

Package ‘dscoreMSM’

July 22, 2025

Type Package

Title Survival Proximity Score Matching in Multi-State Survival Model

Version 0.1.0

Maintainer Atanu Bhattacharjee <atanustat@gmail.com>

Imports rjags, stats, timeROC, ggplot2, survival, mstate

Description Implements survival proximity score matching in multi-state survival models. Includes tools for simulating survival data and estimating transition-specific coxph models with frailty terms. The primary methodological work on multistate censored data modeling using propensity score matching has been published by Bhattacharjee et al.(2024) <[doi:10.1038/s41598-024-54149-y](https://doi.org/10.1038/s41598-024-54149-y)>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Depends R (>= 2.10)

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Atanu Bhattacharjee [aut, cre, ctb],
Bhriku Kumar Rajbongshi [aut, ctb],
Gajendra K Vishwakarma [aut, ctb]

Repository CRAN

Date/Publication 2024-12-13 16:40:02 UTC

Contents

cphGM	2
dscore	4
EBMTdata	5

EBMTupdate	6
expbh	7
ggplot_roc	7
ggplot_surv	9
gompbh	10
print.cphGM	11
simfdata	12
simulated_data	13
weibulbh	14
Index	15

cphGM	<i>CoxPH model with parametric baseline and frailty terms</i>
-------	---

Description

Function for estimating the parameters of coxPH model with frailty terms

Usage

```
cphGM(  
  formula,  
  fterm,  
  Time,  
  status,  
  id,  
  data,  
  bhdist,  
  method = "L-BFGS-B",  
  maxit = 200  
)
```

Arguments

formula	survival model formula like Surv(time,status)~x1+x2
fterm	frailty term like c('gamma','center'). Currently we have the option for gamma distribution.
Time	survival time column
status	survival status column
id	id column
data	dataset
bhdist	distribution of survival time at baseline. Available option 'weibull','exponential','gompertz',
method	options are 'LFGS','L-BFGS-G','CG' etc. for more details see optim
maxit	maximum number of iteration

Details

The hazard model is as follows:

$$h_i(t) = z_i h_0(t) \exp(\mathbf{x}_i \beta) ; i = 1, 2, 3, \dots, n$$

where baseline survival distribution could be Weibull distribution and the hazard function is:

$$h_0(t) = \rho \lambda t^{\rho-1}$$

. Similarly we can have Exponential, log logistic distribution. The following are the formula for hazard and cumulative hazard function For exponential: $h_0(t) = \lambda$ and $H_0(t) = \lambda t$; $\lambda > 0$ Gompertz: $h_0(t) = \lambda \exp(\gamma t)$ and $H_0(t) = \frac{\lambda}{\gamma} (\exp(\gamma t) - 1)$; $\lambda, \gamma > 0$ The frailty term z_i follows Gamma distribution with parameter θ . The parameter estimates are obtained by maximising the log likelihood

$$\prod_{i=1}^n l_i(\beta, \theta, \lambda, \rho)$$

The method argument allows the user to select suitable optimisation method available in optim function.

Value

Estimates obtained from coxph model with the frailty terms.

Author(s)

Atanu Bhattacharjee, Bhriku Kumar Rajbongshi and Gajendra K. Vishwakarma

References

Vishwakarma, G. K., Bhattacharjee, A., Rajbongshi, B. K., & Tripathy, A. (2024). Censored imputation of time to event outcome through survival proximity score method. *Journal of Computational and Applied Mathematics*, 116103;

Bhattacharjee, A., Vishwakarma, G. K., Tripathy, A., & Rajbongshi, B. K. (2024). Competing risk multistate censored data modeling by propensity score matching method. *Scientific Reports*, 14(1), 4368.

See Also

[dscore](#), [simfdata](#)

Examples

```
##
X1<-matrix(rnorm(1000*2),1000,2)
simulated_data<-simfdata(n=1000,beta=c(0.5,0.5),fvar=0.5,
X=X1)
model1<-cphGM(formula=Surv(time,status)~X1+X2,
fterm<-c('gamma','id'),Time="time",status="status",
id="id",data=simulated_data,bhdist='weibull')
model1
```

##

dscore	<i>Survival Proximity Score matching for MSM</i>
--------	--

Description

function for survival proximity score matching in multistate model with three state.

Usage

dscore(status, data, prob, m, n, method = "euclidean")

Arguments

status	status column name in the survival data
data	survival data
prob	threshold probability
m	starting column number
n	ending column number
method	distance metric name e.g. "euclidean", "minkowski", "canberra"

Value

list with newdataset updated using dscore

Author(s)

Atanu Bhattacharjee, Bhritu Kumar Rajbongshi and Gajendra K. Vishwakarma

References

Vishwakarma, G. K., Bhattacharjee, A., Rajbongshi, B. K., & Tripathy, A. (2024). Censored imputation of time to event outcome through survival proximity score method. *Journal of Computational and Applied Mathematics*, 116103;

Bhattacharjee, A., Vishwakarma, G. K., Tripathy, A., & Rajbongshi, B. K. (2024). Competing risk multistate censored data modeling by propensity score matching method. *Scientific Reports*, 14(1), 4368.

See Also

[cphGM](#), [simfdata](#)

Examples

```
##s
data(simulated_data)
udata<-dscore(status="status",data=simulated_data,prob=0.65,m=4,n=7)
##
```

EBMTdata	<i>European Bone Marrow Transplantation data obtained from mstate r package</i>
----------	---

Description

A multi state dataset

Usage

```
data(EBMTdata)
```

Format

a tibble of 13 columns and 2204 observations,

id id value for subjects

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

prstate Platelet recovery status; 1 = platelet recovery, 0 = censored

rfstime Time in days from transplantation to relapse or death or last follow-up (relapse-free survival time)

rfsstate Relapse-free survival status; 1 = relapsed or dead, 0 = censored

dissub Disease subclassification; factor with levels "AML", "ALL", "CML"

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

drmatch Donor-recipient gender match; factor with levels "No gender mismatch", "Gender mismatch"

tcd T-cell depletion; factor with levels "No TCD", "TCD"

x1,x2,x3,x4 simulated covariate information used for SPSM

Source

We acknowledge that this data set is obtained from the `r` package `mstate`. We have included four continuous covariates in the dataset to demonstrate SPSM method in multistate survival model.

References

de Wreede, L. C., Fiocco, M., & Putter, H. (2011). mstate: an R package for the analysis of competing risks and multi-state models. Journal of statistical software, 38, 1-30.

Vishwakarma, G. K., Bhattacharjee, A., Rajbongshi, B. K., & Tripathy, A. (2024). Censored imputation of time to event outcome through survival proximity score method. Journal of Computational and Applied Mathematics, 116103;

Bhattacharjee, A., Vishwakarma, G. K., Tripathy, A., & Rajbongshi, B. K. (2024). Competing risk multistate censored data modeling by propensity score matching method. Scientific Reports, 14(1), 4368.

EBMTupdate	<i>European Bone Marrow Transplantation data obtained from mstate r package. This is the updated data obtained after applying SPSM.</i>
------------	---

Description

A multi state dataset

Usage

data(EBMTupdate)

Format

- a tibble of 13 columns and 2204 observations,
- id** id value for subjects
 - prtime** Time in days from transplantation to platelet recovery or last follow-up
 - prstate** Platelet recovery status; 1 = platelet recovery, 0 = censored
 - rfstime** Time in days from transplantation to relapse or death or last follow-up (relapse-free survival time)
 - rfstate** Relapse-free survival status; 1 = relapsed or dead, 0 = censored
 - dissub** Disease subclassification; factor with levels "AML", "ALL", "CML"
 - age** Patient age at transplant; factor with levels "<=20", "20-40", ">40"
 - drmatch** Donor-recipient gender match; factor with levels "No gender mismatch", "Gender mismatch"
 - tcd** T-cell depletion; factor with levels "No TCD", "TCD"
 - x1,x2,x3,x4** simulated covariate information used for SPSM

Source

We acknowledge that this data set is obtained from the r package mstate. We have included four continuous covariates in the dataset to demonstrate SPSM method in multistate survival model.

References

de Wreede, L. C., Fiocco, M., & Putter, H. (2011). mstate: an R package for the analysis of competing risks and multi-state models. Journal of statistical software, 38, 1-30.

Vishwakarma, G. K., Bhattacharjee, A., Rajbongshi, B. K., & Tripathy, A. (2024). Censored imputation of time to event outcome through survival proximity score method. Journal of Computational and Applied Mathematics, 116103;

Bhattacharjee, A., Vishwakarma, G. K., Tripathy, A., & Rajbongshi, B. K. (2024). Competing risk multistate censored data modeling by propensity score matching method. Scientific Reports, 14(1), 4368.

expbh	<i>Exponential baseline hazard</i>
-------	------------------------------------

Description

Exponential baseline hazard

Usage

```
expbh(t, shape = 2)
```

Arguments

t	time
shape	shape parameter

Value

hazard function value under Exponential distribution

ggplot_roc	<i>Reciever Operating Curve</i>
------------	---------------------------------

Description

this function provides roc plot for coxph model fitted before and after survival proximity score matching.

Usage

```
ggplot_roc(  
  trns,  
  model1,  
  model2,  
  data1,  
  data2,  
  folder_path = NULL,  
  times = NULL  
)
```

Arguments

trns	transition number for the multistate model
model1	fitted object from coxPH (before SPSM)
model2	fitted object from coxPH (after SPSM)
data1	dataset used for model1
data2	dataset used for model2
folder_path	default is NULL. if folder_path is provided then plots will be saved there automatically.
times	default is NULL. time at which TP and FP values are calculated.

Value

returns roc plot for model1 and model2

Author(s)

Atanu Bhattacharjee, Bhritu Kumar Rajbongshi and Gajendra Kumar Vishwakarma

References

Vishwakarma, G. K., Bhattacharjee, A., Rajbongshi, B. K., & Tripathy, A. (2024). Censored imputation of time to event outcome through survival proximity score method. *Journal of Computational and Applied Mathematics*, 116103;

Bhattacharjee, A., Vishwakarma, G. K., Tripathy, A., & Rajbongshi, B. K. (2024). Competing risk multistate censored data modeling by propensity score matching method. *Scientific Reports*, 14(1), 4368.

See Also

[dscore](#), [simfdata](#), [cphGM](#)

Examples

```
##
library(mstate)
data(EBMTdata)
data(EBMTupdate)
tmat<-transMat(x=list(c(2,3),c(3),c()),
               names=c("Tx","Rec","Death"))
covs<-c("dissub","age","drmatch","tcd","prtime","x1","x2","x3","x4")
msbmt<-msprep(time=c(NA,"prtime","rfstime"),
              status=c(NA,"prstat","rfsstat"),
              data=EBMTdata,trans=tmat,keep=covs)
msbmt1<-msprep(time=c(NA,"prtime","rfstime"),
               status=c(NA,"prstat","rfsstat"),
               data=EBMTupdate,trans=tmat,keep=covs)
msph3<-coxph(Surv(time,status)~dissub+age+drmatch+tcd+
             frailty(id,distribution='gamma'),data=msbmt[msbmt$trans==3,])
msph33<-coxph(Surv(Tstart,Tstop,status)~dissub+age +drmatch+ tcd+
             frailty(id,distribution='gamma'),data=msbmt1[msbmt1$trans==3,])
ggplot_roc(trns=3,model1=msph3,model2=msph33,
           data1=msbmt,data2=msbmt1)
##
```

ggplot_surv

*Survival probability plot***Description**

it gives plot with fitted survival curve obtained from two different coxPH model fitted before and after SPSM

Usage

```
ggplot_surv(model1, model2, data1, data2, n_trans, id)
```

Arguments

model1	coxPH fitted model object (before SPSM)
model2	coxPH fitted model object (after SPSM)
data1	multistate data used in model1
data2	multistate data used in model2
n_trans	number of transition
id	particular id from the dataset

Value

plot for survival curve of a particular id obtained from both the model

Author(s)

Atanu Bhattacharjee, Bhriгу Kumar Rajbongshi and Gajendra Kumar Vishwakarma

See Also

[dscore](#), [simfdata](#), [cphGM](#)

Examples

```
##
library(mstate)
data(EBMTdata)
data(EBMTupdate)
tmat<-transMat(x=list(c(2,3),c(3),c()),names=c("Tx","Rec","Death"))
covs<-c("dissub","age","drmatch","tcd","prtime","x1","x2","x3","x4")
msbmt<-msprep(time=c(NA,"prtime","rfstime"),status=c(NA,"prstat","rfsstat"),
              data=EBMTdata,trans=tmat,keep=covs)
msbmt1<-msprep(time=c(NA,"prtime","rfstime"),status=c(NA,"prstat","rfsstat"),
              data=EBMTupdate,trans=tmat,keep=covs)
msph3<-coxph(Surv(time,status)~dissub+age+drmatch+tcd+
             frailty(id,distribution='gamma'),data=msbmt[msbmt$trans==3,])
msph33<-coxph(Surv(Tstart,Tstop,status)~dissub+age +drmatch+ tcd+
             frailty(id,distribution='gamma'),data=msbmt1[msbmt1$trans==3,])
ggplot_surv(model1=msph3,model2=msph33,data1=msbmt,
            data2=msbmt1,n_trans=3,id=1)

#####
# plot1<-ggplot_surv(model1=msph3,model2=msph33,data1=msbmt,data2=msbmt1,
# ggsave("plot1.jpg",path="C:/Users/.....")
#####
##
```

gompbh	<i>Gompartz baseline hazard</i>
--------	---------------------------------

Description

Gompartz baseline hazard

Usage

gompbh(t, shape = 2, scale = 1)

Arguments

t	time
shape	shape parameter
scale	scale parameter

Value

hazard function value under Gompertz distribution

print.cphGM	<i>print function for cphGM</i>
-------------	---------------------------------

Description

S3 print method for class 'cphGM'

Usage

```
## S3 method for class 'cphGM'
print(x, ...)
```

Arguments

x	object
...	others

Value

prints table containing various parameter estimates, SE, P-value.

Examples

```
##
n1<-1000
p1<-2
X1<-matrix(rnorm(n1*p1),n1,p1)
simulated_data<-simfdata(n=1000,beta=c(0.5,0.5),fvar=0.5,X=X1)
model1<-cphGM(formula=Surv(time,status)~X1+X2,
fterm=c('gamma','id'),Time="time",status="status",
id="id",data=simulated_data,bhdist='weibull')
print(model1)
##
```

simfdata	<i>simulation of survival data</i>
----------	------------------------------------

Description

function for simulation of survival data assuming the data comes from a parametric coxph model with gamma frailty distribution

Usage

```
simfdata(n, beta, fvar, bhdist = "weibull", X, fdist = "gamma", ...)
```

Arguments

n	number of individual
beta	vector of regression coefficient for coxph model
fvar	frailty variance value(currently the function works for gamma frailty only)
bhdist	distribution of survival time at baseline e.g. "weibull","exponential","llogistic"
X	model matrix for the coxPH model with particular choice of beta
fdist	distribution of frailty terms e.g. "gamma"
...	user can assume the shape and scale parameter of baseline survival distribution

Details

The process for simulation of multistate survival data is described in our manuscript. As the process includes transition through different states and it involves simulating survival time in different transition. So we have demonstrated the code for simulation of simple survival model. Suppose we want to simulate a survival data with parametric baseline hazard and parametric frailty model. The hazard model is as follows:

$$h_i(t) = z_i h_0(t) \exp(\mathbf{x}_i \beta) ; i = 1, 2, 3, \dots, n$$

where the baseline survival time follow Weibull distribution and the hazard is

$$h_0(t) = \rho \lambda t^{\rho-1}$$

. Similarly we can have Gompertz, log logistic distribution. The following are the formula for hazard and cumulative hazard function For exponential: $h_0(t) = \lambda$ and $H_0(t) = \lambda t; \lambda > 0$
Gompertz: $h_0(t) = \lambda \exp(\gamma t)$ and $H_0(t) = \frac{\lambda}{\gamma} (\exp(\gamma t) - 1); \lambda, \gamma > 0$

Value

simulated survival data for a single transition

Author(s)

Atanu Bhattacharjee, Bhriku Kumar Rajbongshi and Gajendra K. Vishwakarma

References

Vishwakarma, G. K., Bhattacharjee, A., Rajbongshi, B. K., & Tripathy, A. (2024). Censored imputation of time to event outcome through survival proximity score method. *Journal of Computational and Applied Mathematics*, 116103;

Bhattacharjee, A., Vishwakarma, G. K., Tripathy, A., & Rajbongshi, B. K. (2024). Competing risk multistate censored data modeling by propensity score matching method. *Scientific Reports*, 14(1), 4368.

See Also

[cphGM](#)

Examples

```
##
n1<-1000
p1<-2
X1<-matrix(rnorm(n1*p1),n1,p1)
simulated_data<-simfdata(n=1000,beta=c(0.5,0.5),fvar=0.5,
X=X1)
##
```

simulated_data	<i>Simulated multistate data</i>
----------------	----------------------------------

Description

A simulated multi state dataset used for demonstration purpose.

Usage

```
data(simulated_data)
```

Format

- a tibble of 13 columns and 2204 observations,
- id** id value for subjects
 - status** survival status
 - time** survival time
 - x1** Numeric covariate
 - x2** Numeric covariate
 - x3** Numeric covariate
 - x4** Numeric covariate

References

Vishwakarma, G. K., Bhattacharjee, A., Rajbongshi, B. K., & Tripathy, A. (2024). Censored imputation of time to event outcome through survival proximity score method. *Journal of Computational and Applied Mathematics*, 116103;

Bhattacharjee, A., Vishwakarma, G. K., Tripathy, A., & Rajbongshi, B. K. (2024). Competing risk multistate censored data modeling by propensity score matching method. *Scientific Reports*, 14(1), 4368.

weibulbh	<i>Weibull baseline hazard</i>
----------	--------------------------------

Description

Weibull baseline hazard

Usage

```
weibulbh(t, shape = 2, scale = 1)
```

Arguments

t	time
shape	shape parameter
scale	scale parameter

Value

hazard function value under Weibull distibution

Index

* **datasets**

EBMTdata, [5](#)

EBMTupdate, [6](#)

simulated_data, [13](#)

cphGM, [2](#), [4](#), [8](#), [10](#), [13](#)

dscore, [3](#), [4](#), [8](#), [10](#)

EBMTdata, [5](#)

EBMTupdate, [6](#)

expbh, [7](#)

ggplot_roc, [7](#)

ggplot_surv, [9](#)

gompbh, [10](#)

optim, [2](#)

print.cphGM, [11](#)

simfdata, [3](#), [4](#), [8](#), [10](#), [12](#)

simulated_data, [13](#)

weibulbh, [14](#)