
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==13]<-1
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==123]<-12
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==14]<-13
dataDIVAT_3state$trajectory[dataDIVAT_3state$trajectory==124]<-123

# import the expected mortality rates
```

```

data(fr.ratetable)

# 3-state parametric additive relative survival semi-Markov model including one
# explicative variable ('z') on all transitions

# Note: a semi-Markovian process with sojourn times exponentially distributed
# is a time-homogeneous Markov process

smrs3(t1=dataDIVAT_3state$time1,
      t2=dataDIVAT_3state$time2,
      sequence=dataDIVAT_3state$trajectory,
      dist=c("E", "E", "E"),
      cuts.12=NULL,
      cuts.13=NULL,
      cuts.23=NULL,
      ini.base.12=c(10.70),
      ini.base.13=c(11.10),
      ini.base.23=c(0.04),
      cov.12=dataDIVAT_3state$z,
      init.cov.12=c(0.04),
      names.12=c("beta12_z") ,
      cov.13=dataDIVAT_3state$z,
      init.cov.13=c(1.04),
      names.13=c("beta1E_z") ,
      cov.23=dataDIVAT_3state$z,
      init.cov.23=c(0.29),
      names.23=c("beta2E_z"),
      p.age=dataDIVAT_3state$ageR*365.24,
      p.sex=dataDIVAT_3state$sexR,
      p.year=as.date(paste("01", "01", dataDIVAT_3state$year.tx), order = "mdy"),
      p.rate.table=fr.ratetable,
      conf.int=TRUE,
      silent=FALSE,
      precision=0.001,
      maxiter=1000)

```

smrs4

Computes a 4-state parametric additive relative survival semi-Markov model by likelihood maximization

Description

The 4-state SMRS model includes an initial state ($X=1$), a transient state ($X=2$) and two absorbing states including death ($X=3$, and $X=4$ for observed death regardless of cause). The possible transitions are: $1 \rightarrow 2$, $1 \rightarrow 3$, $1 \rightarrow 4$, $2 \rightarrow 3$ and $2 \rightarrow 4$. In additive relative survival analysis, the observed mortality hazard ($X=4$) is assumed to be the sum of two components: the expected population mortality ($X=P$) and the excess mortality related to the disease under study ($X=E$). The expected population mortality hazard ($X=P$) can be obtained from the death rates provided by life tables of statistical national institutes. These tables indicate the proportion of people dead in a calendar year stratified by birthdate and gender. The instantaneous hazard functions for times to death related to the disease ($X=E$) are modelled here using parametric distributions.

Usage

```
smrs4(t1, t2, sequence, dist, cuts.12=NULL, cuts.13=NULL, cuts.14=NULL,
cuts.23=NULL, cuts.24=NULL, ini.base.12=NULL, ini.base.13=NULL, ini.base.14=NULL,
ini.base.23=NULL, ini.base.24=NULL, cov.12=NULL, init.cov.12=NULL, names.12=NULL,
cov.13=NULL, init.cov.13=NULL, names.13=NULL, cov.14=NULL, init.cov.14=NULL,
names.14=NULL, cov.23=NULL, init.cov.23=NULL, names.23=NULL, cov.24=NULL,
init.cov.24=NULL, names.24=NULL, p.age, p.sex, p.year, p.rate.table,
conf.int=TRUE, silent=TRUE, precision=10^(-6), maxiter=10^4)
```

Arguments

t1	a numeric vector with the observed times in days from baseline to the first transition (X=2 or direct transition to X=3 or X=4) or to the right-censoring in X=1.
t2	a numeric vector with the observed times in days from baseline to the second transition or to the right censoring in X=2. NA for individuals right-censored in X=1 or individuals who directly transited from X=1 to X=3 or to X=1 to X=4.
sequence	a numeric vector with the sequence of observed states. Six possible values are allowed: 1 (individual right-censored in X=1), 12 (individual right-censored in X=2), 13 (individual who directly transited from X=1 to X=3), 14 (individual who directly transited from X=1 to X=4), 123 (individual who transited from X=1 to X=3 through X=2), 124 (individual who transited from X=1 to X=4 through X=2).
dist	a character vector with five arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3 and 2->4. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution.
cuts.12	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=2. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.13	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.14	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=4. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.23	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=3. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
cuts.24	a numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=2 to X=4. Only internal timepoints are allowed: timepoints cannot be 0 or Inf. Default is NULL which means that the

	distribution is not piecewise. Piecewise model is only allowed for exponential distribution.
<code>ini.base.12</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=2. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.13</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.14</code>	a numeric vector of initial logarithm values for parameters of distribution from X=1 to X=4. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.23</code>	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=3. Default initial value is 1 for all parameters. See details for more details.
<code>ini.base.24</code>	a numeric vector of initial logarithm values for parameters of distribution from X=2 to X=4. Default initial value is 1 for all parameters. See details for more details.
<code>cov.12</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=2.
<code>init.cov.12</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.12</code> . Default initial value is 0 for all parameters.
<code>names.12</code>	a character vector with name of explicative variables associated to <code>cov.12</code> (optional).
<code>cov.13</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.13</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.13</code> . Default initial value is 0 for all parameters.
<code>names.13</code>	a character vector with name of explicative variables associated to <code>cov.13</code> (optional).
<code>cov.14</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=3.
<code>init.cov.14</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.14</code> . Default initial value is 0 for all parameters.
<code>names.14</code>	a character vector with name of explicative variables associated to <code>cov.14</code> (optional).
<code>cov.23</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=3.
<code>init.cov.23</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.23</code> . Default initial value is 0 for all parameters.
<code>names.23</code>	a character vector with name of explicative variables associated to <code>cov.23</code> (optional).
<code>cov.24</code>	a numeric matrix (or data.frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=3.
<code>init.cov.24</code>	a numeric vector of initial values for regression coefficients associated to <code>cov.24</code> . Default initial value is 0 for all parameters.
<code>names.24</code>	a character vector with name of explicative variables associated to <code>cov.24</code> (optional).

p.age	a numeric vector with the ages in days at baseline (X=1).
p.sex	a character vector with the gender: male or female.
p.year	a numeric vector with entry date in the study in the date format, i.e. in number of days since 01.01.1960.
p.rate.table	a table of event rates, organized as a ratetable object.
conf.int	a logical value specifying if the pointwise confidence intervals for parameters and the variance-covariance matrix should be returned. Default is TRUE.
silent	a logical value specifying if the log-likelihood value should be returned at each iteration. Default is TRUE, which corresponds to silent mode (no display).
precision	a numeric positive value indicating the required precision for the log-likelihood maximization between each iteration. Default is 10^{-6} .
maxiter	a numeric positive value indicating the maximum number of iterations for the log-likelihood maximization. Default is 10^4 .

Details

Hazard functions available are:

Exponential distribution	$\lambda(t) = 1/\sigma$
Weibull distribution	$\lambda(t) = \nu(\frac{1}{\sigma})^\nu t^{\nu-1}$
Generalized Weibull distribution	$\lambda(t) = \frac{1}{\theta} \left(1 + \left(\frac{t}{\sigma}\right)^\nu\right)^{\frac{1}{\theta}-1} \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1}$

with σ, ν , and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. To avoid negative values, these parameters are exponentiated in the model. Therefore, the user must initialize logarithm values of parameters in `ini.base.12`, `ini.base.13`, `ini.base.14`, `ini.base.23` and `ini.base.24`. Moreover, the estimated values also correspond to the logarithms. Maximization of the log-likelihood function is performed using the algorithms of Nelder and Mead (`optim` function). Works only with ratetables that are split by age, sex and year.

Value

Table	a data.frame containing the parameter estimations of the model (<code>Estimate</code>). When the option <code>conf.int=TRUE</code> is specified, this data.frame includes three additional columns: the Standard Errors of parameters (<code>Std.Error</code>), the value of the Wald statistic (<code>t.value</code>), and the p-value for the Wald test (<code>Pr(> t)</code>).
CovarianceMatrix	a data.frame corresponding to variance-covariance matrix of the estimated model.
LogLik	a numeric value corresponding to the log-likelihood of the estimated model.
AIC	a numeric value corresponding to the Akaike Information Criterion of the estimated model.

Author(s)

Yohann Foucher <Yohann.Foucher@univ-nantes.fr> and
 Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. Manuscript submitted. 2015.

Pohar M., Stare J. (2006) "Relative survival analysis in R." Computer Methods and Programs in Biomedicine, 81: 272-278

Examples

```
# import the observed data
# (X=1 corresponds to initial state with a functioning graft, X=2 to acute rejection episode,
# X=3 to return to dialysis, X=4 to death with a functioning graft)

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
dataDIVATsample<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, trunc(nrow(dataDIVAT)/10),
  replace = FALSE),]

# import the expected mortality rates

data(fr.ratetable)

# 4-state parametric additive relative survival semi-Markov model including one
# explicative variable ('z') on all transitions

# Note: a semi-Markovian process with sojourn times exponentially distributed
# is a time-homogeneous Markov process

smrs4(t1=dataDIVATsample$time1,
      t2=dataDIVATsample$time2,
      sequence=dataDIVATsample$trajectory,
      dist=c("E", "E", "E", "E", "E"),
      cuts.12=NULL,
      cuts.13=NULL,
      cuts.14=NULL,
      cuts.23=NULL,
      cuts.24=NULL,
      ini.base.12=c(8.34),
      ini.base.13=c(10.44),
      ini.base.14=c(10.70),
      ini.base.23=c(9.43),
      ini.base.24=c(11.11),
      cov.12=dataDIVATsample$z,
      init.cov.12=c(0.04),
      names.12=c("beta12_z") ,
      cov.13=dataDIVATsample$z,
      init.cov.13=c(1.02),
      names.13=c("beta13_z") ,
      cov.14=dataDIVATsample$z,
      init.cov.14=c(1.04),
      names.14=c("beta1E_z") ,
      cov.23=dataDIVATsample$z,
      init.cov.23=c(0.94),
      names.23=c("beta23_z") ,
      cov.24=dataDIVATsample$z,
      init.cov.24=c(0.30),
```

```

names.24=c("beta2E_z"),
p.age=dataDIVATsample$ageR*365.24,
p.sex=dataDIVATsample$sexR,
p.year=as.date(paste("01", "01", dataDIVATsample$year.tx), order = "mdy"),
p.rate.table=fr.ratetable,
conf.int=TRUE,
silent=FALSE,
precision=0.001,
maxiter=1000)

```

usa.ratetable	<i>Expected mortality rates of the general United States population.</i>
---------------	--

Description

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

Usage

```
data(usa.ratetable)
```

Format

The format is "ratetable". The attributes are:

<code>dim</code>	a numeric vector with the length of each dimension.
<code>dimnames</code>	a vector with the names of each variable of the three dimensions.
<code>dimid</code>	a character vector with the identification of the dimensions: age, year and sex.
<code>factor</code>	a vector of indicators equals to 1 if the corresponding dimension does not vary according to the time. Only the dimension related to sex is associated to 1.
<code>cutpoints</code>	a list of the thresholds to identify the changes in mortality rates according to the time-dependent dimensions (NULL for sex).
<code>class</code>	the class of the object: <code>ratetable</code> .

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

Examples

```
data(usa.ratetable)
```

```
is.ratetable(usa.ratetable)
```

Index

- *Topic **DIVAT**
 - dataDIVAT, [3](#)
- *Topic **France**
 - dataDIVAT, [3](#)
 - fr.ratetable, [4](#)
- *Topic **Markov**
 - m3, [5](#)
 - m4, [9](#)
 - mrs3, [13](#)
 - mrs4, [16](#)
 - Multistate-package, [2](#)
- *Topic **United States**
 - usa.ratetable, [37](#)
- *Topic **datasets**
 - dataDIVAT, [3](#)
 - fr.ratetable, [4](#)
 - usa.ratetable, [37](#)
- *Topic **illness-death**
 - m3, [5](#)
 - sm3, [21](#)
- *Topic **multistate**
 - m3, [5](#)
 - m4, [9](#)
 - mrs3, [13](#)
 - mrs4, [16](#)
 - Multistate-package, [2](#)
 - sm3, [21](#)
 - sm4, [24](#)
 - smrs3, [28](#)
 - smrs4, [32](#)
- *Topic **ratetable**
 - fr.ratetable, [4](#)
 - usa.ratetable, [37](#)
- *Topic **relative survival**
 - mrs3, [13](#)
 - mrs4, [16](#)
 - Multistate-package, [2](#)
 - smrs3, [28](#)
 - smrs4, [32](#)
- *Topic **semi-Markov**
 - Multistate-package, [2](#)
 - sm3, [21](#)
 - sm4, [24](#)
 - smrs3, [28](#)
 - smrs4, [32](#)
- dataDIVAT, [3](#)
- fr.ratetable, [4](#)
- m3, [5](#)
- m4, [9](#)
- mrs3, [13](#)
- mrs4, [16](#)
- Multistate (Multistate-package), [2](#)
- Multistate-package, [2](#)
- sm3, [21](#)
- sm4, [24](#)
- smrs3, [28](#)
- smrs4, [32](#)
- usa.ratetable, [37](#)