

Package ‘multistate’

August 3, 2017

Type Package

Title Fitting Multistate Models

Version 0.2

Date 2017-08-02

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 multiple times-to-event. 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. In this package, we propose functions implementing such 3-state and 4-state multivariable and multistate models. The user can introduce multiple covariates to estimate conditional (subject-specific) effects. We also propose to adjust for possible confounding factors by using the Inverse Probability Weighting (IPW). When a state is patient death, the user can consider to take into account the mortality of the general population (relative survival approach). Finally, in the particular situation of one initial transient state and two competing and absorbing states, this package allows for estimating mixture models.

License GPL (>= 2)

LazyLoad yes

Depends R (>= 2.10), survival, statmod, date, relsurv

URL www.r-project.org, www.labcom-risca.com

NeedsCompilation no

Repository CRAN

Date/Publication 2017-08-03 14:34:43 UTC

R topics documented:

multistate-package	2
dataDIVAT	4
fr.ratetable	5

lrs.multistate	6
m3	7
m3rs	11
m4	15
m4rs	20
mm2	25
pred.mm2	28
sm3	31
sm3ic	34
sm3rs	39
sm4	43
sm4rs	48
usa.ratetable	53

Index	55
--------------	-----------

multistate-package	<i>Fitting 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) and to estimate the marginal causal effects by adjusting for possible confounding factors using Inverse Probability Weighting (IPW). Finally, in the particular situation of one initial transient state and two competing and absorbing states, this package allows for estimating mixture models.

The methods proposed are limited to parametric models. The 3-state models model include one initial state ($X=1$), one transient state ($X=2$), and an absorbing state ($X=3$). In other words, this is an illness-death model. 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. The 2-state mixture model includes on initial transient state ($X=1$) and two absorbing states in competition ($X=2$ and $X=3$). The possible transitions are: 1->2 and 1->3.

Details

Package:	multistate
Type:	package
Version:	0.2

Date: 2017-08-02
 License: GPL (>=2)
 LazyLoad: yes

mm2 2-state mixture model for data with competing events
 m3 3-state time-inhomogeneous Markov model
 m3rs 3-state relative survival time-inhomogeneous Markov model with additive risks
 m4 4-state time-inhomogeneous Markov model
 m4rs 4-state relative survival time-inhomogeneous Markov model with additive risks
 sm3 3-state semi-Markov model
 sm3rs 3-state relative survival semi-Markov model with additive risks
 sm3ic 3-state semi-Markov model with interval-censored data
 sm4 4-state semi-Markov model
 sm4rs 4-state relative survival semi-Markov model with additive risks
 lrs.multistate Likelihood ratio statistic to compare two embedded multistate models
 pred.mm2 Predictions from a 2-state mixture 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. *Statistical methods in medical research* Jun 2015. <DOI: 10.1177/ 0962280215586456>
- Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2016; 81: 272-278. <DOI: 10.1016/ j.cmpb.2006.01.004>
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/ 00273171.2011.568786>
- Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript in revision. 2016.
- Trebern-Launay K, Kessler M, Bayat-Makoei S, Querard AH, Briancon S, Giral M, Foucher Y. Horizontal mixture model for competing risks: a method to obtain easily interpretable results by both physicians and patients-illustration for waitlisted renal transplant candidates in a perspective of patient-centered decision making. Manuscript submitted. 2017.

See Also

URL: www.labcom-risca.com

 dataDIVAT

Sample of the DIVAT Cohort

Description

A data frame with 5943 French kidney transplant recipients from the DIVAT cohort. The patient evolution can be described according to a 4-state structure: X=2 represents the acute rejection episode, X=3 the definitive return to dialysis and X=4 the death. These times can be right-censored. A vector of covariates is also collected at the transplantation, i.e. the baseline of the cohort.

Usage

```
data(dataDIVAT)
```

Format

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

trajectory	A numeric vector with the sequences of observed states.
time1	A numeric vector with the times (in days) between the transplantation and the first clinical event (acute rejection episode, return to dialysis, or death with a functioning graft), or the times to censoring if trajectory=1
time2	A numeric vector with the time between the transplantation and the second clinical event (return to dialysis or death with a functioning graft), or the the time to censoring if trajectory=12
ageR	A numeric vector with the recipient age (in years) at the transplantation.
sexR	A character vector with the recipient gender.
year.tx	A numeric vector with the calendar year of the transplantation.
z	A 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$time<-NA
dataDIVAT$time<-ifelse(dataDIVAT$trajectory %in% c(1,12,13,14),
dataDIVAT$time1,dataDIVAT$time1+dataDIVAT$time2)

plot(survfit(Surv(time/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:

<code>dim</code>	A numeric vector with the length of each dimension.
<code>dimnames</code>	A character 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 numeric vector of indicators equals to 1 if the corresponding dimension does not vary according to the time. Only the dimension related to sex equals 1.
<code>cutpoints</code>	A list of the thresholds to identify the changes in the 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). The original data and updates can be downloaded from the Human Life-Table Database (HMD, The Human Mortality Database).

Source

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)
```

lrs.multistate	<i>Likelihood Ratio Statistic to Compare Embedded Multistate Models</i>
----------------	---

Description

This function computes a Likelihood Ratio Statistic to compare two embedded multistate models.

Usage

```
lrs.multistate(model1, model0)
```

Arguments

model1	A list containing the results after using a function included in the multistate package.
model0	A list containing the results after using a function included in the multistate package. The function used to obtain the model0 have to be the same than the one used to obtain the model1. The model0 have to be embedded in the model1.

Value

statistic	The value of the statistic.
ddl	Degrees of freedom.
pvalue	p-value.

Author(s)

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

Examples

```

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d3<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 250, 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

d3$trajectory[d3$trajectory==13]<-1
d3$trajectory[d3$trajectory==123]<-12
d3$trajectory[d3$trajectory==14]<-13
d3$trajectory[d3$trajectory==124]<-123

# 3-state parametric semi-Markov model : does 'z' influence both the
# transition 1->3 ? We only reduced the precision and the number of iteration
# to save time in this example, prefer the default values.

m1 <- sm3(t1=d3$time1, t2=d3$time2, sequence=d3$trajectory, dist=c("E","E","E"),
  ini.dist.12=c(9.93), ini.dist.13=c(11.54), ini.dist.23=c(10.21),
  cov.12=d3$z, init.cov.12=c(-0.13), names.12=c("beta12_z"),
  cov.13=d3$z, init.cov.13=c(1.61), names.13=c("beta13_z"),
  conf.int=TRUE, silent=FALSE, precision=0.001)

m1

m0 <- sm3(t1=d3$time1, t2=d3$time2, sequence=d3$trajectory, dist=c("E","E","E"),
  ini.dist.12=c(9.93), ini.dist.13=c(11.54), ini.dist.23=c(10.21),
  cov.12=d3$z, init.cov.12=c(-0.13), names.12=c("beta12_z"),
  conf.int=TRUE, silent=FALSE, precision=0.001)

m0

lrs.multistate(model1=m1, model0=m0)

```

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: $1 \rightarrow 2$, $1 \rightarrow 3$ and $2 \rightarrow 3$.

Usage

```
m3(t1, t2, sequence, weights=NULL, dist,
   cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
   ini.dist.12=NULL, ini.dist.13=NULL, ini.dist.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))
```

Arguments

t1	A numeric vector with the observed times (in days) from baseline to the first transition (in X=2 or X=3) or to the right-censoring (in X=1 at the last follow-up).
t2	A numeric vector with the observed times (in days) from baseline to the second transition (in X=3) or to the right censoring (in X=2 at the last follow-up). 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).
weights	A numeric vector with the weights for correcting the contribution of each individual. Default is NULL which means that no weighting is applied.
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. When the user chooses "PE", the arguments "cut.XX" have also to be defined.
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.dist.12	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.

ini.dist.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.23	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
cov.12	A 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 (logarithm of the cause-specific hazards ratios) associated to cov.12. Default initial value is 0.
names.12	An optional character vector with name of explicative variables associated to cov.12.
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 (logarithm of the cause-specific hazards ratios) associated to cov.13. Default initial value is 0.
names.13	An optional character vector with name of explicative variables associated to cov.13.
cov.23	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.23	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.23. Default initial value is 0.
names.23	An optional character vector with name of explicative variables associated to cov.23.
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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in ini.dist.12, ini.dist.13 and ini.dist.23.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic

model.

Value

object	The character string indicating the estimated model: "m3 (3-state time-inhomogeneous markov model)".
dist	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3.
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.
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.
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.
covariates	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3 and 2->3.
table	A data frame containing the estimated parameters 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 related p-value (<code>Pr(> t)</code>).
cov.matrix	A data frame corresponding to variance-covariance matrix of the parameters.
LogLik	A numeric value corresponding to the (weighted) log-likelihood of the model.
AIC	A numeric value corresponding to the Akaike Information Criterion of the model.

Author(s)

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

References

- Huszi E, Abrahamowicz M, Alioum A, Binquet C, Quantin C. Relative survival multistate Markov model. *Stat Med.* 2012 Feb 10;31(3):269-86. <DOI: 10.1002/sim.4392>
- Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript submitted. 2016.
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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, and X=4 to death with a functioning graft
```

```

data(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d3<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 300, replace = FALSE),]

# Individuals with trajectory 13 and 123 are
# censored at the time of transition into state X=3

d3$trajectory[d3$trajectory==13]<-1
d3$trajectory[d3$trajectory==123]<-12
d3$trajectory[d3$trajectory==14]<-13
d3$trajectory[d3$trajectory==124]<-123

# 3-state parametric Markov model including one explicative variable
# (z is the delayed graft function) on the transition 1->2. We only reduced
# the precision and the number of iteration to save time in this example,
# prefer the default values.

m3(t1=d3$time1, t2=d3$time2, sequence=d3$trajectory, weights=NULL,
   dist=c("E", "E", "E"), ini.dist.12=c(9.93),
   ini.dist.13=c(11.54), ini.dist.23=c(10.21),
   cov.12=d3$z, init.cov.12=c(-0.13), names.12=c("beta12_z"),
   conf.int=TRUE, silent=FALSE, precision=0.001)

```

m3rs

3-state Relative Survival Markov Model with Additive Risks

Description

The 3-state Markov relative survival model includes an initial state ($X=1$), a transient state ($X=2$), and the death ($X=3$). The possible transitions are: 1->2, 1->3 and 2->3. Assuming additive risks, the observed mortality hazard is 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.

Usage

```

m3rs(t1, t2, sequence, weights=NULL, dist,
     cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
     ini.dist.12=NULL, ini.dist.13=NULL, ini.dist.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⁾)

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 at the last follow-up).
t2	A numeric vector with the observed times in days from baseline to the second transition or to the right censoring (in X=2 at the last follow-up). 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).
weights	A numeric vector with the weights for correcting the contribution of each individual. Default is NULL which means that no weighting is applied.
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. When the user chooses "PE", the arguments "cut.xx" have also to be defined.
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.dist.12	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.23	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
cov.12	A 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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.12</code> . Default initial value is 0.
<code>names.12</code>	An optional character vector with name of explicative variables associated to <code>cov.12</code> .
<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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.13</code> . Default initial value is 0.
<code>names.13</code>	An optional character vector with name of explicative variables associated to <code>cov.13</code> .
<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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.23</code> . Default initial value is 0.
<code>names.23</code>	An optional character vector with name of explicative variables associated to <code>cov.23</code> .
<code>p.age</code>	A numeric vector with the patient ages in days at baseline ($X=1$).
<code>p.sex</code>	A character vector with the genders: male or female.
<code>p.year</code>	A numeric vector with the entry dates in the study respecting the date format, i.e. in number of days since 01.01.1960.
<code>p.rate.table</code>	A list containing the information related to the expected rates of mortality. This list is 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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.dist.12`, `ini.dist.13` and `ini.dist.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary vari-

able is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

object	The character string indicating the estimated model: "m3 (3-state time-inhomogeneous markov model)".
dist	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3.
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.
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.
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.
covariates	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3 and 2->3.
table	A data frame containing the estimated parameters 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 (Std.Error), the value of the Wald statistic (<code>t.value</code>), and the related p-value (<code>Pr(> t)</code>).
cov.matrix	A data frame corresponding to variance-covariance matrix of the parameters.
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

- Huszti E, Abrahamowicz M, Alioum A, Biquet C, Quantin C. Relative survival multistate Markov model. *Stat Med*. 2012 Feb 10;31(3):269-86. <DOI: 10.1002/sim.4392>
- Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2016; 81: 272-278. <DOI: 10.1016/j.cmpb.2006.01.004>
- Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript submitted. 2016.
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d3<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 200, replace = FALSE),]

# Individuals with trajectory 13 and 123 are
# censored at the time of transition into state X=3

d3$trajectory[d3$trajectory==13]<-1
d3$trajectory[d3$trajectory==123]<-12
d3$trajectory[d3$trajectory==14]<-13
d3$trajectory[d3$trajectory==124]<-123

# import the expected mortality rates

data(fr.ratetable)

# 3-state Markov model with additive risks including one explicative variable
# (z is the delayed graft function) on all transitions. We only reduced
# the precision and the number of iteration to save time in this example,
# prefer the default values.

m3rs(t1=d3$time1, t2=d3$time2, sequence=d3$trajectory,
     dist=c("E", "E", "E"), ini.dist.12=c(8.34),
     ini.dist.13=c(10.70), ini.dist.23=c(11.10),
     cov.12=d3$z, init.cov.12=c(0.04), names.12=c("beta12_z"),
     cov.13=d3$z, init.cov.13=c(1.04), names.13=c("beta1E_z"),
     cov.23=d3$z, init.cov.23=c(0.29), names.23=c("beta2E_z"),
     p.age=d3$ageR*365.24, p.sex=d3$sexR,
     p.year=as.date(paste("01", "01", d3$year.tx), order = "mdy"),
     p.rate.table=fr.ratetable, conf.int=TRUE,
     silent=FALSE, precision=0.001)

```

Description

The 4-state Markov model includes an initial state ($X=1$), a transient state ($X=2$) and two absorbing states ($X=3$ and $X=4$). The possible transitions are: $1 \rightarrow 2$, $1 \rightarrow 3$, $1 \rightarrow 4$, $2 \rightarrow 3$ and $2 \rightarrow 4$.

Usage

```
m4(t1, t2, sequence, weights=NULL, 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))
```

Arguments

t1	A numeric vector with the observed times in days from baseline to the first transition (X=2, X=3 or X=4) or to the right-censoring (in X=1 at the last follow-up).
t2	A numeric vector with the observed times in days from baseline to the second transition or to the right censoring (in X=2 at the last follow-up). NA for individuals right-censored in X=1 or individuals who directly transitioned from X=1 to X=3 or 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).
weights	A numeric vector with the weights for correcting the contribution of each individual. When the vector is completed, the IPW estimator is implemented. Default is NULL which means that no weighting is applied.
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. When the user chooses "PE", the arguments "cut.XX" have also to be defined.
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 values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.14	A numeric vector of initial values for the distribution from X=1 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.23	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.24	A numeric vector of initial values for the distribution from X=2 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
cov.12	A 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 (logarithm of the cause-specific hazards ratios) associated to cov.12. Default initial value is 0.
names.12	An optional character vector with name of explicative variables associated to cov.12.
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 (logarithm of the cause-specific hazards ratios) associated to cov.13. Default initial value is 0.
names.13	An optional character vector with name of explicative variables associated to cov.13.
cov.14	A numeric matrix (or data frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=4.
init.cov.14	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.14. Default initial value is 0.
names.14	An optional character vector with name of explicative variables associated to cov.14.

cov.23	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.23	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.23. Default initial value is 0.
names.23	An optional character vector with name of explicative variables associated to cov.23.
cov.24	A numeric matrix (or data frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=4.
init.cov.24	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.24. Default initial value is 0.
names.24	An optional character vector with name of explicative variables associated to cov.24.
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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.base.12`, `ini.base.13` and `ini.base.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

object	A character string indicating the estimated model: "m4 (4-state time-inhomogeneous markov model)".
dist	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3, and 2->4.
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.

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.
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.
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.
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.
covariates	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3, 1->4, 2->3, and 2->4.
table	A data frame containing the estimated parameters 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 related p-value (<code>Pr(> t)</code>).
cov.matrix	A data frame corresponding to variance-covariance matrix of the parameters.
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

- Huszti E, Abrahamowicz M, Alioum A, Binquet C, Quantin C. Relative survival multistate Markov model. *Stat Med*. 2012 Feb 10;31(3):269-86. <DOI: 10.1002/sim.4392>
- Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2016; 81: 272-278. <DOI: 10.1016/j.cmpb.2006.01.004>
- Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript submitted. 2016.
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d4<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 200, replace = FALSE),]

# 4-state parametric Markov model including one explicative variable ('z')
# on the transition from X=1 to X=2. We only reduced
# the precision and the number of iteration to save time in this example,
# prefer the default values.

m4(t1=d4$time1, t2=d4$time2, sequence=d4$trajectory, dist=c("E","E","E","E","E"),
  ini.base.12=c(8.31), ini.base.13=c(10.46), ini.base.14=c(10.83),
  ini.base.23=c(9.01), ini.base.24=c(10.81),
  cov.12=d4$z, init.cov.12=c(-0.02), names.12=c("beta12_z") ,
  conf.int=TRUE, silent=FALSE, precision=0.001)$table

```

m4rs

4-state Relative Survival Markov Model with Additive Risks

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 death). The possible transitions are: $1 \rightarrow 2$, $1 \rightarrow 3$, $1 \rightarrow 4$, $2 \rightarrow 3$ and $2 \rightarrow 4$. Assuming additive risks, the observed mortality hazard ($X=4$) is 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.

Usage

```

m4rs(t1, t2, sequence, weights=NULL, dist,
cuts.12=NULL, cuts.13=NULL, cuts.14=NULL, cuts.23=NULL,
cuts.24=NULL, ini.dist.12=NULL, ini.dist.13=NULL,
ini.dist.14=NULL, ini.dist.23=NULL, ini.dist.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))

```

Arguments

t1 A numeric vector with the observed times in days from baseline to the first transition ($X=2$, $X=3$ or $X=4$) or to the right-censoring (in $X=1$ at the last follow-up).

t2	A numeric vector with the observed times in days from baseline to the second transition or to the right censoring (in X=2 at the last follow-up). NA for individuals right-censored in X=1 or individuals who directly transitioned from X=1 to X=3 or 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).
weights	a numeric vector with the weights for correcting the contribution of each individual. Default is NULL which means that no weighting is applied.
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. When the user chooses "PE", the arguments "cut.XX" have also to be defined.
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.dist.12	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.

<code>ini.dist.14</code>	A numeric vector of initial values for the distribution from X=1 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
<code>ini.dist.23</code>	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
<code>ini.dist.24</code>	A numeric vector of initial values for the distribution from X=2 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
<code>cov.12</code>	A 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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.12</code> . Default initial value is 0.
<code>names.12</code>	An optional character vector with name of explicative variables associated to <code>cov.12</code> .
<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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.13</code> . Default initial value is 0.
<code>names.13</code>	An optional character vector with name of explicative variables associated to <code>cov.13</code> .
<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=4.
<code>init.cov.14</code>	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to <code>cov.14</code> . Default initial value is 0.
<code>names.14</code>	An optional character vector with name of explicative variables associated to <code>cov.14</code> .
<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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.23</code> . Default initial value is 0.
<code>names.23</code>	An optional character vector with name of explicative variables associated to <code>cov.23</code> .
<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=4.
<code>init.cov.24</code>	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to <code>cov.24</code> . Default initial value is 0.
<code>names.24</code>	An optional character vector with name of explicative variables associated to <code>cov.24</code> .
<code>p.age</code>	A numeric vector with the patient ages in days at baseline (X=1).
<code>p.sex</code>	A character vector with the genders: male or female.
<code>p.year</code>	A numeric vector with the entry dates in the study respecting the date format, i.e. in number of days since 01.01.1960.
<code>p.rate.table</code>	A list containing the information related to the expected rates of mortality. This list is organized as a <code>ratetable</code> 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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.dist.12`, `ini.dist.13` and `ini.dist.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

object	A character string indicating the estimated model: "m4 (4-state relative survival markov model with additive risks)".
dist	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3, and 2->4.
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.
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.
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.
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.
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.
covariates	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3, 1->4, 2->3, and 2->4.
table	A data frame containing the estimated parameters 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 related p-value (<code>Pr(> t)</code>).

cov.matrix	A data frame corresponding to variance-covariance matrix of the parameters.
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

- Huszti E, Abrahamowicz M, Alioum A, Binquet C, Quantin C. Relative survival multistate Markov model. *Stat Med*. 2012 Feb 10;31(3):269-86. <DOI: 10.1002/sim.4392>
- Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2016; 81: 272-278. <DOI: 10.1016/j.cmpb.2006.01.004>
- Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript submitted. 2016.
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d3<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 200, replace = FALSE),]

# import the expected mortality rates

data(fr.ratetable)

# 4-state parametric additive relative survival Markov model including one
# explicative variable ('z') on the transition 1->2. We only reduced
# the precision and the number of iteration to save time in this example,
# prefer the default values.

m4rs(t1=d3$time1, t2=d3$time2, sequence=d3$trajectory, dist=c("E","E","E","E","E"),
  ini.dist.12=c(8.34), ini.dist.13=c(10.44), ini.dist.14=c(10.70),
  ini.dist.23=c(9.43), ini.dist.24=c(11.11),
```

```
cov.12=d3$z, init.cov.12=c(0.04), names.12=c("beta12_z"),
p.age=d3$ageR*365.24, p.sex=d3$sexR,
p.year=as.date(paste("01","01",d3$year.tx), order = "mdy"),
p.rate.table=fr.ratetable, conf.int=TRUE,
silent=FALSE, precision=0.001)
```

mm2

*Horizontal Mixture Model for Two Competing Events***Description**

The 2-state mixture model which includes an initial state (X=1) and two absorbing states in competition (X=2 and X=3). Parameters are estimated by (weighted) Likelihood maximization.

Usage

```
mm2(t, sequence, weights=NULL, dist, cuts.12=NULL, cuts.13=NULL,
ini.dist.12=NULL, ini.dist.13=NULL, cov.12=NULL, init.cov.12=NULL,
names.12=NULL, cov.13=NULL, init.cov.13=NULL, names.13=NULL,
cov.p=NULL, init.cov.p=NULL, names.p=NULL, init.intercept.p=NULL,
conf.int=TRUE, silent=TRUE, precision=10^(-6))
```

Arguments

t	A numeric vector with the observed times in days from baseline to the last observation.
sequence	A numeric vector with the sequence of observed states. Three possible values are allowed: 1 (the individual is right-censored in X=1), 12 (the individual transits to X=2) and 13 (the individual transits to X=3).
weights	A numeric vector with the weights for correcting the contribution of each individual. When the vector is completed, the IPW estimator is implemented. Default is NULL which means that no weighting is applied.
dist	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2 and 1->3. Arguments allowed are "E" for Exponential distribution, "PE" for the piecewise exponential distribution, "W" for Weibull distribution or "WG" for Generalized Weibull distribution. When the user choose "PE", the arguments "cut.XX" have also to be defined.
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>ini.dist.12</code>	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
<code>ini.dist.13</code>	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
<code>cov.12</code>	A 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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.12</code> . Default initial value is 0.
<code>names.12</code>	An optional character vector with name of explicative variables associated to <code>cov.12</code> .
<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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.13</code> . Default initial value is 0.
<code>names.13</code>	An optional character vector with name of explicative variables associated to <code>cov.13</code> .
<code>cov.p</code>	A matrix (or data frame) with the explicative time-fixed variable(s) related to the probability $P(X=2)$, which is regressing according to a logistic function.
<code>init.cov.p</code>	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to <code>cov.p</code> . Default initial value is 0.
<code>names.p</code>	An optional character vector with name of explicative variables associated to <code>cov.p</code> .
<code>init.intercept.p</code>	A numeric value to initiate the intercept of the logit of $P(X=2)$. Default value is 0.
<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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.dist.12`, `ini.dist.13` and `ini.dist.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is

the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

object	The character string indicating the estimated model: "mm2 (mixture model with two competing events)".
dist	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2 and 1->3.
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.
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.
covariates	A numeric vector indicating the numbers of covariates respectively related to the time to the event X=2, the time to the event X=3, the long-term probability $P(X=2)$.
table	A data frame containing the estimated parameters 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 (Std.Error), the value of the Wald statistic (<code>t.value</code>), and the related p-value for the Wald test (<code>Pr(> t)</code>).
cov.matrix	A data frame corresponding to variance-covariance matrix of the parameters.
LogLik	A numeric value corresponding to the (weighted) log-likelihood of the model.
AIC	A numeric value corresponding to the Akaike Information Criterion of the model.

Author(s)

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

References

Trebern-Launay K, Kessler M, Bayat-Makoei S, Querard AH, Briancon S, Giral M, Foucher Y. Horizontal mixture model for competing risks: a method to obtain easily interpretable results by both physicians and patients-illustration for waitlisted renal transplant candidates in a perspective of patient-centered decision making. Manuscript submitted. 2017.

Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

Examples

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

```

data(dataDIVAT)

dim(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d2<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 300, replace = FALSE),]

# Data-management: two competing events
# the patient death is now X=2
# the return in dialysis is now X=3

d2$time<-NA
d2$time[d2$trajectory==1]<-d2$time1[d2$trajectory==1]

d2$time[d2$trajectory==12]<-d2$time2[d2$trajectory==12]
d2$trajectory[d2$trajectory==12]<-1

d2$time[d2$trajectory==13]<-d2$time1[d2$trajectory==13]

d2$time[d2$trajectory==123]<-d2$time2[d2$trajectory==123]
d2$trajectory[d2$trajectory==123]<-13

d2$time[d2$trajectory==14]<-d2$time1[d2$trajectory==14]

d2$time[d2$trajectory==124]<-d2$time2[d2$trajectory==124]
d2$trajectory[d2$trajectory==124]<-14

d2$trajectory[d2$trajectory==14]<-12

table(d2$trajectory)

# Univariable horizontal mixture model one binary explicative variable
# z is 1 if delayed graft function and 0 otherwise

mm2.test <- mm2(t=d2$time, sequence=d2$trajectory, weights=NULL,
  dist=c("E","W"), cuts.12=NULL, cuts.13=NULL,
  ini.dist.12=c(9.28), ini.dist.13=c(9.92, -0.23),
  cov.12=d2$z, init.cov.12=0.84, names.12="beta_12",
  cov.13=d2$z, init.cov.13=0.76, names.13="beta_13",
  cov.p=NULL, init.cov.p=NULL, names.p=NULL, init.intercept.p=-0.75,
  conf.int=TRUE, silent=FALSE)

mm2.test$table

```

Description

This function allows to estimate a cumulative incidence function (CIF) from an horizontal mixture model with two competing events, i.e. the results obtained from the function mm2.

Usage

```
pred.mm2(model, event, times, cov.12=NULL, cov.13=NULL, cov.p=NULL)
```

Arguments

model	A list obtained by using the function mm2.
event	A numeric value for identifying the event for which the CIF has to be computed. Two possible values are allowed: 2 (for the CIF related to X=2) and 3 (for the CIF related to X=3).
times	A numeric vector with positive values related to the times for which the CIF has to be computed.
cov.12	A vector, matrix or data frame in which to look for variables related to the time from X=1 to X=2 with which to predict the CIF.
cov.13	A vector, matrix or data frame in which to look for variables related to the time from X=1 to X=3 with which to predict the CIF.
cov.p	A vector, matrix or data frame in which to look for variables related to the probability P(X=2).

Details

The covariates has to be identical than the ones included in the mixture model declared in the argument model. More precisely, the columns of cov.12, cov.13 and cov.p must correspond to the same variables.

Value

times	A numeric vector with the times for which the CIF has to be computed.
cif	A matrix with the predicted CIF for the times in columns and the individuals in rows.

Author(s)

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

References

Trebern-Launay K, KesslerM, Bayat-Makoei S, Querard AH, Briancon S, Giral M, Foucher Y. Horizontal mixture model for competing risks: a method to obtain easily interpretable results by both physicians and patients-illustration for waitlisted renal transplant candidates in a perspective of patient-centered decision making. Manuscript submitted. 2017.

Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

Examples

```

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

data(dataDIVAT)

dim(dataDIVAT)

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

dataDIVAT$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d2<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 300, replace = FALSE),]

# Data-management: two competing events
# the patient death is now X=2
# the return in dialysis is now X=3

d2$time<-NA
d2$time[d2$trajectory==1]<-d2$time1[d2$trajectory==1]

d2$time[d2$trajectory==12]<-d2$time2[d2$trajectory==12]
d2$trajectory[d2$trajectory==12]<-1

d2$time[d2$trajectory==13]<-d2$time1[d2$trajectory==13]

d2$time[d2$trajectory==123]<-d2$time2[d2$trajectory==123]
d2$trajectory[d2$trajectory==123]<-13

d2$time[d2$trajectory==14]<-d2$time1[d2$trajectory==14]

d2$time[d2$trajectory==124]<-d2$time2[d2$trajectory==124]
d2$trajectory[d2$trajectory==124]<-14

d2$trajectory[d2$trajectory==14]<-12

table(d2$trajectory)

# Univariable horizontal mixture model one binary explicative variable
# z is 1 if delayed graft function and 0 otherwise

mm2.model <- mm2(t=d2$time, sequence=d2$trajectory, weights=NULL,
  dist=c("E","W"), cuts.12=NULL, cuts.13=NULL,
  ini.dist.12=c(9.28), ini.dist.13=c(9.92, -0.23),
  cov.12=d2$z, init.cov.12=0.84, names.12="beta_12",
  cov.13=d2$z, init.cov.13=0.76, names.13="beta_13",
  cov.p=NULL, init.cov.p=NULL, names.p=NULL, init.intercept.p=-0.75,
  conf.int=TRUE, silent=FALSE)

cif2.mm2 <- pred.mm2(mm2.model, event=2, times=seq(0, 4000, by=30),

```

```

cov.12=c(0,1), cov.13=c(0,1), cov.p=NULL)

plot(cif2.mm2$times/365.25, cif2.mm2$cif[1,], col = 1, type="l", lty = 1,
ylim=c(0,1), lwd =2, ylab="Cumulative Incidence Function",
xlab="Times (years)", main="", xlim=c(0, 11), legend=FALSE)

lines(cif2.mm2$times/365.25, cif2.mm2$cif[2,], lwd=2, col=2)

```

sm3

*3-state Semi-Markov Model***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: $1 \rightarrow 2$, $1 \rightarrow 3$ and $2 \rightarrow 3$.

Usage

```

sm3(t1, t2, sequence, weights=NULL, dist,
cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
ini.dist.12=NULL, ini.dist.13=NULL, ini.dist.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))

```

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 $X=1$ at the last follow-up).
t2	A numeric vector with the observed times in days from baseline to the second transition or to the right censoring (in $X=2$ at the last follow-up). 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 $X=2$).
weights	A numeric vector with the weights for correcting the contribution of each individual. When the vector is completed, the IPW estimator is implemented. Default is NULL which means that no weighting is applied.
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. When the user choose "PE", the arguments "cut.XX" have also to be defined.

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.dist.12	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.23	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
cov.12	A 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 (logarithm of the cause-specific hazards ratios) associated to cov.12. Default initial value is 0.
names.12	An optional character vector with name of explicative variables associated to cov.12.
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 (logarithm of the cause-specific hazards ratios) associated to cov.13. Default initial value is 0.
names.13	An optional character vector with name of explicative variables associated to cov.13.
cov.23	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.23	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.23. Default initial value is 0.
names.23	An optional character vector with name of explicative variables associated to cov.23.
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} .

Details

Hazard functions available are:

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

with $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.dist.12`, `ini.dist.13` and `ini.dist.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

<code>object</code>	The character string indicating the estimated model: "sm3 (3-state semi-markov model)".
<code>dist</code>	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3.
<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.
<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.
<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.
<code>covariates</code>	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3 and 2->3.
<code>table</code>	A data frame containing the estimated parameters 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 (Std.Error), the value of the Wald statistic (<code>t.value</code>), and the related p-value for the Wald test (<code>Pr(> t)</code>).
<code>cov.matrix</code>	A data frame corresponding to variance-covariance matrix of the parameters.
<code>LogLik</code>	A numeric value corresponding to the (weighted) log-likelihood of the model.
<code>AIC</code>	A numeric value corresponding to the Akaike Information Criterion of the 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. *Statistical methods in medical research* Jun 2015. <DOI: 10.1177/ 0962280215586456>

Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d3<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 250, 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

d3$trajectory[d3$trajectory==13]<-1
d3$trajectory[d3$trajectory==123]<-12
d3$trajectory[d3$trajectory==14]<-13
d3$trajectory[d3$trajectory==124]<-123

# 3-state parametric semi-Markov model including one explicative variable
# on the transition 1->2 (z is 1 if delayed graft function and 0 otherwise).
# We only reduced the precision and the number of iteration to save time in this example,
# prefer the default values.

sm3(t1=d3$time1, t2=d3$time2, sequence=d3$trajectory, dist=c("E","E","E"),
    ini.dist.12=c(9.93), ini.dist.13=c(11.54), ini.dist.23=c(10.21),
    cov.12=d3$z, init.cov.12=c(-0.13), names.12=c("beta12_z"),
    conf.int=TRUE, silent=FALSE, precision=0.001)$table
```

sm3ic

3-State Semi-Markov Model with Interval-Censored Data

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. The time from X=1 to X=2 is interval-censored. Parameters are estimated by (weighted) Likelihood maximization.

Usage

```
sm3ic(t0, t1, t2, sequence, weights=NULL, dist,
cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
ini.dist.12=NULL, ini.dist.13=NULL, ini.dist.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),
legendre=30, homogeneous=TRUE)
```

Arguments

t0	A numeric vector with the observed times in days from baseline to the last observation time in X=1.
t1	A numeric vector with the observed times in days from baseline to the first observation time in X=2. NA for individuals right-censored in X=1 or individuals who are directly in X=3 after X=1 (without any observation in X=2).
t2	A numeric vector with the observed times in days from baseline to the last follow-up.
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 observed in X=3 after X=3, without any observation of X=2), 123 (individual who transited from X=1 to X=3 through X=2).
weights	A numeric vector with the weights for correcting the contribution of each individual. When the vector is completed, the IPW estimator is implemented. Default is NULL which means that no weighting is applied.
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. When the user choose "PE", the arguments "cut.XX" have also to be defined.
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.dist.12	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.23	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
cov.12	A 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 (logarithm of the cause-specific hazards ratios) associated to cov.12. Default initial value is 0.
names.12	An optional character vector with name of explicative variables associated to cov.12.
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 (logarithm of the cause-specific hazards ratios) associated to cov.13. Default initial value is 0.
names.13	An optional character vector with name of explicative variables associated to cov.13.
cov.23	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.23	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.23. Default initial value is 0.
names.23	An optional character vector with name of explicative variables associated to cov.23.
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} .
legendre	A numeric value indicating the number of knots and weights for Gaussian quadrature used in convolution products. Default is 30.
homogeneous	A logical value specifying if the time spent in the state X=1 is considered as non-associated with the distribution of the time from the entry in the state X=2 to the transition in the state X=3. Default is TRUE, assuming no association.

Details

Hazard functions available are:

Exponential distribution	$\lambda(t) = 1/\sigma$
Weibull distribution	$\lambda(t) = \nu \left(\frac{1}{\sigma}\right)^\nu t^{\nu-1}$
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. We advise to initialize the logarithm of these parameters in `ini.dist.12`, `ini.dist.13` and `ini.dist.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Two kinds of model can be estimated: homogeneous and non-homogeneous semi-Markov model. In the first one, the hazard functions only depend on the times spent in the corresponding state. Note that for the transitions from the state $X=1$, the time spent in the state corresponds to the chronological time from the baseline of the study, as for Markov models. In the second one, the hazard function of the transition from the state $X=2$ to $X=3$ depends on two time scales: the time spent in the state 2 which is the random variable of interest, and the time spend in the state $X=1$ as a covariate.

Value

<code>object</code>	The character string indicating the estimated model: "sm3ic (3-state semi-markov model with interval-censored data)".
<code>dist</code>	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3.
<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 $X=2$.
<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 $X=3$.
<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 $X=3$.
<code>covariates</code>	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3 and 2->3.
<code>table</code>	A data frame containing the estimated parameters 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 related p-value for the Wald test (<code>Pr(> t)</code>).
<code>cov.matrix</code>	A data frame corresponding to variance-covariance matrix of the parameters.
<code>LogLik</code>	A numeric value corresponding to the (weighted) log-likelihood of the model.
<code>AIC</code>	A numeric value corresponding to the Akaike Information Criterion of the model.

Author(s)

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

Florence Gillaizeau <Florence.Gillaizeau@univ-nantes.fr>

References

Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript submitted. 2016.

Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. *Statistical methods in medical research* Jun 2015. <DOI: 10.1177/0962280215586456>

Examples

```
# The example is too long to compute for a submission on the CRAN
# Remove the characters '#'

# import the observed data (read the application in Gillaizeau et al. for more details)
# 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$id<-c(1:nrow(dataDIVAT))
# set.seed(2)
# d3<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 100, replace = FALSE),]

# To illustrate the use of a 3-state model, the return in dialysis are right-censored

# d3$trajectory[d3$trajectory==13]<-1
# d3$trajectory[d3$trajectory==123]<-12
# d3$trajectory[d3$trajectory==14]<-13
# d3$trajectory[d3$trajectory==124]<-123
# table(d3$trajectory)

# X=2 is supposed to be interval-censored between 't0' and 't1' because
# health examinations take place each year after inclusion

# d3$t0<-NA
# d3$t1<-NA
# d3$time2_<-NA

# i<-d3$trajectory==1
# d3$t0[i]<-trunc(d3$time1[i]/365.24)*365.24+1
# d3$t1[i]<-NA
# d3$t2[i]<- d3$time1[i]+1

# i<-d3$trajectory==12
# d3$t0[i]<-trunc(d3$time1[i]/365.24)*365.24+1
# d3$t1[i]<-(trunc(d3$time1[d3$trajectory==12]/365.24)+1)*365.24
```

```

# d3$t2[i]<-pmax(d3$time2[i], (trunc(d3$time1[i]/365.24)+2)*365.24)

# i<-d3$trajectory==13
# d3$t0[i]<-trunc(d3$time1[i]/365.24)*365.24+1
# d3$t1[i]<-NA
# d3$t2[i]<-d3$time1[i]

# i<-d3$trajectory==123
# d3$t0[i]<-trunc(d3$time1[i]/365.24)*365.24+1
# d3$t1[i]<-(trunc(d3$time1[i]/365.24)+1)*365.24
# d3$t2[i]<- pmax(d3$time2[i], (trunc(d3$time1[i]/365.24)+2)*365.24)

# 3-state homogeneous semi-Markov model with interval-censored data
# including one binary explicative variable (z is 1 if delayed graft function and
# 0 otherwise).
# Estimation of the marginal effect of z on the transition from X=1 to X=2
# by adjusting for 2 possible confounding factors (age and gender)
# We only reduced the precision and the number of iteration to save time in this example,
# prefer the default values.

# propensity.score <- glm(z ~ ageR + sexR, family=binomial(link="logit"),data=d3)
# d3$fit<-propensity.score$fitted.values

# p1<-mean(d3$z)
# d3$w <- p1/d3$fit
# d3$w[d3$z==0]<-(1-p1)/(1-d3$fit[d3$z==0])

# sm3ic(t0=d3$t0, t1=d3$t1, t2=d3$t2, sequence=d3$trajectory, weights=d3$w,
# dist=c("E","E","E"), cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,
# ini.dist.12=c(8.23), ini.dist.13=c(10.92), ini.dist.23=c(10.67),
# cov.12=d3$z, init.cov.12=c(0.02), names.12=c("beta12_z"),
# conf.int=TRUE, silent=FALSE, precision=0.001, legendre=20)$table

```

sm3rs

3-State Relative Survival Semi-Markov Model with Additive Risks

Description

The 3-state SMRS model includes an initial state ($X=1$), a transient state ($X=2$), and the death ($X=3$). The possible transitions are: $1 \rightarrow 2$, $1 \rightarrow 3$ and $2 \rightarrow 3$. Assuming additive risks, the observed mortality hazard is 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.

Usage

```

sm3rs(t1, t2, sequence, weights=NULL, dist,
cuts.12=NULL, cuts.13=NULL, cuts.23=NULL,

```

```

ini.dist.12=NULL, ini.dist.13=NULL, ini.dist.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)

```

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 at the last follow-up).
t2	A numeric vector with the observed times in days from baseline to the second transition or to the right censoring (in X=2 at the last follow-up). 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).
weights	A numeric vector with the weights for correcting the contribution of each individual. Default is NULL which means that no weighting is applied.
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. When the user chooses "PE", the arguments "cut.XX" have also to be defined.
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.dist.12	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.

<code>ini.dist.23</code>	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
<code>cov.12</code>	A 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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.12</code> . Default initial value is 0.
<code>names.12</code>	An optional character vector with name of explicative variables associated to <code>cov.12</code> .
<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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.13</code> . Default initial value is 0.
<code>names.13</code>	An optional character vector with name of explicative variables associated to <code>cov.13</code> .
<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 (logarithm of the cause-specific hazards ratios) associated to <code>cov.23</code> . Default initial value is 0.
<code>names.23</code>	An optional character vector with name of explicative variables associated to <code>cov.23</code> .
<code>p.age</code>	A numeric vector with the patient ages in days at baseline (X=1).
<code>p.sex</code>	A character vector with the genders: male or female.
<code>p.year</code>	A numeric vector with the entry dates in the study respecting the date format, i.e. in number of days since 01.01.1960.
<code>p.rate.table</code>	A list containing the information related to the expected rates of mortality. This list is 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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.dist.12`, `ini.dist.13`

and `ini.dist.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

<code>object</code>	The character string indicating the estimated model: "sm3rs (3-state relative survival semi-Markov model with additive risks)".
<code>dist</code>	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3 and 2->3.
<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.
<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.
<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.
<code>covariates</code>	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3 and 2->3.
<code>table</code>	A data frame containing the estimated parameters 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 related p-value (<code>Pr(> t)</code>).
<code>cov.matrix</code>	A data frame corresponding to variance-covariance matrix of the parameters.
<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

- Gillaizeau F, Dantan E, Giral M, Foucher Y. A multistate additive relative survival semi-Markov model. *Statistical methods in medical research* Jun 2015. <DOI: 10.1177/ 0962280215586456>
- Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2016; 81: 272-278. <DOI: 10.1016/j.cmpb.2006.01.004>
- Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript submitted. 2016.

Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d3<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 150, 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

d3$trajectory[d3$trajectory==13]<-1
d3$trajectory[d3$trajectory==123]<-12
d3$trajectory[d3$trajectory==14]<-13
d3$trajectory[d3$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 is the delayed graft function) on all transitions. We only reduced
# the precision and the number of iteration to save time in this example,
# prefer the default values.

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

sm3rs(t1=d3$time1, t2=d3$time2, sequence=d3$trajectory, dist=c("E","E","E"),
  ini.dist.12=c(10.70), ini.dist.13=c(11.10), ini.dist.23=c(0.04),
  cov.12=d3$z, init.cov.12=c(0.04), names.12=c("beta12_z"),
  cov.13=d3$z, init.cov.13=c(1.04), names.13=c("beta1E_z"),
  cov.23=d3$z, init.cov.23=c(0.29), names.23=c("beta2E_z"),
  p.age=d3$ageR*365.24, p.sex=d3$sexR,
  p.year=as.date(paste("01","01",d3$year.tx),order = "mdy"),
  p.rate.table=fr.ratetable,
  conf.int=TRUE, silent=FALSE, precision=0.001)
```

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, weights=NULL, 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))
```

Arguments

t1	A numeric vector with the observed times in days from baseline to the first transition ($X=2$, $X=3$ or $X=4$) or to the right-censoring (in $X=1$ at the last follow-up).
t2	A numeric vector with the observed times in days from baseline to the second transition or to the right censoring (in $X=2$ at the last follow-up). NA for individuals right-censored in $X=1$ or individuals who directly transitioned from $X=1$ to $X=3$ or $X=4$.
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 $X=2$).
weights	A numeric vector with the weights for correcting the contribution of each individual. When the vector is completed, the IPW estimator is implemented. Default is NULL which means that no weighting is applied.
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. When the user chooses "PE", the arguments "cut.xx" have also to be defined.
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 values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.14	A numeric vector of initial values for the distribution from X=1 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.23	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.base.24	A numeric vector of initial values for the distribution from X=2 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
cov.12	A 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 (logarithm of the cause-specific hazards ratios) associated to cov.12. Default initial value is 0.
names.12	An optional character vector with name of explicative variables associated to cov.12.
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 (logarithm of the cause-specific hazards ratios) associated to cov.13. Default initial value is 0.
names.13	An optional character vector with name of explicative variables associated to cov.13.
cov.14	A numeric matrix (or data frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=4.

<code>init.cov.14</code>	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to <code>cov.14</code> . Default initial value is 0.
<code>names.14</code>	An optional character vector with name of explicative variables associated to <code>cov.14</code> .
<code>cov.23</code>	A numeric matrix (or data frame) with the explicative time-fixed variable(s) related to the time from <code>X=2</code> to <code>X=3</code> .
<code>init.cov.23</code>	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to <code>cov.23</code> . Default initial value is 0.
<code>names.23</code>	An optional character vector with name of explicative variables associated to <code>cov.23</code> .
<code>cov.24</code>	A numeric matrix (or data frame) with the explicative time-fixed variable(s) related to the time from <code>X=2</code> to <code>X=4</code> .
<code>init.cov.24</code>	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to <code>cov.24</code> . Default initial value is 0.
<code>names.24</code>	An optional character vector with name of explicative variables associated to <code>cov.24</code> .
<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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.base.12`, `ini.base.13` and `ini.base.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

`object` A character string indicating the estimated model: "sm4 (4-state semi-Markov model)".

<code>dist</code>	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3, and 2->4.
<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.
<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.
<code>cuts.14</code>	A numeric vector indicating the timepoints in days for the piecewise exponential distribution related to the time from X=1 to X=4.
<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.
<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.
<code>covariates</code>	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3, 1->4, 2->3, and 2->4.
<code>table</code>	A data frame containing the estimated parameters 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 related p-value for the Wald test (<code>Pr(> t)</code>).
<code>cov.matrix</code>	A data frame corresponding to variance-covariance matrix of the parameters.
<code>LogLik</code>	A numeric value corresponding to the (weighted) log-likelihood of the model.
<code>AIC</code>	A numeric value corresponding to the Akaike Information Criterion of the 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. *Statistical methods in medical research* Jun 2015. <DOI: 10.1177/ 0962280215586456>
 Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. *Manuscript submitted*. 2016.
 Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d4<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 200, replace = FALSE),]

# 4-state parametric semi-Markov model including one explicative variable
# (z is the delayed graft function) on the transition from X=1 to X=2

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

# We only reduced the precision and the number of iteration to save time in this example,
# prefer the default values.

sm4(t1=d4$time1, t2=d4$time2, sequence=d4$trajectory, dist=c("E","E","E","E","E"),
    ini.base.12=c(8.31), ini.base.13=c(10.46), ini.base.14=c(10.83),
    ini.base.23=c(9.01), ini.base.24=c(10.81),
    cov.12=d4$z, init.cov.12=c(-0.02), names.12=c("beta12_z"),
    conf.int=TRUE, silent=FALSE, precision=0.001)$table

```

sm4rs

4-State Relative Survival Semi-Markov Model with Additive Risks

Description

The 4-state SMRS model includes an initial state ($X=1$), a transient state ($X=2$) and two absorbing states ($X=3$ $X=4$ for death). The possible transitions are: 1->2, 1->3, 1->4, 2->3 and 2->4. Assuming additive risks, the observed mortality hazard is 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.

Usage

```

sm4rs(t1, t2, sequence, weights=NULL, dist,
cuts.12=NULL, cuts.13=NULL, cuts.14=NULL, cuts.23=NULL, cuts.24=NULL,
ini.dist.12=NULL, ini.dist.13=NULL, ini.dist.14=NULL,
ini.dist.23=NULL, ini.dist.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))

```

Arguments

t1	A numeric vector with the observed times in days from baseline to the first transition (X=2, X=3 or X=4) or to the right-censoring (in X=1 at the last follow-up).
t2	A numeric vector with the observed times in days from baseline to the second transition or to the right censoring (in X=2 at the last follow-up). NA for individuals right-censored in X=1 or individuals who directly transitioned from X=1 to X=3 or 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).
weights	A numeric vector with the weights for correcting the contribution of each individual. When the vector is completed, the IPW estimator is implemented. Default is NULL which means that no weighting is applied.
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. When the user chooses "PE", the arguments "cut.XX" have also to be defined.
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.dist.12	A numeric vector of initial values for the distribution from X=1 to X=2. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.13	A numeric vector of initial values for the distribution from X=1 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.14	A numeric vector of initial values for the distribution from X=1 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.23	A numeric vector of initial values for the distribution from X=2 to X=3. The logarithm of the parameters have to be declared. Default value is 1.
ini.dist.24	A numeric vector of initial values for the distribution from X=2 to X=4. The logarithm of the parameters have to be declared. Default value is 1.
cov.12	A 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 (logarithm of the cause-specific hazards ratios) associated to cov.12. Default initial value is 0.
names.12	An optional character vector with name of explicative variables associated to cov.12.
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 (logarithm of the cause-specific hazards ratios) associated to cov.13. Default initial value is 0.
names.13	An optional character vector with name of explicative variables associated to cov.13.
cov.14	A numeric matrix (or data frame) with the explicative time-fixed variable(s) related to the time from X=1 to X=4.
init.cov.14	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.14. Default initial value is 0.
names.14	An optional character vector with name of explicative variables associated to cov.14.
cov.23	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.23	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.23. Default initial value is 0.
names.23	An optional character vector with name of explicative variables associated to cov.23.
cov.24	A numeric matrix (or data frame) with the explicative time-fixed variable(s) related to the time from X=2 to X=4.
init.cov.24	A numeric vector of initial values for regression coefficients (logarithm of the cause-specific hazards ratios) associated to cov.24. Default initial value is 0.
names.24	An optional character vector with name of explicative variables associated to cov.24.

p.age	A numeric vector with the patient ages in days at baseline (X=1).
p.sex	A character vector with the genders: male or female.
p.year	A numeric vector with the entry dates in the study respecting the date format, i.e. in number of days since 01.01.1960.
p.rate.table	A list containing the information related to the expected rates of mortality. This list is 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} .

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 $\sigma, \nu,$ and $\theta > 0$. The parameter σ varies for each interval when the distribution is piecewise Exponential. We advise to initialize the logarithm of these parameters in `ini.dist.12`, `ini.dist.13` and `ini.dist.23`.

To estimate the marginal effect of a binary exposure, the weights may be equal to $1/p$, where p is the estimated probability that the individual belongs to his or her own observed group of exposure. The probabilities p are often estimated by a logistic regression in which the dependent binary variable is the exposure. The possible confounding factors are the explanatory variables of this logistic model.

Value

object	A character string indicating the estimated model: "sm4rs (4-state relative survival semi-Markov model with additive risks)".
dist	A character vector with two arguments describing respectively the distributions of duration time for transitions 1->2, 1->3, 1->4, 2->3, and 2->4.
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.
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.
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.
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.

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.
covariates	A numeric vector indicating the numbers of covariates respectively related to the transition 1->2, 1->3, 1->4, 2->3, and 2->4.
table	A data frame containing the estimated parameters 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 related p-value for the Wald test (<code>Pr(> t)</code>).
cov.matrix	A data frame corresponding to variance-covariance matrix of the parameters.
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. *Statistical methods in medical research* Jun 2015. <DOI: 10.1177/0962280215586456>
- Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2016; 81: 272-278. <DOI: 10.1016/j.cmpb.2006.01.004>
- Gillaizeau F, Senage T, Le Borgne F, Le Tourneau T, Roussel JC, Leffondre K, Porcher R, Giraudeau B, Dantan E, Foucher Y. Inverse Probability Weighting to control confounding in an illness-death model for interval-censored data. Manuscript submitted. 2016.
- Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* May 2011; 46(3): 399-424. <DOI: 10.1080/00273171.2011.568786>

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$id<-c(1:nrow(dataDIVAT))
set.seed(2)
d4<-dataDIVAT[dataDIVAT$id %in% sample(dataDIVAT$id, 300, replace = FALSE),]

# import the expected mortality rates

data(fr.ratetable)
```

```

# 4-state parametric additive relative survival semi-Markov model including one
# explicative variable (z is the delayed graft function) on the transition from X=1 to X=2

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

# We only reduced the precision and the number of iteration to save time in this example,
# prefer the default values.

sm4rs(t1=d4$time1, t2=d4$time2, sequence=d4$trajectory, dist=c("E","E","E","E","E"),
      ini.dist.12=c(8.34), ini.dist.13=c(10.44), ini.dist.14=c(10.70),
      ini.dist.23=c(9.43), ini.dist.24=c(11.11),
      cov.12=d4$z, init.cov.12=c(0.04), names.12=c("beta12_z"),
      p.age=d4$ageR*365.24, p.sex=d4$sexR,
      p.year=as.date(paste("01","01",d4$year.tx), order = "mdy"),
      p.rate.table=fr.ratetable, conf.int=TRUE,
      silent=FALSE, precision=0.001)

```

usa.ratetable

Expected Mortality Rates of the General United States Population.

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). The original data and updates can be downloaded from the Human Life-Table Database (HMD, The Human Mortality Database).

Source

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.

Pohar M, Stare J. Relative survival analysis in R. *Computer Methods and Programs in Biomedicine* 2016; 81: 272-278. <DOI: 10.1016/j.cmpb.2006.01.004>

Examples

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

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

Index

- *Topic **Markov**
 - m3, [7](#)
 - m3rs, [11](#)
 - m4, [15](#)
 - m4rs, [20](#)
 - *Topic **competing events**
 - mm2, [25](#)
 - pred.mm2, [28](#)
 - *Topic **datasets**
 - usa.ratetable, [53](#)
 - *Topic **dataset**
 - dataDIVAT, [4](#)
 - fr.ratetable, [5](#)
 - *Topic **illness-death**
 - sm3, [31](#)
 - sm3ic, [34](#)
 - *Topic **interval-censoring**
 - sm3ic, [34](#)
 - *Topic **ratetable**
 - fr.ratetable, [5](#)
 - usa.ratetable, [53](#)
 - *Topic **relative survival**
 - m3rs, [11](#)
 - m4rs, [20](#)
 - sm3rs, [39](#)
 - sm4rs, [48](#)
 - *Topic **semi-Markov**
 - sm3, [31](#)
 - sm3ic, [34](#)
 - sm3rs, [39](#)
 - sm4, [43](#)
 - sm4rs, [48](#)
- dataDIVAT, [4](#)
- fr.ratetable, [5](#)
- lrs.multistate, [6](#)
- m3, [7](#)
- m3rs, [11](#)
- m4, [15](#)
- m4rs, [20](#)
- mm2, [25](#)
- multistate (multistate-package), [2](#)
- multistate-package, [2](#)
- pred.mm2, [28](#)
- sm3, [31](#)
- sm3ic, [34](#)
- sm3rs, [39](#)
- sm4, [43](#)
- sm4rs, [48](#)
- usa.ratetable, [53](#)