Package 'survtd'

October 17, 2017

Version 0.0.1

Title Survival analysis with time-dependent covariates

Depends R (>= 3.1.0), lme4, survival, timereg, mice

Imports boot, ipw, stringr, MASS, Matrix

Description Fit semi-parametric Cox or additive hazards regression models with time-fixed covariates of any type and multiple continuous time-dependent covariates subject to missing data, measurement error or simply observed in discrete time: all these issues are handled with the two-stage Multiple Imputation for Joint Modeling (MIJM) approach developed by Moreno-Betancur et al. (2017). The package also provides easy-to-use implementations of an unadapted version (unMIJM); a simple two-stage approach (simple2S), based on single imputation of the continuous markers from linear mixed models; and the last observation carried forward (LOCF) approach for time-dependent covariates of any type, which is the traditional method to incoporate these in hazard models. Data simulation functions also included.

License GPL (>= 3)

LazyData TRUE

RoxygenNote 6.0.1

Author Margarita Moreno-Betancur [cre, aut],
Samuel L Brilleman [ctb]

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

R topics documented:

`simjm` *Simulate data from a joint model of multiple continuous time-dependent markers and a time-to-event*

Description

simjm simulates data from a Cox proportional hazards or semi-parametric additive model with two time-fixed covariates (Z1 and Z2) and three time-dependent covariates (Yij_1,Yij_2,Yij_3). The user can specify various characteristics of these distributions.

Usage

```
simjm(n = 200, surv_model = "Cox", marker_model = "RE", MErr = "High",
  Miss = "Low", effects = "Weak", corr = "Low")
```

Arguments

n	Size of dataset to be generated
surv_model	Model for the time-to-event. Options are "Cox" for Cox proportional hazard model and "Add" for semi-parametric additive hazards model.
marker_model	Model for the multiple markers. Options are "RE" for the correlated random effects model and "PN" for the product normal model.
MErr	Degree of measurement error in the multiple markers. Options are "Low", "Mod" and "High".
Miss	Degree of missing at-risk measurements. Options are "None", "Low" and "High".
effects	Strength of effects of the markers on the hazard. Options are "Null", "Weak" and "Strong".
corr	Degree of marginal pairwise correlations between the markers. Options are "Low", "Mod" and "High".

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 survtd. That is, in the long format, with one row per individual and per visit time at which any of the time-dependent covariates were measured, with the corresponding measurements. The dataset also includes a variable that uniquely identifies observations from the same individual; a variable that indicates the timing of each measurement visit; and the fixed variables (time-to-event, event indicator, time-fixed covariates) which are constant across rows of the same individual. The final columns of the dataset (from Xij_1 onwards) are to recover the true values of the markers as per the data generation model for use with function simjm_benchmark. Specifically, the variables in the dataset are:

ID Unique identifier of observations from the same individual.

tt Time to event, possibly right-censored.

event Indicator of event, with event=1 if an event occurred at tt and event=0 if the individual is censored.

Z1 Time-fixed binary covariate.

Z2 Time-fixed continuous covariate.

tj Timing of the measurement visit.

Yij_1 Measured value of marker 1 at time tj

Yij_2 Measured value of marker 2 at time tj

Yij_3 Measured value of marker 3 at time tj

Xij_1 True value of marker 1 at time tj

Xij_2 True value of marker 2 at time tj

Xij_3 True value of marker 3 at time tj

fixed_1 Time-fixed part of the linear predictor of the linear mixed model from which Yij_1 is generated.

tim_1 Time-dependent part of the linear predictor of the linear mixed model from which Yij_1 is generated, excluding terms for other markers in the case of product-normal model.

fixed_2 Time-fixed part of the linear predictor of the linear mixed model from which Yij_2 is generated.

tim_2 Time-dependent part of the linear predictor of the linear mixed model from which Yij_2 is generated, excluding terms for other markers in the case of product-normal model.

fixed_3 Time-fixed part of the linear predictor of the linear mixed model from which Yij_3 is generated.

tim_3 Time-dependent part of the linear predictor of the linear mixed model from which Yij_3 is generated, excluding terms for other markers in the case of product-normal model.

References

Moreno-Betancur M, Carlin JB, Brilleman SL, Tanamas S, Peeters A, Wolfe R (2017). Survival analysis with time-dependent covariates subject to missing data or measurement error: Multiple Imputation for Joint Modeling (MIJM). *Biostatistics* [Epub ahead of print 12 Oct 2017].

Examples

```
dat<-simjm(n=200,surv_model="Cox",marker_model="PN",
           MErr="High",Miss="None",effects="Weak",corr="Low")

head(dat)

dat<-simjm(n=200,surv_model="Add",marker_model="RE",
          MErr="Low",Miss="High",effects="Strong",corr="Mod")

head(dat)
```

`simjm_benchmark` *Benchmark "True" analysis of data simulated by* `simjm` *function*

Description

`simjm_benchmark` fits Cox proportional hazards and semi-parametric additive hazards models to data generated by `simjm` using the perfect data, i.e. the true values of the time-dependent markers.

Usage

```
simjm_benchmark(data, surv_model = "Cox", marker_model = "RE",
  corr = "Low")
```

Arguments

`data` A data.frame produced by `simjm`.

`surv_model` Time-to-event model that was used to generate `data`. Options are "Cox" for Cox proportional hazard model and "Add" for semi-parametric additive hazards model.

`marker_model` Multiple marker model that was used to generate `data`. Options are "RE" for the correlated random effects model and "PN" for the product normal model.

`corr` Degree of marginal pairwise correlations between the markers that was used to generate `data`. Options are "Low", "Mod" and "High".

Details

The function is used to perform what is referred to as the "True" analysis in Moreno-Betancur et al. (2017) which is based on perfect data, i.e. the true values of the markers at each of the event times as drawn from the simulation model. This provides a benchmark for analyses performed with error-polluted and incomplete data using `survtd`.

Value

Returns regression coefficient estimates for each covariate based on the "True" analysis, along with 95 confidence intervals and p-values.

References

Moreno-Betancur M, Carlin JB, Brilleman SL, Tanamas S, Peeters A, Wolfe R (2017). Survival analysis with time-dependent covariates subject to missing data or measurement error: Multiple Imputation for Joint Modeling (MIJM). *Biostatistics* [Epub ahead of print 12 Oct 2017].

Examples

```
dat<-simjm(n=200,surv_model="Cox",marker_model="PN",
          MErr="High",Miss="None",effects="Weak",corr="Low")

simjm_benchmark(dat,surv_model="Cox",marker_model="PN",corr="Low")
```

`survtd` *Fit survival hazard models with time-dependent covariates*

Description

`survtd` fits semi-parametric Cox proportional hazards or additive hazards models with time-fixed covariates of any type and time-dependent covariates with either of these approaches: Multiple Imputation for Joint Modeling (MIJM);Unadapted version of that approach (unMIJM); Simple two-stage approach (simple2S); Last observation carried forward approach (LOCF).

Usage

```
survtd(formula, data, id, visit.time, model = "Cox", method = "MIJM",
  M = 5, G = 5, time.trend = as.formula("~x"))
```

Arguments

`formula` A formula object, with the response on the left of a ~ operator, and the terms on the right as regressors. The response must be a survival object as returned by the Surv function of type "right" (other types are not supported at this stage). Time-dependent regressors are specified by the wrapper `td()` (see **Examples** section below). This development version does not support interactions nor terms constructed using I(). The user thus needs to create the necessary interaction/derived variables in the dataset before using this function.

`data` A data.frame in which to interpret the variables named in the formula. The dataset needs to be in long format, with one row per individual and per visit time at which any of the time-dependent covariates were measured, with the corresponding measurements. The dataset must also include a variable that uniquely identifies observations from the same individual (see parameter `id` below); a variable that indicates the timing of each measurement visit (see parameter `visit.time` below); and other fixed variables (time-to-event, event indicator, time-fixed covariates) which are constant across rows of the same individual.

`id` The name of the variable in `data` that uniquely identifies observations from the same individual.

`visit.time` The name of the variable in `data` that indicates the timing of each visit.

`model` Indicates which hazard model to fit. Options are "Cox" for the Cox proportional hazard model and "Add" for the semi-parametric additive hazards model.

`method` Indicates which method to use to fit the model. Options are "MIJM", "unMIJM", "LOCF" and "simple2S". Both "MIJM" and "unMIJM" use a two-stage joint modeling approach based on multiple imputation to incorporate time-varying continuous covariates (see **Details** section below). Method "LOCF" uses the Last Observation Carried Forward (LOCF) approach. Method "simple2S" is a simple two-stage approach (see **Details** section below).

`M` Number of imputations to perform for methods "MIJM" and "unMIJM" (ignored if method="LOCF" or "simple2S").

`G` Number of iterations to perform in multiple imputation by chained equations (MICE) algorithm for methods "MIJM" and "unMIJM" (ignored if method is "LOCF" or "simple2S").

`time.trend` Formula object with empty left hand side and right hand side expressing a polynomial of "x" which determines the way time is modeled in the fixed effects part of the linear mixed model for each time-depdent marker (ignored if method="LOCF"). Default is a linear trend (i.e. include `visit.time` as predictor in the model). The random effects part includes a random intercept and slope. Future development plans are to allow a different `time.trend` argument for each marker, the possibility to include natural cubic splines in this argument and more general random effects structures.

Details

The `survtd` function can be used to fit the Cox proportional hazards or semi-parametric additive hazard models with time-depedent covariates.

Methods "MIJM" and "unMIJM" can be used to fit models with time-fixed covariates of any type and multiple continuous time-dependent covariates. To deal with the discrete-time observation of the time-varying covariates, particularly measurement error and missing data, a two-stage joint modeling approach is used that is based on multiple imputation. Details are provided in Moreno-Betancur et al. (2017), but briefly, in Stage 1 the true (error-corrected) values of the time-dependent covariates at each event time at which an individual is at risk are multiply imputed by drawing iteratively from smoothed trajectories based on interdependent linear mixed models using Multiple Imputation by Chained Equations (MICE). An adaptation of the mice function from the *mice* package is used for this step, largely based on the procedure developed by Moreno-Betancur and Chavance (2016). In Stage 2, the time-to-event model is fitted to each of the imputed datasets (see below) and estimates are pooled using Rubin's MI formulas.

The two methods "MIJM" and "unMIJM" differ in the way information about the event occurrence is included in the imputation models: the "unMIJM" method includes the the event indicator, while the "MIJM" approach is based on a more refined approximation including a modified event indicator (see Moreno-Betancur et al. 2017). Hence, the approach "MIJM" is generally preferable.

The "simple2S" method can be used to fit models with time-fixed covariates of any type and multiple continuous time-dependent covariates using a simple two-stage approach. This method singly imputes each continuous marker at each event time from its estimated trajectory obtained from a linear mixed model including fixed and random effects for time (see description of time.trend argument) and fixed effects for the time-fixed covariates appearing in the formula argument. The method thus ignores: the uncertainty in these imputed values, the interrelations between the time-dependent markers and the relation between the time-dependent markers and the time-to-event process (see Moreno-Betancur et al. 2017).

The "LOCF" method can be used with any type of time-fixed and time-varying covariates (categorical or continuous). This approach uses the last available measurement of the time-varying covariate to singly impute its value at each event time at which an individual is at risk. It can perform very poorly if the observations are not synchronously updated across individuals (e.g. due to missing data) and if the time-varying covariates are measured with error (see Moreno-Betancur et al. 2017).

Once the values of the time-dependent covariates are imputed at each of the event-times at which the individual is at risk according to either of the four methods, the function uses coxph from the *survival* package to fit the Cox model and aalen from the *timereg* package to fit the additive model.

Value

This development version returns a data frame with regression coefficient estimates for each covariate based on the method chosen, along with 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.

References

Moreno-Betancur M, Carlin JB, Brilleman SL, Tanamas S, Peeters A, Wolfe R (2017). Survival analysis with time-dependent covariates subject to missing data or measurement error: Multiple Imputation for Joint Modeling (MIJM). *Biostatistics* [Epub ahead of print 12 Oct 2017].

Moreno-Betancur M., Chavance M. (2016) Sensitivity analysis of incomplete longitudinal data departing from the missing at random assumption: Methodology and application in a clinical trial with drop-outs. *Statistical methods in medical research*, 25 (4), 1471-1489 [Epub ahead of print, May 22 2013]

Examples

```
## Example with additive model ##

dat<-simjm(n=200,surv_model="Add",marker_model="RE",
          MErr="Low",Miss="High",effects="Strong",corr="Mod")
```

```
survtd(Surv(time=tt,event=event)~Z1+Z2+td(Yij_1)+td(Yij_2)+td(Yij_3),
       data=dat,  id="ID", visit.time="tj", model="Add",
       method="MIJM", M=5, G=5,time.trend=as.formula("~x+I(x^2)"))

survtd(Surv(time=tt,event=event)~Z1+Z2+td(Yij_1)+td(Yij_2)+td(Yij_3),
       data=dat,  id="ID", visit.time="tj", model="Add",
       method="simple2S", M=5, G=5,time.trend=as.formula("~x+I(x^2)"))

survtd(Surv(time=tt,event=event)~Z1+Z2+td(Yij_1)+td(Yij_2)+td(Yij_3),
       data=dat,  id="ID", visit.time="tj", model="Add",
       method="LOCF", M=5, G=5)

simjm_benchmark(dat,surv_model="Add",marker_model="RE",corr="Mod")
```

Index