Package 'SemiMarkov'

October 12, 2022

Type Package

Title Multi-States Semi-Markov Models

Version 1.4.6

Date 2019-06-27

Author Agnieszka Listwon-Krol, Philippe Saint-Pierre

Maintainer Agnieszka Listwon-Krol <krol@lunenfeld.ca>

Description Functions for fitting multi-state semi-Markov models to longitudinal data. A parametric maximum likelihood estimation method adapted to deal with Exponential, Weibull and Exponentiated Weibull distributions is considered. Right-censoring can be taken into account and both constant and time-varying covariates can be included using a Cox proportional model. Reference: A. Krol and P. Saint-Pierre (2015) <doi:10.18637/jss.v066.i06>.

Depends R (>= 2.10.0), numDeriv, MASS, Rsolnp

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2019-07-02 11:10:03 UTC

R topics documented:

| asthma | *Asthma control data* |

Description

Data from a follow-up study of severe asthmatic patients. At each visit, covariates are recorded and asthma was evaluated using the concept of control scores. Such scores reflect a global judgement of the disease gravity based on official criteria. Three levels are considered (optimal, suboptimal and unacceptable control) and can be used to define the subject's state at each visit. The aim is to investigate the evolution of asthma control and to evaluate the effect of covariates. The data contains an extraction of 371 patients with at least two visits. The table is presented in long format with one row for each observed transition between two states. The rows corresponding to the same subject are ordered chronologically. The last sojourn time is right-censored by the end of the study and represent the time until censoring. A censored transition is defined as a transition to the same state h->h.

Usage

```
data(asthma)
```

Format

A data frame containing 928 rows.

Each row represents a patient examination and contains several covariates.

id Patient identification number

state.h Starting state (1 for optimal, 2 for suboptimal and 3 for unacceptable control state)

state.j Arrival state (1 for optimal, 2 for suboptimal and 3 for unacceptable control state)

time Waiting (sojourn) time in state state.h

Severity Disease severity (1=severe, 0=mild-moderate asthma)

BMI Body Mass Index (1=BMI>=25, 0=otherwise)

Sex Sex (1=men, 0=women)

This presentation of the data implies that, for a given patient, the visited states are the sequence of state.h and the follow-up time is the the cumulated sum of time.

Source

ARIA (Association pour la Recherche en Intelligence Artificielle), France.

References

Saint-Pierre P., Combescure C., Daures J.P., Godard P. (2003). The analysis of asthma control under a Markov assumption with use of covariates. *Statistics in Medicine*, 22(24):3755-70.

Examples

```
data(asthma)
head(asthma)
```

---

hazard	*Computes hazard rates using an object of class* `semiMarkov` *or* `param.init`

Description

For a given vector of times, the function computes the hazard rates values of an object of class `semiMarkov` or `param.init` (which provided the hazard rates). Both, values of hazard rate of waiting time of semi-Markov process can be obtained.

Usage

```
hazard(object, type = "alpha", time = NULL,
      cov = NULL, s = 0, t = "last", Length = 1000)
```

Arguments

object	Object of class `semiMarkov` or `param.init`.
type	Type of hazard to be computed: `"alpha"` for the hazard rates of waiting times of the embedded Markov chain and `"lambda"` for the hazard rates of the semi-Markov process. Default is `"alpha"`.
time	A vector containing the time values for which the hazard rate is computed. Default value is a vector `seq(0, last, length = Length)` where `last` is the largest duration observed in the data set and `Length` is the length of the vector.
cov	A list with one component for each covariate. Each component gives values of covariates that are to be used for the hazard rates computation. For a time-fixed covariate a single value can be given whereas a whole vector of values is required for time dependent covariates. Default is `NULL` which corresponds to time-independent covariates all equal to 0. Note that the same covariates values are used for all transitions.
s	Starting value of the time interval $[s, t]$ which is used to compute the hazard rate. This argument is not considered when the vector `time` is defined. Default value is 0.
t	Ending value of the time interval $[s, t]$ which is used to compute the hazard rate. This argument is not considered when the vector `time` is defined. Default value is `last` which is the the largest duration observed in the data set.
Length	The number of points of the time interval $[s, t]$ for which the hazard rate is computed. These points are equally distributed in the time interval $[s, t]$. This argument is not considered when the vector `time` is defined. Default value is 1000.

Details

This function computes the hazard rates of waiting (or sojourn) times and the hazard rates of semi-Markov process defined in the parametric multi-state semi-Markov model described in Listwon and Saint-Pierre (2013). Additional details about the methodology behind the `SemiMarkov` package can be found in Limnios and Oprisan (2001), Foucher *et al.* (2006) and Perez-Ocon and Ruiz-Castro (1999).

The hazard rate of waiting time at time t represents the conditional probability that a transition from state h to state j is observed given that no event occurs until time t. In a parametric framework, the expression of the hazard rates can easily be obtained as the distributions of waiting time belong to a parametric family. The hazard rate values are calculated using the chosen distribution and the given values of the parameters. The effects of both constant and time-varying covariates on the hazard of waiting time can be studied using a proportional intensities model. The effects of covariates can then be interpreted in terms of relative risk.

The hazard rate of the semi-Markov process at time t represents the conditional probability that a transition into state j is observed given that the subject is in state h and that no event occurs until time t. The hazard rate of the semi-Markov process can be interpreted as the subject's risk of passing from state h to state j. This quantity can be deduced from the transition probabilities of the embedded Markov chain and from the distributions of waiting times.

This function can be used to compute the hazard rates for different values of the covariates or different values of the parameters. These hazard rates can then be plotted using plot.hazard.

Objects of classes `semiMarkov` and `param.init` can be used in the function `hazard`. These objects contain informations on the model and the values of the parameters for the waiting time distribution, the transition probability of Markov chain and the regression coefficients.

Value

Returns an object of class `hazard`.

Type	The type of hazard computed by the function `hazard`: the hazard of waiting time (`alpha`) or the hazard of the semi-Markov process (`lambda`).
vector	A data frame containing one vector for each possible transition. A vector contains values of the hazard rate associated to the vector of times.
Time	The vector of times used to compute the hazard rate.
Covariates	A list containing the values of the covariates (fixed or time-dependent).
Summary	A list of data frames (one for each possible transition). A dataframe contains quantiles, means, minimums and maximums of the hazard rate values.
Transition_matrix	
	A matrix containing informations on the model: the possible transitions and the distribution of waiting times for each transition (Exponential, Weibull or Exponentiated Weibull).
call	Recall the name of the model.

Author(s)

Agnieszka Listwon-Krol

References

Krol, A., Saint-Pierre P. (2015). *SemiMarkov : An R Package for Parametric Estimation in Multi-State Semi-Markov Models*. 66(6), 1-16.

Limnios, N., Oprisan, G. (2001). *Semi-Markov processes and reliability*. Statistics for Industry and Technology. Birkhauser Boston.

Foucher, Y., Mathieu, E., Saint-Pierre, P., Durand, J.F., Daures, J.P. (2006). A semi-Markov model based on Generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal*, 47(6), 825-833.

Perez-Ocon, R., Ruiz-Castro, J. E. (1999). Semi-markov models and applications, chapter 14, pages 229-238. *Kluwer Academic Publishers*.

See Also

plot.hazard, semiMarkov, param.init, summary.hazard, print.hazard

Examples

```
## Asthma control data
data(asthma)

## Definition of the model:  states, names,
# possible transtions and waiting times distributions
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- c("E","E")
mtrans_1[2, c(1,3)] <- c("E","E")
mtrans_1[3, c(1,2)] <- c("W","E")

## semi-Markov model without covariates
fit1 <- semiMarkov(data = asthma, states = states_1, mtrans = mtrans_1)

## Hazard rates of waiting time
alpha1 <- hazard(fit1)
plot(alpha1)

## Hazard rates of the semi-Markov process
lambda1 <- hazard(fit1, type = "lambda")
plot(lambda1)

## Defining a vector of equally distributed times
alpha2 <- hazard(fit1, s=0, t=3, Length=300)
plot(alpha2)

## Considering times observed in the data set
alpha3 <- hazard(fit1, time=sort(unique(asthma$time)))
plot(alpha3)
```

```
## semi-Markov model with a covariate "BMI"
fit2 <- semiMarkov(data = asthma, cov = as.data.frame(asthma$BMI),
        states = states_1, mtrans = mtrans_1)

## Time fixed covariate
## Covariate equal to 0 and 1 for each transition
alpha4 <- hazard(fit2)
alpha5 <- hazard(fit2, cov=1)
plot(alpha4,alpha5)

## Time dependent covariate
## Suppose that the covariate value is known for all times values
Time<-sort(unique(asthma$time))          # observed times in ascending order
Cov1<-sort(rbinom(length(Time), 1, 0.3))   # simulation of binary covariate
Cov2<-sort(rexp(length(Time), 5))          # simulation of numeric covariate
alpha6 <- hazard(fit2, time=Time, cov=Cov1)
plot(alpha6)
alpha7 <- hazard(fit2, time=Time, cov=Cov2)
plot(alpha7)

## semi-Markov model with two covariates
## "BMI" affects transitions "1->3" and "3->1"
## "Sex" affects transition "3->1"
SEX <- as.data.frame(asthma$Sex)
BMI <- as.data.frame(asthma$BMI)
fit3 <- semiMarkov(data = asthma, cov = as.data.frame(cbind(BMI,SEX)),
                   states = states_1, mtrans = mtrans_1,
                   cov_tra = list(c("13","31"),c("31")))
alpha8 <- hazard(fit3, cov=c(0,0))
alpha9 <- hazard(fit3, cov=c(1,1))
plot(alpha8,alpha9)
```

---

param.init   *Defines the initial values of parameters for a semi-Markov model*

Description

Function defining initial values of parameters of the waiting time distributions, probabilities of the
Markov chain and optional regression coefficients associated with covariates. The function can
either provides the default initial values (the same as those considered in the function semiMarkov)
or can be used to specify particular initial values.

Usage

```
param.init(data = NULL, cov = NULL, states, mtrans,
           cov_tra = NULL, cens = NULL, dist_init = NULL,
           proba_init=NULL, coef_init = NULL)
```

Arguments

data	data frame of the form `data.frame(id,state.h,state.j,time)`, where

- `id`: the individual identification number
- `state.h`: state left by the process
- `state.j`: state entered by the process
- `time`: waiting time in state `state.h`

The data.frame containts one row per transition (possibly several rows per patient). The data frame `data` is not needed if `proba_init` is provided.

cov	Optional data frame containing the covariates values.
states	A numeric vector giving the names of the states (names are values used in `state.h`).
mtrans	A quadratic matrix of characters describing the possible transitions and the distributions of waiting time. The rows represent the left states, and the columns represent the entered states. If an instantaneous transition is not allowed from state `h` to state `j`, then `mtrans` should have (h, j) entry `FALSE`, otherwise it should be `"E"` (or `"Exp"` or `"Exponential"`) for Exponential distribution, `"W"` (or `"Weibull"`) for Weibull distribution or `"EW"` (or `"EWeibull"` or `"Exponentiated Weibull"`) for Exponentiated Weibull distribution. If `TRUE` is used instead of the name of the distribution, then a Weibull distribution is considered. By definition of a semi-Markov model, the transitions into the same state are not possible. The diagonal elements of `mtrans` must be set to `FALSE` otherwise the function will stop.
cov_tra	Optional list of vectors: a vector is associated with covariates included in the model. For a given covariate, the vector contains the transitions `"hj"` for which the covariate have an effect (only the transitions not equal to `FALSE` in `mtrans` are allowed). The effect of covariates can then be considered only on specific transitions. By default, the effects of covariates on all the possible transitions are included in a model.
cens	A character giving the code for censored observations in the column `state.j` of the data. Default is `NULL` which means that the censoring is defined as a transition fron state h to state h.
dist_init	Optional numeric vector giving the initial values of the distribution parameters. Default is 1 for each distribution parameter. The length of the vector depends on the chosen distribution, number of transitions and states.
proba_init	Optional numeric vector giving the initial values of the transition probabilities. The sum of the probabilities in the same raw must be equal to 1. According to semi-Markov model, the probability to stay in the same state must be equal to 0. The default values for the transition probabilities are estimated from the data. If `data = NULL`, the argument `proba_init` is obligatory.
coef_init	Optional numeric vector giving the initial values of the regression coefficients associated with the covariates. Default is 0 for each regression coefficient meaning no effect of the covariate.

Details

This function returns a data frame containing the initial values of parameters of a semi-Markov model. The model parameters are the distribution parameters, the transition probabilities of the Markov chain and the regression coefficients associated with covariates. The number of parameters depends on the chosen model: the distributions of the sojourn times, the number of states and the transitions between states specified with the matrix `mtrans`, the number of covariates (`cov`) and their effects or not on transitions (`cov_tra`).

The default initial values for the distribution parameters are fixed to 1. As the three possible distributions are nested for respective parameters equal to 1 (See details of the `semiMarkov` function), the initial distribution corresponds to the exponential distribution with parameter equal to 1 (whatever the chosen distribution). The default initial values for the regression coefficients are fixed to 0 meaning that the covariates have no effect on the hazard rates of the sojourn times. These initial values may be changed using the arguments `dist_init` and `coef_init`.

By default, the initial probabilities are calculated by simple proportions. The probability associated to the transition from h to j is estimed by the number of observed transitions from state h to state j divided by the total number of transitions from state `h` observed in the data. The results are displayed in matrix `matrix.P`. The number of parameters for transition probabilities is smaller than the number of possible transitions as the probabilities in the same row sum up to one. Considering this point and that the probability to stay in the same state is zero, the user can change the initial values using the argument `proba_init`.

Value

This function returns an object of class `param.init` to be used in functions [semiMarkov](#) and [hazard](#). An object of class `param.init` consists of

nstates	The length of vector `states` interpreted as the number of possible states for the process.
table.state	A table, with starting states as rows and arrival states as columns, which provides the number of times that a transition between two states is observed. This argument is only returned when `data` is provided. It can be used to quickly summarize multi-state data.
Ncens	Number of individuals subjected to censoring.
matrix.P	Quadratic matrix, with starting states as rows and arrival states as columns, giving the default initial values for the transition propabilities of the embedded Markov chain. All diagonal values are zero. The sum of all probabilities of the same row is equal to one.
last	The largest duration observed in `data` if `data` is given.
Transition_matrix	A matrix containing the informations on the model definition : the possible transitions o and the distribution of waiting times for each transition (Exponential, Weibull or Exponentiated Weibull).
dist.init	A data frame giving the names of the parameters, transitions associated with and initial values of the distribution parameters.
proba.init	A data frame giving names of the parameters, transitions associated with and initial values of the probabilities of the embedded Markov chain.

coef.init   A data frame giving the names of covariates, transitions associated with and initial values of the regression coefficients.

Author(s)

Agnieszka Listwon-Krol

See Also

hazard, semiMarkov

Examples

```
## Asthma control data
data(asthma)

## Definition of the model:  states, names,
# possible transtions and waiting time distributions
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- c("W","W")
mtrans_1[2, c(1,3)] <- c("EW","EW")
mtrans_1[3, c(1,2)] <- c("W","W")

## Default initial values in a model without covariates
init_1 <- param.init(data = asthma, states = states_1, mtrans = mtrans_1)

## Definition of initial values in a model without covariates
init_2 <- param.init(data = asthma, states = states_1, mtrans = mtrans_1,
         dist_init=c(rep(1.5,6),rep(1.8,6),rep(2,2)),
         proba_init=c(0.2,0.8,0.3,0.7,0.35,0.65))

## Default initial values with a covariate "Sex"
# influencing transitions " 1->2" and "3->2"
init_3 <- param.init(data = asthma, cov=as.data.frame(asthma$Sex),
         states = states_1, mtrans = mtrans_1, cov_tra=list(c("12","32")))

## Definition of initial values with a covariate "Sex"
# influencing transitions " 1->2" and "3->2"
init_4 <- param.init(data = asthma, cov=as.data.frame(asthma$Sex),
         states = states_1, mtrans = mtrans_1, cov_tra=list(c("12","32")),
         dist_init=c(rep(1.5,6),rep(1.8,6),rep(2,2)),
         proba_init=c(0.2,0.8,0.3,0.7,0.35,0.65), coef_init=rep(0.3,2))

init_5 <- param.init(data = asthma, cov=as.data.frame(asthma$Sex),
         states = states_1, mtrans = mtrans_1, cov_tra=list(c("12","32")),
         coef_init=c(0.2,0.5))

## Definition of initial values without dataset in an illness-death model
## 1 - healthy, 2 - illness, 3 - death
states_2 <- c("1","2","3")
```

```
mtrans_2 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_2[1,c(2,3)] <- c("E","W")
mtrans_2[2,c(1,3)] <- c("EW","EW")
init_6<-param.init(states=states_2, mtrans=mtrans_2, proba_init=c(0.7,0.3,0.2,0.8))
```

---

plot.hazard   *Plot method for objects of class* hazard

Description

Plot method for one or several (maximum 10) objects of class hazard. Depending on the hazard rate chosen in the function hazard, the function plots either the hazard rates of sojourn times or the semi-Markov process hazard rate for each considered transition (one plot for each transition).

Usage

```
## S3 method for class 'hazard'
plot(x, x2 = NULL, x3 = NULL, x4 = NULL, x5 = NULL, x6 = NULL, x7 = NULL,
      x8 = NULL, x9 = NULL, x10 = NULL, transitions = NULL, names = NULL,
      legend = TRUE, legend.pos = NULL, cex = NULL, colors = NULL,
   xlab = "Time", ylab = "Hazard function", lwd = 3, type = "p", ...)
```

Arguments

x	Object of class hazard. At least one hazard object is needed.
x2	Object of class hazard. Default is NULL.
x3	Object of class hazard. Default is NULL.
x4	Object of class hazard. Default is NULL.
x5	Object of class hazard. Default is NULL.
x6	Object of class hazard. Default is NULL.
x7	Object of class hazard. Default is NULL.
x8	Object of class hazard. Default is NULL.
x9	Object of class hazard. Default is NULL.
x10	Object of class hazard. A maximum of ten hazard objects is possible. Default is NULL.
transitions	A character vector giving the transitions to be plotted. Default is NULL which means that all the possible transitions are displayed.
names	Names of the hazard rates. Default is NULL which means that the names used in the semiMarkov object are applied.
legend	A logical value specifying if a legend should be added. Default is TRUE.
legend.pos	A vector giving the legend position.
cex	character expansion factor relative to current par("cex").

colors	A vector of colours for the hazard rates.
xlab	x-axis label. Default is Time.
ylab	y-axis label. Default is Hazard function.
lwd	Thickness of lines or points.
type	Type of graph. Default are points p.
...	Further arguments for plot.

Value

No value returned.

Author(s)

Agnieszka Listwon-Krol

See Also

hazard, semiMarkov

Examples

```
## Asthma control data
data(asthma)

## Definition of the model:  states, names, possible transtions
# and waiting times distributions
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- c("E","E")
mtrans_1[2, c(1,3)] <- c("E","E")
mtrans_1[3, c(1,2)] <- c("W","E")
fit <- semiMarkov(data = asthma, states = states_1, mtrans = mtrans_1)
lambda<-hazard (fit, type = "lambda")

plot(lambda, names = c("lambda"),legend=FALSE)
plot(lambda, transitions = c("13","31"), names = c("lambda"),
legend.pos=c(2,0.09,2,0.4))

## semi-Markov model in each stratum of Severity
fit0 <- semiMarkov(data = asthma[asthma$Severity==0,],
        states = states_1, mtrans = mtrans_1)
fit1 <- semiMarkov(data = asthma[asthma$Severity==1,],
        states = states_1, mtrans = mtrans_1)
lambda0<-hazard (fit0, type = "lambda",s=0,t=5,Length=1000)
lambda1<-hazard (fit1, type = "lambda",s=0,t=5,Length=1000)
plot(lambda0,lambda1, names = c("lambda0", "lambda1"),
legend.pos=c(4,0.18,4,0.8,4,0.2,4,0.09,4,0.7,4,0.21))
```

```
## semi-Markov model with covariate "BMI"
fitcov <- semiMarkov(data = asthma, cov = as.data.frame(asthma$BMI),
        states = states_1, mtrans = mtrans_1)
lambda0<-hazard (fitcov, type = "lambda",cov = c(0))
lambda1<-hazard (fitcov, type = "lambda",cov = c(1))
plot(lambda0,lambda1, names = c("lambda0", "lambda1"))
```

---

print.hazard   *Print method for object of class* hazard

Description

Print method for objects of class hazard.

Usage

```
## S3 method for class 'hazard'
print(x, whole = FALSE, ...)
```

Arguments

x   An object of class hazard.

whole   A logical value indicating if the whole vectors of the object hazard should be
             displayed. Default is FALSE which means that only the first six values are re-
             turned.

...   Further arguments for print or summary.

Value

No value returned.

Author(s)

Agnieszka Listwon-Krol

See Also

hazard, plot.hazard

---

print.semiMarkov   *Print method for object of class* semiMarkov

---

Description

Print method for objects of class semiMarkov.

Usage

```
## S3 method for class 'semiMarkov'
print(x, CI=TRUE, Wald.test=TRUE, ...)
```

Arguments

x	An object of class semiMarkov.
CI	A logical value indicating if the confidence intervals for each parameter should be returned. Default is TRUE. The confidence level is chosen in semiMarkov.
Wald.test	A logical value indicating if the results of the Wald test for each parameter should be returned. Default is TRUE.
...	Further arguments for print or summary.

Value

No value returned.

Author(s)

Agnieszka Listwon-Krol

See Also

semiMarkov, summary.semiMarkov

---

semiMarkov   *Parametric estimation in multi-state semi-Markov models*

Description

This function computes the parametric maximum likelihood estimation in multi-state semi-Markov models in continuous-time. The effect of time varying or fixed covariates can be studied using a proportional intensities model for the hazard of the sojourn time.

Usage

```
semiMarkov(data, cov = NULL, states, mtrans, cov_tra = NULL,
          cens = NULL, dist_init=NULL, proba_init = NULL, coef_init = NULL,
          alpha_ci = 0.95, Wald_par = NULL, eqfun = NULL,
          ineqLB = NULL,ineqUB = NULL, control = list() )
```

Arguments

`data`	data frame of the form `data.frame(id,state.h,state.j,time)`, where

  - `id`: the individual identification number
  - `state.h`: state left by the process
  - `state.j`: state entered by the process
  - `time`: waiting time in state `state.h`

  This data.frame containts one row per transition (possibly several rows per patient).

`cov`	Optional data frame containing covariates values.
`states`	A numeric vector giving names of the states (names are values used in `state.h`).
`mtrans`	A quadratic matrix of characters describing the possible transitions and the distributions of waiting time. The rows represent the left states, and the columns represent the entered states. If an instantaneous transition is not allowed from state `h` to state `j`, then `mtrans` should have (h, j) entry `FALSE`, otherwise it should be `"E"` (or `"Exp"` or `"Exponential"`) for Exponential distribution, `"W"` (or `"Weibull"`) for Weibull distribution or `"EW"` (or `"EWeibull"` or `"Exponentiated Weibull"`) for Exponentiated Weibull distribution. If `TRUE` is used instead of the name of the distribution, then a Weibull distribution is considered. By definition of a semi-Markov model, the transitions into the same state are not possible. The diagonal elements of `mtrans` must be set to `FALSE` otherwise the function will stop.
`cov_tra`	Optional list of vectors: a vector is associated with covariate included in the model. For a given covariate, the vector contains the transitions `"hj"` for which the covariate have an effect (only the transitions specified in `mtrans` are allowed). The effect of covariates can then be considered only on specific transitions. By default, the effects of covariates on all the possible transitions are studied.
`cens`	A character giving the code for censored observations in the column `state.j` of the data. Default is `NULL` which means that the censoring is defined as a transition fron state h to state h.
`dist_init`	Optional numeric vector giving the initial values of the distribution parameters. Default is 1 for each distribution parameter. The length of the vector depend on the chosen distribution, number of transitions and states.
`proba_init`	Optional numeric vector giving the initial values of the transition probabilities. The sum of the probabilities in the same row must be equal to 1. According to semi-Markov model, the probability to stay in the same state must be equal to 0. The default values for the transition probabilities are estimated from the data. If `data = NULL`, the argument `proba_init` is obligatory.

coef_init   Optional numeric vector giving the initial values of the regression coefficients associated with the covariates. Default is 0 for each regression coefficient which means that the covariate has no effect.

alpha_ci   Confidence level to be considered for the confidence intervals. The default value is 0.95.

Wald_par   Optional numeric vector giving the values to be tested (null hypothesis) by the Wald test for each parameter. The Wald statistics are evaluated only for the parameters of distributions and regression coefficients. The length of this vector must then be equal to the number of those parameters. The order of the values must be as in the parameters table given by objects semiMarkov or param.init (excluding the parameters associated to the transition probabilities). The default values for the elements of Wald_par vector are 1 for the distribution parameters and 0 for the regression coefficients.

eqfun   Optional list given equality constraints between parameters. These constraints are passed using the equality constraint function that can be defined in the solnp optimization function. See below for details.

ineqLB   Optional list given values of lower bound for parameters. These values are used in the inequality constraint that can be defined in the solnp optimization function. See below for details.

ineqUB   Optional list given values of upper bound for parameters. These values are used in the inequality constraint that can be defined in the solnp optimization function. See below for details.

control   The control list of optimization parameters for solnp optimization function.

Details

This function fits parametric multi-state semi-Markov model described in Listwon and Saint-Pierre (2013) to longitudinal data. Additional details about the methodology behind the SemiMarkov package can be found in Limnios and Oprisan (2001), Foucher *et al.* (2006) and Perez-Ocon and Ruiz-Castro (1999).

Consider an homogeneous semi-Markov process with a finite state space. In a parametric framework, distributions of the waiting time belong to parametric families. The distribution of the waiting time can be chosen between the exponential, the Weibull and the exponentiated Weibull distributions. The exponential distribution with scale parameter $\sigma > 0$ has a density defined as follows

$$f(x) = (1/\sigma)exp(-x/\sigma).$$

The Weibull distribution with scale parameter $\sigma > 0$ and shape parameter $\nu > 0$ has a density given by (same as one defined in dweibull)

$$g(x) = (\nu/\sigma)(x/\sigma)^{\nu-1}exp(-(x/\sigma)^{\nu}).$$

The exponentiated Weibull distribution (or generalized Weibull) with scale parameter $\sigma > 0$, shape parameter $\nu > 0$ and family parameter equal to $\theta > 0$ has a density given by (same as one defined in function dgweibull from the R package rmutil)

$$h(x) = \theta(\nu/\sigma)(x/\sigma)^{\nu-1}exp(-(x/\sigma)^{\nu})(1 - exp(-(x/\sigma)^{\nu}))^{\theta-1}.$$

These three distributions are nested. The exponentiated Weibull density with $\theta = 1$ gives a Weibull distribution and the Weibull density with $\nu = 1$ gives the exponential density.

Note that the effects of both constant and time-varying covariates on the hazards of sojourn time can be studied using a proportional intensities model. The effects of covariates can then be interpreted in terms of relative risk.

The model parameters are the distribution parameters, the transition probabilities of the Markov chain and the regression coefficients associated with covariates. The number of parameters depends on the chosen model: the distributions of the sojourn times, the number of states and the transitions between states specified with the matrix `mtrans`, the number of covariates (`cov`) and their effects or not on transitions (`cov_tra`).

The default initial values for the distribution parameters are fixed to 1. As the three possible distributions are nested for parameters equal to 1 (See details of the `semiMarkov` function), the initial distribution corresponds to an exponential with parameter equal to 1 (whatever the chosen distribution). The default initial values for the regression coefficients are fixed to 0 meaning that the covariates have no effect on the hazard rates of the sojourn times. These initial values may be changed using the arguments `dist_init` and `coef_init`.

By default, the initial probabilities are calculated by simple proportions. The probability associated to the transition from h to j is estimed by the number of observed transitions from state h to state j divided by the total number of transitions from state `h` observed in the data. The results are displayed in matrix `matrix.P`. The number of parameters for transition probabilities is smaller than the number of possible transitions as the probabilities in the same row sum up to one. Considering this point and that the probability to stay in the same state is zero, the user can changed the initial values using the argument `proba_init`.

The Yinyu Ye optimization solver to nonlinear problem is applied to maximize the log-likelihood using the function `solnp` created by A. Ghalanos and S. Theussl. In order to modify the optimization parameters refer to the package `Rsolnp` documentation.

Some optimization difficulties can arrise when there is not enough information in the data to estimate each transition rate. The users can change the optimization parameters and the initial values. It may be appropriate to reduce the number of states in the model, the number of allowed transitions, or the number of covariate effects, to ensure convergence.

Some additionals constraints can be introduced using `eqfun`, `ineqLB` and `ineqUB`. These constraints on distribution parameters, transition probabilities and regression coefficients can be defined using lists of vectors. The argument `eqfun` gives the possibility to add constraints of type $par1 = a * par2$ (a is a constant). This equality constraint must be expressed with a vector of 3 elements where the first element is the identifier of the parameters type (`"dist"` for distribution parameters, `"proba"` for the transition probabilities and `"coef"` for the regression coefficients), the second and the third elements are the index of $par1$ and $par2$, respectively. The index values of distribution parameters, transition probabilities and regression coefficients can be found in the table provided by an object `semiMarkov`. The last element of the vector corresponds to the constant a. The arguments `ineqLB` and `ineqUB` allow to add constraints of type $par \geq a$ and $par \leq a$, respectively. These arguments are lists of vectors of length 3 where the first element is the type of the parameter (`"dist"`, `"proba"` or `"coef"`), the second element is the index of parameter par and the last one is the constant a. If a chosen constraint corresponds to a transition probability, it should be considered that the last probabilities in a row of the transition matrix are not estimated but obtained directly since the sum of transition probabilities in the same row is equal to 1. Thus, no additional constraints related to these parameters are permitted. Moreover, note that the argument `eqfun` does not allowed to define

relationships between parameters of different types (for instance, a transition probability can not be equal to a regression coefficient). The optional constraints on parameters should be used prudently since they may induce problems in the convergence of the optimization method. In particular, the Wald statistic and the standard deviation may not be computed for some parameters due to negative values in the hessian matrix. Note that the default constraints induce by the model definition are treated in priority.

Value

call	The original call to `semiMarkov`.
minus2loglik	Minus twice the maximized log-likelihood.
solution	Etimations of the distribution parameters, probabilities of the embedded Markov chain and regression coefficients (if any considered) for each transition specified in the matrix `mtrans`. This is a data frame with three columns: the label of the parameter, the transition associated with and the estimated value.
opt.message	The message giving the information on the optimization result returned by the `constrOptim.nl` function.
opt.iter	Number of outer iterations of the optimization method.
nstates	The length of vector `states` interpreted as the number of possible states for the process.
table.state	A table, with starting states as rows and arrival states as columns, which provides the number of observed transitions between two states. This argument can be used to quickly summarize multi-state data.
Ncens	Number of individuals subjected to censoring.
Transition_matrix	A matrix containing the informations on the model definition : the possible transitions and the distribution of waiting times for each transition (Exponential, Weibull or Exponentiated Weibull).
param.init	Recall the initial values of the parameters. The third column of this object can be used in `hazard` function.
table.dist	Statistics for the estimations of distribution parameters of waiting time distributions. For the exponential distribution one data frame for the parameter `sigma` is returned, for the Weibull distribution two data frames for `sigma` and `nu` are returned, and for the Exponentiated Weibull distribution three data frames for `sigma`, `nu` and `theta` are returned. The columns of each data frame are the possible transitions, the estimations, the standard deviations, the lower and upper bounds of confidence intervals, the Wald test null hypothesis, the Wald test statistics and the p-values of the Wald test when testing hypothesis `sigma=1`, `nu=1` or `theta=1`.
table.proba	A data frame giving the estimations of the transition probabilities of the Markov chain and their standard deviations. By definition, the probability associated to the last possible transition of each row of the matrix `mtrans` is equal to $1 - pr$, where pr is the sum of all other probabilities from the row.
table.coef	If some covariates are included in the model it returns a data frame with the statistics for the estimated values of the regression coefficients. The columns

of the data frame are the transitions associated with the coefficients, the estimations, the standard deviations, the lower and upper bounds of confidence intervals, the Wald test null hypothesis, the Wald test statistics and the p-values of the Wald test when testing hypothesis coef=0.

table.param   Data frame with the statistics for all the model parameters, that is `table.dist`, `table.proba` and `table.coef` in a single data frame.

Note

Printing a `semiMarkov` object by typing the object's name at the command line implicitly invokes `print.semiMarkov`.

Author(s)

Agnieszka Listwon-Krol, Philippe Saint-Pierre

References

Krol, A., Saint-Pierre P. (2015). *SemiMarkov : An R Package for Parametric Estimation in Multi-State Semi-Markov Models*. 66(6), 1-16.

Ghalanos, A. and Theussl S. (2012). Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method. R package version 1.14.

Limnios, N., Oprisan, G. (2001). *Semi-Markov processes and reliability*. Statistics for Industry and Technology. Birkhauser Boston.

Foucher, Y., Mathieu, E., Saint-Pierre, P., Durand, J.F., Daures, J.P. (2006). A semi-Markov model based on Generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal*, 47(6), 825-833.

Perez-Ocon, R., Ruiz-Castro, J. E. (1999). Semi-markov models and applications, chapter 14, pages 229-238. *Kluwer Academic Publishers*.

See Also

param.init, hazard, summary.semiMarkov, print.semiMarkov

Examples

```
## Asthma control data
data(asthma)

## Definition of the model:  states, names, possible transtions and waiting time
## distributions
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- c("E","E")
mtrans_1[2, c(1,3)] <- c("E","E")
mtrans_1[3, c(1,2)] <- c("W","E")
```

```
## semi-Markov model without covariates
fit1 <- semiMarkov(data = asthma, states = states_1, mtrans = mtrans_1)

## semi-Markov model with one covariate
## "BMI" affects all transitions
BMI <- as.data.frame(asthma$BMI)
fit2 <- semiMarkov(data = asthma, cov = BMI, states = states_1, mtrans = mtrans_1)
## semi-Markov model with one covariate
## "BMI" affects the transitions "1->3" and "3->1"
fit3 <- semiMarkov(data = asthma, cov = BMI, states = states_1, mtrans = mtrans_1,
                   cov_tra = list(c("13","31")))

## semi-Markov model with two covariates
## "BMI" affects the transitions "1->3" and "3->1"
## "Sex" affects the transition "3->1"
SEX <- as.data.frame(asthma$Sex)
fit4 <- semiMarkov(data = asthma, cov = as.data.frame(cbind(BMI,SEX)),
                   states = states_1, mtrans = mtrans_1,
                   cov_tra = list(c("13","31"),c("31")))

## semi-Markov model using specific initial values
## same model as "fit1" but using different initial values
## "fit5" and "fit6" are equivalent

init <- param.init(data = asthma, states = states_1, mtrans = mtrans_1,
        dist_init=c(rep(1.5,6),c(1.8)), proba_init=c(0.2,0.8,0.3,0.7,0.35,0.65))
fit5 <- semiMarkov(data = asthma, states = states_1, mtrans = mtrans_1,
                   dist_init=init$dist.init[,3], proba_init=init$proba.init[,3])
fit6 <- semiMarkov(data = asthma, states = states_1, mtrans = mtrans_1,
                   dist_init=c(rep(1.5,6),c(1.8)),
                   proba_init=c(0.2,0.8,0.3,0.7,0.35,0.65))


## The Wald test null hypothesis is modified
## Wald statistics when testing nullity of distribution parameters
## and regression coefficients equal to -1
fit7 <- semiMarkov(data = asthma, cov = BMI, states = states_1, mtrans = mtrans_1,
                   Wald_par = c(rep(0,7),rep(-1,6)))

## semi-Markov model with additional constraints
## distribution parameters sigma for transition "1->3" = sigma for transition "2->1"
fit8 <- semiMarkov(data = asthma, cov = BMI, states = states_1, mtrans = mtrans_1,
                   eqfun = list(c("dist",2,3,1)))

## semi-Markov model with additional constraints
## regression coefficients beta for transition "1->2" = beta for transition "2->1"
##                                                     = beta for transition "2->3"
fit9 <- semiMarkov(data = asthma, cov = BMI, states = states_1, mtrans = mtrans_1,
                   eqfun = list(c("coef",1,3,1),c("coef",1,4,1)))

## semi-Markov model with additional constraints
## regression coeficient beta for transition "1->2" belongs to [-0.2,0.2]
```

```
## and regression coeficient beta for transition "2->3" belongs to [-0.05,0.05]
fit10 <- semiMarkov(data = asthma, cov = BMI, states = states_1, mtrans = mtrans_1,
                    ineqLB = list(c("coef",1,-0.2),c("coef",4,-0.05)),
                    ineqUB = list(c("coef",1,0.2),c("coef",4,0.05)))
```

---

summary.hazard   *Summary method for objects of class* hazard

---

Description

Summary method for objects of class hazard.

Usage

```
## S3 method for class 'hazard'
summary(object, ...)
```

Arguments

object   An object of class hazard.

...   Further arguments for summary.

Details

For an object of class hazard, this function gives the informations on the type of hazard rates
(sojourn time or semi-Markov process), the chosen model, the distribution of the sojourn times, the
covariates and the vector of times.

Value

No value returned.

Author(s)

Agnieszka Listwon-Krol

See Also

hazard, print.hazard

---

summary.semiMarkov   *Summary method for objects of class* semiMarkov

---

Description

Summary method for objects of class semiMarkov.

Usage

```
## S3 method for class 'semiMarkov'
summary(object, all = TRUE, transitions = NULL, ...)
```

Arguments

object   An object of class semiMarkov.

all   A logical value indicating if the results should be displayed for all the possible transitions. If set to FALSE, the transitions to be displayed must be specified using the argument transitions. Default is TRUE.

transitions   A vector of characters specifying the transitions to be displayed when the argument all is set to FALSE.

...   Further arguments for summary.

Value

A list of data frames giving

Transition_matrix

A matrix containing the informations on the model definition : the possible transitions and the distribution of waiting times for each transition (Exponential, Weibull or Exponentiated Weibull).

param.init   Recall the initial values of the parameters. The third column of this object can be used in hazard function.

table.state   A table, with starting states as rows and arrival states as columns, which provides the number of observed transitions between two states. This argument can be used to quickly summarize multi-state data.

Ncens   Number of individuals subjected to censoring.

table.param   List of data frames (one for each transition). A data frame includes, for each parameter (distribution parameters, the transition probabilities and the regression coefficients), the estimation, the standard deviation, the lower and upper bounds of confidence interval, the Wald test statistic and Wald test p-value (for the distribution parameters and the regression coefficients).

Author(s)

Agnieszka Listwon-Krol

See Also

semiMarkov, print.semiMarkov

---

table.state   *Table giving the numbers of observed transitions*

Description

Function returning a table with numbers of transitions between two states observed in the data set. This table can be a used to summarize a multi-state data or to define the matrix mtrans required in the semiMarkov function.

Usage

```
table.state( data, states = NULL, mtrans = NULL, cens = NULL)
```

Arguments

data   data frame in form data.frame(id,state.h,state.j,time), where

- id: the individual identification number
- state.h: state left by the process
- state.j: state entered by the process
- time: waiting time in state state.h

This data.frame containts one row per transition (possibly several rows per patient).

states   A numeric vector giving the names of the states (names are values used in state.h).

mtrans   A quadratic matrix of logical values describing the possible transitions. The rows represent the left states, and the columns represent the entered states. If an instantaneous transition is not allowed from state h to state j, then mtrans should have (h, j) entry FALSE, otherwise it should be TRUE. Default value is a matrix which allows all the possible transitions between states.

cens   A character giving the code for censored observations in the column state.j of the data. Default is NULL which means that the censoring is defined as a transition fron state i to state i (by definition of a semi-Markov model, the transitions into the same state are not possible).

Value

table.state   A table, with starting states as rows and arrival states as columns, which provides the number of observed transitions between two states. This argument can be used to quickly summarize multi-state data.

Ncens   Number of individuals subjected to censoring.

Author(s)

Agnieszka Listwon-Krol

See Also

[param.init](#), [semiMarkov](#)

Examples

```
## Asthma control data
data(asthma)

# default description
# censoring is implicitly defined as a transition "h->h"
table.state(asthma)
table.state(asthma)$Ncens

# censoring defined as a transition to state "4"
asthma_bis<-asthma
for(i in 1:dim(asthma)[1]){if(asthma[i,2]==asthma[i,3]) asthma_bis[i,3]<-4}
table.state (asthma_bis, cens = 4)

## Definition of the model: states names and possible transtions
states_1 <- c("1","2","3")
mtrans_1 <- matrix(FALSE, nrow = 3, ncol = 3)
mtrans_1[1, 2:3] <- TRUE
mtrans_1[2, c(1,3)] <- c("W","E")
table.state(asthma, states = states_1, mtrans = mtrans_1)
```

Index