Package 'presmoothedTP'

July 23, 2025

Type Package

Title Presmoothed Landmark Aalen-Johansen Estimator of Transition
Probabilities for Complex Multi-State Models

Version 0.1.0

Description Multi-state models are essential tools in longitudinal data analysis.
One primary goal of these models is the estimation of transition
probabilities, a critical metric for predicting clinical prognosis
across various stages of diseases or medical conditions. Traditionally,
inference in multi-state models relies on the Aalen-Johansen (AJ)
estimator which is consistent under the Markov assumption. However,
in many practical applications, the Markovian nature of the process
is often not guaranteed, limiting the applicability of the AJ
estimator in more complex scenarios. This package extends the landmark Aalen-Johansen
estimator (Putter, H, Spitoni, C (2018) <doi:10.1177/0962280216674497>)
incorporating presmoothing techniques described by Soutinho,
Meira-Machado and Oliveira (2020) <doi:10.1080/03610918.2020.1762895>,
offering a robust alternative for estimating transition probabilities
in non-Markovian multi-state models with multiple states and potential
reversible transitions.

License GPL-3

Depends R (>= 3.5.0)

Imports survival, mstate, plyr

RoxygenNote 7.3.1

Encoding UTF-8

LazyLoad yes

LazyData yes

NeedsCompilation yes

Author Gustavo Soutinho [aut, cre] (ORCID:
<https://orcid.org/0000-0002-0559-1327>),
Luis Meira-Machado [aut] (ORCID:
<https://orcid.org/0000-0002-8577-7665>)

Maintainer Gustavo Soutinho <gustavosoutinho@sapo.pt>

Repository CRAN

Date/Publication 2025-01-22 17:10:02 UTC

Contents

---

colonMSM   *Chemotherapy for Stage B/C colon cancer.*

Description

These are data from one of the first successful trials of adjuvant chemotherapy for colon cancer. Levamisole is a low-toxicity compound previously used to treat worm infestations in animals; 5-FU is a moderately toxic (as these things go) chemotherapy agent.

Usage

```
data("colonMSM")
```

Format

A data frame with 929 observations on the following 15 variables. Below a brief description is given for some of these variables.

time1 Time to recurrence/censoring/death, whichever occurs first.

event1 Recurrence/censoring indicator (recurrence=1, alive=0).

Stime Time to censoring/death, whichever occurs first.

event Death/censoring indicator (death=1, alive=0).

rx Treatment - Obs(ervation), Lev(amisole), Lev(amisole)+5-FU.

sex Sex indicator (male=1, female=0).

age Age in years.

obstruct Obstruction of colon by tumour.

perfor Perforation of colon.

adhere Adherence to nearby organs.

nodes Number of lymph nodes with detectable cancer.

differ Differentiation of tumour (1=well, 2=moderate, 3=poor).

extent Extent of local spread (1=submucosa, 2=muscle, 3=serosa, 4=contiguous structures).

surg Time from surgery to registration (0=short, 1=long).

node4 More than 4 positive lymph nodes.

Source

The study is originally described in Laurie (1989).The main report is found in Moertel (1990). This data set is closest to that of the final report in Moertel (1991). A version of the data with less follow-up time was used in the paper by Lin (1994).

References

JA Laurie, CG Moertel, TR Fleming, HS Wieand, JE Leigh, J Rubin, GW McCormack, JB Gerstner, JE Krook and J Malliard. Surgical adjuvant therapy of large-bowel carcinoma: An evaluation of levamisole and the combination of levamisole and fluorouracil: The North Central Cancer Treatment Group and the Mayo Clinic. Journal of Clinical Oncology, 7:1447-1456, 1989.

DY Lin. Cox regression analysis of multivariate failure time data: the marginal approach. Statistics in Medicine, 13:2233-2247, 1994.

CG Moertel, TR Fleming, JS MacDonald, DG Haller, JA Laurie, PJ Goodman, JS Ungerleider, WA Emerson, DC Tormey, JH Glick, MH Veeder and JA Maillard. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. New England Journal of Medicine, 332:352-358, 1990.

CG Moertel, TR Fleming, JS MacDonald, DG Haller, JA Laurie, CM Tangen, JS Ungerleider, WA Emerson, DC Tormey, JH Glick, MH Veeder and JA Maillard. Fluorouracil plus Levamisole as and effective adjuvant therapy after resection of stage II colon carcinoma: a final report. Annals of Internal Medicine, 122:321-326, 1991.

Examples

```
data(colonMSM)
head(colonMSM)
```

---

data	*data*

Description

Simulated data in which the individuals can move from the initial state (State 1) to an intermediate state (State 2). From there, they may either proceed to the absorbing state (State 4), passing through another intermediate state (State 3), or move directly from the initial state to the absorbing state.

Usage

```
data("data")
```

Format

A data frame with 500 observations on the following 6 variables.

y1 First gap time.

state1 Indicator of the first gap time; 0 if the first time is censored and 1 otherwise.

y2 Second gap time.

state2 Censoring indicator of the second gap time; 0 if the total time is censored and 1 otherwise.

y3 Third gap time.

state3 Censoring indicator of the third gap time; 0 if the total time is censored and 1 otherwise.

Examples

```
data(data)
```

---

ebmt4   *Data from the European Society for Blood and Marrow Transplantation (EBMT).*

Description

A data frame of 2279 patients transplanted at the EBMT between 1985 and 1998. These data were used in Fiocco, Putter & van Houwelingen (2008), van Houwelingen & Putter (2008, 2012) and de Wreede, Fiocco & Putter (2011). The included variables are

id Patient identification number

rec Time in days from transplantation to recovery or last follow-up

rec.s Recovery status; 1 = recovery, 0 = censored

ae Time in days from transplantation to adverse event (AE) or last follow-up

ae.s Adverse event status; 1 = adverse event, 0 = censored

recae Time in days from transplantation to both recovery and AE or last follow-up

recae.s Recovery and AE status; 1 = both recovery and AE, 0 = no recovery or no AE or censored

rel Time in days from transplantation to relapse or last follow-up

rel.s Relapse status; 1 = relapse, 0 = censored

srv Time in days from transplantation to death or last follow-up

srv.s Relapse status; 1 = dead, 0 = censored

year Year of transplantation; factor with levels "1985-1989", "1990-1994", "1995-1998"

agecl Patient age at transplant; factor with levels "<=20", "20-40", ">40"

proph Prophylaxis; factor with levels "no", "yes"

match Donor-recipient gender match; factor with levels "no gender mismatch", "gender mismatch"

Usage

```
data("ebmt4")
```

Format

A data frame, see data.frame.

Author(s)

Gustavo Soutinho and Luis Meira-Machado.

Source

We acknowledge the European Society for Blood and Marrow Transplantation (EBMT) for making available these data. Disclaimer: these data were simplified for the purpose of illustration of the analysis of competing risks and multi-state models and do not reflect any real life situation. No clinical conclusions should be drawn from these data.

References

Fiocco M, Putter H, van Houwelingen HC (2008). Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. Statistics in Medicine 27, 4340-4358.

van Houwelingen HC, Putter H (2008). Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. Lifetime Data Anal 14, 447-463.

van Houwelingen HC, Putter H (2012). Dynamic Prediction in Clinical Survival Analaysis. Chapman & Hall/CRC Press, Boca Raton.

de Wreede LC, Fiocco M, and Putter H (2011). mstate: An R Package for the Analysis of Competing Risks and Multi-State Models. Journal of Statistical Software, Volume 38, Issue 7.

Examples

```
data(ebmt4)
head(ebmt4)
```

---

prepTP	*Function to prepare data set for multi-state modeling in long format from data set in wide format.*

Description

This function converts a data set which is in wide format (one subject per line, multiple columns indicating time and status for different states) into a data set in long format (one line for each transition for which a subject is at risk). Selected covariates are replicated per subjects.

Usage

```
prepTP(data, trans, times.names, status.names)
```

Arguments

data	Data frame in wide format in which to interpret time, status, id or keep, if appropriate.
trans	Transition matrix describing the states and transitions in the multi-state model. If S is the number of states in the multi-state model, trans should be an S x S matrix, with (i,j)-element a positive integer if a transition from i to j is possible in the multi-state model, NA otherwise. In particular, all diagonal elements should be NA. The integers indicating the possible transitions in the multi-state model should be sequentially numbered, 1,...,K, with K the number of transitions.
times.names	Either 1) a matrix or data frame of dimension n x S (n being the number of individuals and S the number of states in the multi-state model), containing the times at which the states are visited or last follow-up time, or 2) a character vector of length S containing the column names indicating these times. In the latter cases, some elements of time may be NA, see Details.
status.names	Either 1) a matrix or data frame of dimension n x S, containing, for each of the states, event indicators taking the value 1 if the state is visited or 0 if it is not (censored), or 2) a character vector of length S containing the column names indicating these status variables. In the latter cases, some elements of status may be NA, see Details.

Value

An object of class "msdata", which is a data frame in long (counting process) format containing the subject id, the covariates (replicated per subject), and from the starting state to the receiving state trans the transition number Tstart the starting time of the transition Tstop the stopping time of the transition status status variable, with 1 indicating an event (transition), 0 a censoring

Author(s)

Gustavo Soutinho and Luis Meira-Machado.

Examples

```
data("ebmt4")
db_wide <- ebmt4
positions<-list(c(2, 3, 5, 6), c(4, 5, 6), c(4, 5, 6), c(5, 6), c(6), c())
states.names = c("Tx", "Rec", "AE", "Rec+AE", "Rel", "Death")
trans<-transTP(positions, states.names)
times.names = c(NA, "rec", "ae","recae", "rel", "srv")
status.names=c(NA, "rec.s", "ae.s", "recae.s", "rel.s", "srv.s")
prepTP(data=db_wide, trans, times.names, status.names)
```

presmoothTP	*This function implements presmoothed estimates based on the landmark Aalen-Johansen method for non-parametric estimation of transition probabilities in non-Markov multi-state models.*

Description

This function is used to obtain nonparametric presmoothed estimates of the transition probabilities for any type of multistate model, including models that allow for reversible transitions, and potencially non-Markovian seetings. The methods used to compute the transition probabilities are "PLM" (given by a "logistic" model), "cauchit", "probit". The proposed presmoothed estimator is designed to handle complex multi-state models with more than three states, as well as reversible transitions, accommodating potentially non-Markovian behaviors.

Usage

```
presmoothTP(data, s, from, to, method = "LM", trans)
```

Arguments

data	A data frame in the long format containing the subject id; from corresponding to the starting state; the receiving state, to; the transition number, trans; the starting time of the transition given by Tstart; the stopping time of the transition, Tstop, and status for the status variable, with 1 indicating an event (transition), 0 a censoring.
s	The first time for obtaining estimates for the transition.
from	The starting state of the transition probabilities.
to	The last receiving state considered for the estimation of the transition probabilities.
method	The method used to compute the transition probabilities. Possible options are "PLM", "cauchit" and "probit".
trans	The transition matrix for multi-state model.

Value

Nonparametric presmoothed estimates of transition probabilities based on logistic', 'probit', and 'cauchit' distributions (PLM,cauchit and probit).

Author(s)

Gustavo Soutinho and Luis Meira-Machado.

Examples

```
data("ebmt4")
db_wide <- ebmt4
positions<-list(c(2, 3, 5, 6), c(4, 5, 6), c(4, 5, 6), c(5, 6), c(6), c())
states.names = c("Tx", "Rec", "AE", "Rec+AE", "Rel", "Death")
trans<-transTP(positions, states.names)
times.names = c(NA, "rec", "ae","recae", "rel", "srv")
status.names=c(NA, "rec.s", "ae.s", "recae.s", "rel.s", "srv.s")
data.long<-prepTP(data=db_wide, trans, times.names, status.names)

res1<-presmoothTP(data.long, s = 100, from = 1, to = 1,
method = 'PLM', trans=trans)
res2<-presmoothTP(data.long, s = 100, from = 1, to = 1, method = 'cauchit',
trans=trans)
res3<-presmoothTP(data.long, s = 100, from = 1, to = 1, method = 'probit',
trans=trans)

plot(res1, type='s', xlim=c(100,3000), ylim=c(0.6,1), ylab=c('p11(100,t)'))
lines(res2, type='s', col=2)
lines(res3, type='s', col=4)
legend("topright", legend = c("PLM","Cauchit", "Probit"), lwd = 0.3,
col = c(1,2,4), cex=0.7, lty=c(1, 1))
res4<-presmoothTP(data.long, s = 1000, from = 2, to = 6, method = 'PLM',
trans=trans)
res5<-presmoothTP(data.long, s = 1000, from = 2, to = 6, method = 'cauchit',
trans=trans)
res6<-presmoothTP(data.long, s = 1000, from = 2, to = 6, method = 'probit',
trans=trans)
plot(res4, type='s', xlim=c(1000,3000), ylim=c(0,0.12),ylab=c('p26(1000,t)'))
lines(res5, type='s', col=2)
lines(res6, type='s', col=4)
legend("topright", legend = c("PLM","Cauchit", "Probit"), lwd = 0.3,
col = c(1,2,4), cex=0.7, lty=c(1, 1))
```

---

prothr   *Liver cirrhosis data - Abnormal prothrombin levels in liver cirrhosis.*

Description

A data frame of 488 liver cirrhosis patients from a randomized clinical trial concerning prednisone treatment in these patients. The dataset is in long format. The included variables are

id Patient identification number

from Starting state

to Receiving state

trans Transition number

Tstart Starting time

Tstop Transition time

status Status variable; 1=transition, 0=censored

treat Treatment; factor with levels "Placebo", "Prednisone"

Usage

```
data("prothr")
```

Format

A data frame, see data.frame.

Author(s)

Gustavo Soutinho and Luis Meira-Machado.

Source

This data was kindly provided by Per Kragh Andersen. It was introduced in Andersen, Borgan, Gill & Keiding (1993), Example 1.3.12, and used as illustration for computation of transition probabilities in multi-state models, see Sections IV.4 (Example IV.4.4) and VII.2 (Example VII.2.10).

References

Andersen PK, Borgan O, Gill RD, Keiding N (1993). Statistical Models Based on Counting Processes. Springer, New York.

Examples

```
data(prothr)
head(prothr)
```

---

transTP	*Define transition matrix for multi-state model.*

Description

Define transition matrices for multi-state model. Specific functions for defining such transition matrices are pre-defined for common multi-state models like the competing risks model and the illness-death model.

Usage

```
transTP(positions, states.names)
```

Arguments

positions   List of possible transitions; x[[i]] consists of a vector of state numbers reachable from state i.

states.names   A character vector containing the names of either the competing risks or the states in the multi-state model specified by the competing risks or illness-death model. names should have the same length as the list x (for transMat), or either K or K+1 (for trans.comprisk), or 3 (for trans.illdeath).

Value

A transition matrix describing the states and transitions in the multi-state model.

Author(s)

Gustavo Soutinho and Luis Meira-Machado.

Examples

```
data("ebmt4")
db_wide <- ebmt4
positions<-list(c(2, 3, 5, 6), c(4, 5, 6), c(4, 5, 6), c(5, 6), c(6), c())
states.names = c("Tx", "Rec", "AE", "Rec+AE", "Rel", "Death")
transTP(positions, states.names)
```

---

unsmoothTP   *This function implements unsmoothed estimates of from the Aalen-Johansen estimator (AJ) and landmark Aalen-Johansen method of Putter & Spitoni (2016) for non-parametric estimation of transition probabilities in non-Markov multi-state models.*

Description

This function is used to obtain nonparametric unsmoothed estimates of the transition probabilities. The methods used to compute the transition probabilities are "AJ" and "LM" which can be also applied for any type of multistate model, including models that allow for reversible transitions, and potencially non-Markovian seetings.

Usage

```
unsmoothTP(data, s, from, to, method = "LM", trans = trans)
```

Arguments

data   A data frame in the long format containing the subject id; from corresponding to the starting state; the receiving state, to; the transition number, trans; the starting time of the transition given by Tstart; the stopping time of the transition, Tstop, and status for the status variable, with 1 indicating an event (transition), 0 a censoring.

s	The first time for obtaining estimates for the transition.
from	The starting state of the transition probabilities.
to	The last receiving state considered for the estimation of the transition probabilities.
method	The method used to compute the transition probabilities. Possible options are "AJ" and "LM".
trans	The transition matrix for multi-state model.

Value

Nonparametric estimates of transition probabilities in general multi-state models using the AJ and the LMAJ estimators.

Examples

```
data("ebmt4")
db_wide <- ebmt4
positions<-list(c(2, 3, 5, 6), c(4, 5, 6), c(4, 5, 6), c(5, 6), c(6), c())
states.names = c("Tx", "Rec", "AE", "Rec+AE", "Rel", "Death")
trans<-transTP(positions, states.names)
times.names = c(NA, "rec", "ae","recae", "rel", "srv")
status.names=c(NA, "rec.s", "ae.s", "recae.s", "rel.s", "srv.s")
data.long<-prepTP(data=db_wide, trans, times.names, status.names)

res1<-unsmoothTP(data.long, s = 100, from = 1, to = 1, method = 'AJ',
trans=trans)
head(res1)
res2<-unsmoothTP(data.long, s = 100, from = 1, to = 1, method = 'LM',
trans=trans)
head(res2)
plot(res1, type='s', xlim=c(0,3000), ylim=c(0.6,1), ylab=c('p11(100,t)'))
lines(res2, type='s', col=2)
legend("topright", legend = c("AJ","LM"), lwd = 0.3, col = c(1,2), cex=0.7,
lty=c(1, 1))

res3<-unsmoothTP(data.long, s = 1000, from = 2, to = 6, method = 'AJ',
trans=trans)
res4<-unsmoothTP(data.long, s = 1000, from = 2, to = 6, method = 'LM',
trans=trans)

plot(res3, type='s', xlim=c(1000,3000), ylim=c(0,0.12), ylab=c('p26(100,t)'))
lines(res4, type='s', col=2)
legend("topright", legend = c("AJ","LM"), lwd = 0.3, col = c(1,2), cex=0.7,
lty=c(1, 1))
```

Index