Package 'survMCOD'

April 15, 2017

Version 0.1.0

Title Survival analysis with multiple causes of death

Depends R (>= 3.1.0), survival

Imports Matrix

Description
Fits Cox regression models for the hazard of death due to a disease of interest based on multiple cause-of-
death data, thus acknowledging that death may be caused by several disease processes acting concurrently. This is based on the model for multiple-cause mortality proposed by Moreno-Betancur et al.
(2017) which extends the single-cause competing risks model to that setting.

License GPL (>= 3)

LazyData TRUE

RoxygenNote 6.0.1

Author Margarita Moreno-Betancur [cre, aut]

Maintainer Margarita Moreno-Betancur <margarita.moreno@mcri.edu.au>

URL https://github.com/moreno-betancur/survMCOD

BugReports https://github.com/moreno-betancur/survMCOD/issues

R topics documented:

`check.survMCOD` *Produce plots to check convergence of multiple-cause model*

Description

Produce plots to check convergence of multiple-cause model

Usage

```
check.survMCOD(fit)
```

Arguments

fit	A model fit returned by survMCOD.

Details

The plots produced show Iter+2 estimates of each parameter, where Iter is the number of iterations specified by the user in the call to survMCOD. The first two estimates correspond to a starting value and an improved starting value (see Moreno-Betancur et al. 2017 for details). Healthy convergence is seen by curves showing variation in estimates across the first three points, followed by a stabilisation of the curve around the final estimate.

References

Moreno-Betancur M, Sadaoui H, Piffaretti C, Rey G. Survival analysis with multiple causes of death: Extending the competing risks model. Epidemiology 2017; 28(1): 12-19.

Examples

```
## Example ## uncomment to run

# First we simulate data using the simMCOD function:
# datEx<-simMCOD(n=1000,xi=-1,rho=-2,phi=0,
#       pgen=c(1,0,0.75,0.25,0.125,0.083),
#      lambda=0.001,v=2,pUC=c(1,0.75))

# Run analysis

# fitMCOD<-survMCOD(SurvM(time=TimeEntry,time2=TimeExit,status=Status,
#                         weight=Pi)~X1,
#                   formOther=~Z1,data=datEx,UC_indicator="UC")

# Check convergence of multiple-cause analysis
# check.survMCOD(fitMCOD)
```

---

simMCOD *Simulate survival data with multiple causes of death*

Description

simMCOD simulates data from the multiple-cause model of Moreno-Betancur et al. (2017), specifically under the same assumptions as in their simulation study.

Usage

```
simMCOD(n, xi, rho, phi, pgen, lambda, v, pUC)
```

Arguments

n	Size of dataset to be generated
xi	Log-ratio of the pure baseline hazards, assumed constant
rho	Effect of the binary exposure X1(=Z1) on the pure hazard of the disease of interest
phi	Effect of the binary exposure Z1(=X1) on the pure hazard of other diseases
pgen	Vector of weights corresponding to each of the 6 states in the multistate model.
lambda	Weibull scale parameter used to specify the pure baseline hazard of the disease of interest.
v	Weibull shape parameter used to specify the pure baseline hazard of the disease of interest.
pUC	Vector indicating the values in pgen corresponding to weight values assigned to the underlying cause of death.

Details

The function can be used to generate data from any of the scenarios considered in the the main simulation settings of Moreno-Betancur et al. (2017). See that reference for details.

Value

A data.frame as required by survMCOD function. Specifically, a dataset with one row per individual and the following variables:

X1 Binary exposure (to include in model for disease of interest).

Z1 Binary exposure equal to X1 (to include in model for other diseases).

TimeEntry Time of entry into the study

TimeExit Time of exit from the study

Status Indicator of vital status, with Status=1 if individual died and Status=0 if individual is censored.

Pi Weight indicating the proportion of the death attributed to the disease of interest (missing, i.e. "NA", if Status=0).

UC The Underlying Cause Indicator, which is 1 if the individual died and the disease of interest was selected as underlying cause of death and 0 otherwise (i.e. 0 if individual censored or if individual died but the disease of interest was not selected as underlying cause of death).

References

Moreno-Betancur M, Sadaoui H, Piffaretti C, Rey G. Survival analysis with multiple causes of death: Extending the competing risks model. Epidemiology 2017; 28(1): 12-19.

Examples

```
datEx<-simMCOD(n=1000,xi=-1,rho=-2,phi=0,
        pgen=c(1,0,0.75,0.25,0.125,0.083),
        lambda=0.001,v=2,pUC=c(1,0.75))

head(datEx)
```

SurvM *Create a multiple-cause survival object*

Description

SurvM Creates a multiple-cause survival object, which is used as a response variable in the model formula provided to survMCOD.

Usage

```
SurvM(time, time2, status, weight, type = c("right", "counting"))
```

Arguments

time	For right censored data, this is the follow up time. For counting process data, the first argument is the entry time.
time2	For counting process data only, it is the exit time.
status	The vital status indicator such that 0=alive and 1=dead.
weight	The weight indicating the proportion of the death attributed to the disease of interest. It should be NA if individual alive (status=0) and a number between 0 and 1 if dead (status=1).
type	Character string specifying the type of censoring. Possible values are "right" and "counting".

Details

This function needs to be used to in the formula argument of the survMCOD function.

The user is referred to Moreno-Betancur et al. (2017), Piffaretti et al. (2016) and Rey et al. (2017) for descriptions and discussions of various weight-attribution strategies to determine the weight argument.

Value

An object of class SurvM.

References

Moreno-Betancur M, Sadaoui H, Piffaretti C, Rey G. Survival analysis with multiple causes of death: Extending the competing risks model. Epidemiology 2017; 28(1): 12-19.

Piffaretti C, Moreno-Betancur M, Lamarche-Vadel A, Rey G. Quantifying cause-related mortality by weighting multiple causes of death. Bulletin of the World Health Organization 2016; 94:870-879B.

Rey G, Piffaretti C, Rondet C, Lamarche-Vadel A, Moreno-Betancur M. Analyse de la mortalite par cause : ponderation des causes multiples. Bulletin Epidemiologique Hebdomadaire, 2017; (1): 13-9.

Examples

```
datEx<-simMCOD(n=1000,xi=-1,rho=-2,phi=0,pgen=c(1,0,0.75,0.25,0.125,0.083),
        lambda=0.001,v=2,pUC=c(1,0.75))

SurvM(time=datEx$TimeEntry, time2=datEx$TimeExit, status=datEx$Status,
weight=datEx$Pi, type="counting")
```

`survMCOD` *Fit Cox regression models for multiple cause-of-death data*

Description

`survMCOD` fits Cox regression models for the pure hazard of death due to a disease of interest based on multiple cause-of-death data ("Multiple-cause analysis"), thus acknowledging that death may be caused by several disease processes acting concurrently. The pure hazard is the rate of deaths caused exclusively by the disease of interest, and is thus a quantity that is conceptually closer to the marginal "causal" hazard than the cause-specific hazard. The latter is the quantity modeled when using competing risks Cox regression based on the so-called "underlying cause of death" and ignoring all other diseases mentioned on the death certificate ("Single-cause analysis").

Usage

```
survMCOD(formula, formOther = formula[-2], data, UC_indicator, Iter = 4)
```

Arguments

`formula`	A formula object with the response on the left of the ~ operator being an object as returned by the `SurvM` function (see **Examples** section below). The terms on the right are the regressors to include in the the model for the pure hazard of the disease of interest.
`formOther`	A formula object with empty response, i.e. nothing on the left of the ~ operator (if a response is given it is ignored). The terms on the right are the regressors to include in the the model for the pure hazard of other diseases. If unspecified, this model will include the same regressors as the model for the disease of interest as specified in `formula` above.
`data`	A data.frame in which to interpret the variables named in `formula` and `formOther` above. Specifically, the dataset must contain one row per individual and the following variables: - The time-to-event (for type="right") or the entry and exit times (for type="counting") - A status variable indicating whether the individual died (=1) or is censored (=0) - A weight variable which is missing ("NA") if the individual is censored, and a proportion between 0 and 1 if the individual died. The weight corresponds to the proportion of that death that is attributed to the disease of interest according to the chosen weight attribution strategy (see **Details** section below). - All the regressors named in `formula` and `formOther`. - The Underlying Cause Indicator, specified in `UC_indicator`, which is 1 if the individual died and the disease of interest was selected as underlying cause of death and 0 otherwise (i.e. 0 if individual censored or if individual died but the disease of interest was not selected as underlying cause of death).

`UC_indicator` Name of the Underlying Cause Indicator variable in the dataset.

`Iter` Number of iterations for iteration procedure. Default is 4 which is generally enough to achieve convergence. See **Value** below for more details.

Details

The `survMCOD` function can be used to fit Cox regression models for the pure hazard of death due to a disease of interest based on multiple cause-of-death data. In addition to results from the multiple-cause analysis, the function also provides the results from the single-cause analysis (i.e. from a competing risks Cox regression based on the so-called "underlying cause of death").

The key preliminary step to using this function to perform the multiple-cause analysis is to assign a weight to each death that represents the proportion of the death attributed to the disease of interest. The user is referred to Moreno-Betancur et al. (2017), Piffaretti et al. (2016) and Rey et al. (2017) for descriptions and discussions of various weight-attribution strategies.

The assumptions of the multiple-cause model and details of the estimation procedure are provided in Moreno-Betancur et al. (2017). A key feature of the model is that regression coefficients need to be estimated simultaneaously for a Cox model for the disease of interest and a Cox model for other causes, and deaths with a weight between 0 and 1 will contribute to both. This is why the user needs to specify regressors for each of these models using the two arguments `formula` and `formOther`.

Another aspect of the multiple-cause model is that a fully parametric model for the log ratio of the baseline pure hazards needs to be posited. The current default is to parametrise this a piecewise constant function with cut-offs at the 25th, 50th and 75th percentile of the the user control over this.

Convergence of the multiple-cause model fitting procedure should be checked using `check.survMCOD`.

Value

This development version returns a list with three components. First, a list with the results of the multiple-cause analysis. These are estimates of log hazard ratios for the models for the disease of interest and other diseases, and estimates of the piecewise constant log ratio of the baseline pure hazards. Second, a list with the single-cause analysis log hazard ratio estimates for the disease of interest and other diseases. All estimates are accompanied with their corresponding standard errors, 95% confidence intervals and p-values. This will change in future versions when a proper class of objects and summary and other such methods are developed. Third, a data.frame with Iter+2 estimates of each parameter, the first two corresponding to a starting value and an improved starting value (see Moreno-Betancur et al. 2017 for details). This data.frame is used by function `check.survMCOD`, which enables the user to check convergence (type ?check.survMCOD for details).

References

Moreno-Betancur M, Sadaoui H, Piffaretti C, Rey G. Survival analysis with multiple causes of death: Extending the competing risks model. Epidemiology 2017; 28(1): 12-19.

Piffaretti C, Moreno-Betancur M, Lamarche-Vadel A, Rey G. Quantifying cause-related mortality by weighting multiple causes of death. Bulletin of the World Health Organization 2016; 94:870-879B.

Rey G, Piffaretti C, Rondet C, Lamarche-Vadel A, Moreno-Betancur M. Analyse de la mortalite par cause : ponderation des causes multiples. Bulletin Epidemiologique Hebdomadaire, 2017; (1): 13-9.

Examples

```
## Example ##

# First we simulate data using the simMCOD function:
datEx<-simMCOD(n=1000,xi=-1,rho=-2,phi=0,
       pgen=c(1,0,0.75,0.25,0.125,0.083),
       lambda=0.001,v=2,pUC=c(1,0.75))

# Run analysis

fitMCOD<-survMCOD(SurvM(time=TimeEntry,time2=TimeExit,status=Status,
                        weight=Pi)~X1,
                  formOther=~Z1,data=datEx,UC_indicator="UC")

# Multiple-cause analysis results
 fitMCOD[[1]]

# Single-cause analysis results
 fitMCOD[[2]]

# Check convergence of multiple-cause analysis
  check.survMCOD(fitMCOD)
```

Index