Package ‘PenCoxFrail’

May 7, 2016

Type Package

Title Regularization in Cox Frailty Models

Version 1.0.1

Date 2016-05-06

Author Andreas Groll

Maintainer Andreas Groll <groll@mathematik.uni-muenchen.de>

Description A regularization approach for Cox Frailty Models by penalization methods is provided.

License GPL (>= 2)

Imports Rcpp (>= 0.11.6), Matrix

Depends survival

LinkingTo Rcpp, RcppArmadillo

Repository CRAN

SystemRequirements C++11

NeedsCompilation yes

Date/Publication 2016-05-07 01:06:13

R topics documented:

bs.design *Generate a B-spline design matrix*

Description

The function generates a B-spline design matrix with equidistant knots for given degree of the splines and number of basis functions.

Usage

```
bs.design(x, xl, xr, spline.degree, nbasis, comp = NULL)
```

Arguments

`x`	the positions where spline to be evaluated.
`xl`	lower intervall boundary where spline functions are relevant.
`xr`	upper intervall boundary where spline functions are relevant.
`spline.degree`	(polynomial) degree of the B-splines.
`nbasis`	number of basis functions used.
`comp`	Specify if only specific columns of the B-spline design matrix should be returned. Default is NULL and the whole B-spline design matrix is returned.

Value

The B-spline design matrix is returned.

Author(s)

Andreas Groll <`groll@math.lmu.de`>

See Also

`pencoxfrail`

Examples

```
x <- rnorm(100)
B <- bs.design(x=x, xl=min(x), xr=max(x), spline.degree=3, nbasis=5)
```

int.approx *Approximation of a Cox likelihood intergral*

Description

The function approximates the integral $\int_0^t exp(uB(s)\alpha)ds)$ which appears in the (full) Cox likelihood if the covariate u has a time-varying effect $\beta(t)$, which is expanded in B-splines, i.e. $\beta(t) = B(t)\alpha$.

Usage

```
int.approx(z,time.grid,B,nbasis,alpha)
```

Arguments

`z`	a vector which contains at the first component a time point up to which it should be integrated and the covariates u in the remaining components.
`time.grid`	an equally-spaced time grid on which the B-spline design matrix B has been generated. The maximal value of the time grid should usually be the maximal upper integral border that is of interest.
`B`	a B-spline design matrix, which has been created with the function `bs.design` on the full time grid `time.grid`.
`nbasis`	number of basis functions used when the B-spline design matrix B has been generated.
`alpha`	vector of B-spline coefficients.

Value

The B-spline design matrix is returned.

Author(s)

Andreas Groll <`groll@math.lmu.de`>

See Also

`pencoxfrail`

Examples

```
## generate time grid and corresponding B-spline design matrix
time.grid <- seq(0,200,by=1)
B <- bs.design(x=time.grid, xl=min(time.grid), xr=max(time.grid), spline.degree=3, nbasis=5)

## specify spline coefficients and covariate vector (with upper integral bound as first component)
alpha <- c(0.1,0.2,0.05,0.1,0.15)
z <- c(time=100,age=25)
```

```
## calculate intergal from 0 to 100
int.approx(z=z,time.grid=time.grid,B=B,nbasis=5,alpha=alpha)
```

pencoxfrail *Regularization in Cox Frailty Models.*

Description

A regularization approach for Cox Frailty Models by penalization methods is provided.

Usage

```
pencoxfrail(fix=formula, rnd=formula, vary.coef=formula, data, xi,
              adaptive.weights = NULL, control = list())
```

Arguments

`fix`	a two-sided linear formula object describing the unpenalized fixed (time-constant) effects part of the model, with the response on the left of a ~ operator and the terms, separated by + operators, on the right. The response must be a survival object as returned by the `Surv` function.
`rnd`	a two-sided linear formula object describing the random-effects part of the model, with the grouping factor on the left of a ~ operator and the random terms, separated by + operators, on the right.
`vary.coef`	a one-sided linear formula object describing the time-varying effects part of the model, with the time-varying terms, separated by + operators, on the right side of a ~ operator.
`data`	the data frame containing the variables named in the three preceding `formula` arguments.
`xi`	the overall penalty parameter that controls the strenght of both penalty terms in $\xi \cdot J(\zeta, \alpha)$ and, hence, controls the overall amount of smoothness (up to constant effects) and variable selection for a given proportion ζ. The optimal penalty parameter is a tuning parameter of the procedure that has to be determined, e.g. by K-fold cross validation. (See details or the quick demo for an example.)
`adaptive.weights`	a two-column matrix of adaptive weights passed to the procedure; the first column contains the weights $w_{\Delta,k}$, the second column the weights v_k from $\xi \cdot J(\zeta, \alpha)$. If no adaptive weights are specified all weights are set to one. The recommended strategy is to first fit an unpenalized model (i.e. $\xi = 0$) and then use the obtained adaptive weights (see value section) when fitting the model for all other combinations of ξ and ζ.
`control`	a list of control values for the estimation algorithm to replace the default values returned by the function `pencoxfrailControl`. Defaults to an empty list.

Details

The `pencoxfrail` algorithm is designed to investigate the effect structure in the Cox frailty model, which is a widely used model that accounts for heterogeneity in survival data. Since in survival models one has to account for possible variation of the effect strength over time the selection of the relevant features distinguishes between the folllowing cases: covariates can have time-varying effects, can have time-constant effects or be irrelevant. For this purpose, the following specific penality is applied on the vectors of B-spline coefficients α_k, assuming $k = 1, ..., r$ different, potentially time-varying effects, each expanded in M B-spline basis functions:

$$\xi \cdot J(\zeta, \alpha) = \xi \left(\zeta \sum_{k=1}^{r} \psi \, w_{\Delta,k} ||\Delta_M \alpha_k||_2 + (1-\zeta) \sum_{k=1}^{r} \phi \, v_k ||\alpha_k||_2 \right)$$

This penalty is able to distinguish between these types of effects to obtain a sparse representation that includes the relevant effects in a proper form.

The penalty is depending on two tuning parameters, ξ and ζ, which have to be determined by a suitable technique, e.g. by (2-dimensional) K-fold cross validation.

The first term of the penalty controls the smoothness of the time-varying covariate effects, whereby for values of ξ and ζ large enough, all differences $(\alpha_{k,l} - \alpha_{k,l-1}), l = 2, ..., M$, are removed from the model, resulting in constant covariate effects. As the B-splines of each variable with varying coefficients sum up to one, a constant effect is obtained if all spline coefficients are set equal. Hence, the first penalty term does not affect the spline's global level. The second term penalizes all spline coefficients belonging to a single time-varying effect in the way of a group LASSO and, hence, controls the selection of covariates.

Package:	pencoxfrail
Type:	Package
Version:	1.0.1
Date:	2016-05-06
License:	GPL-2
LazyLoad:	yes

for loading a dataset type data(nameofdataset)

Value

Generic functions such as `print`, `predict`, `plot` and `summary` have methods to show the results of the fit.

The `predict` function uses also estimates of random effects for prediction, if possible (i.e. for known subjects of the grouping factor). Either the survival stepfunction or the baseline hazard (not cumulative!) can be calculated by specifying one of two possible methods: `method=c("hazard","survival")`. By default, for each new subject in `new.data` an individual stepfunction is calculated on a pre-specified time grid, also accounting for covariate changes over time. Alternatively, for `new.data` a single vector of a specific (time-constant) covariate combination can be specified.

Usage: `predict(pencoxfrail.obj,new.data,time.grid,method=c("hazard","survival"))`

The `plot` function plots all time-varying effects, including the baseline hazard.

`call`	a list containing an image of the `pencoxfrail` call that produced the object.
`baseline`	a vector containing the estimated B-spline coefficients of the baseline hazard. If the covariates corresponding to the time-varying effects are centered (and standardized, see `pencoxfrailControl`), the coefficients are transformed back to the original scale.
`time.vary`	a vector containing the estimated B-spline coefficients of all time-varying effects. If the covariates corresponding to the time-varying effects are standardized (see `pencoxfrailControl`) the coefficients are transformed back to the original scale.
`coefficients`	a vector containing the estimated fixed effects.
`ranef`	a vector containing the estimated random effects.
`Q`	a scalar or matrix containing the estimates of the random effects standard deviation or variance-covariance parameters, respectively.
`Delta`	a matrix containing the estimates of fixed and random effects (columns) for each iteration (rows) of the main algorithm (i.e. before the final re-estimation step is performed, see details).
`Q_long`	a list containing the estimates of the random effects variance-covariance parameters for each iteration of the main algorithm.
`iter`	number of iterations until the main algorithm has converged.
`adaptive.weights`	If $\xi = 0$, a two-column matrix of adaptive weights is calculated; the first column contains the weights $w_{\Delta,k}$, the second column the weights v_k from $\xi \cdot J(\zeta, \alpha)$. If $\xi > 0$, the adaptive weights that have been used in the function's argument are displayed.
`knots`	vector of knots used in the B-spline representation.
`Phi.big`	large B-spline design matrix corresponding to the baseline hazard and all time-varying effects. For the time-varying effects, the B-spline functions (as a function of time) have already been multiplied with their associated covariates.
`time.grid`	the time grid used in when approximating the (Riemann) integral involved in the model's full likelihood.
`m`	number of metric covariates with time-varying effects.
`m2`	number of categorical covariates with time-varying effects.

Author(s)

Andreas Groll `<groll@math.lmu.de>`

References

Groll, A., T. Hastie and G. Tutz (2016). Regularization in Cox Frailty Models. Ludwig-Maximilians-University. *Technical Report* 191.

See Also

`pencoxfrailControl`, `Surv`, `pbc`

Examples

```
## Not run:
data(lung)

# remove NAs
lung <- lung[!is.na(lung$inst),]

# transform inst into factor variable
lung$inst <- as.factor(lung$inst)

# Random institutional effect
fix.form <- as.formula("Surv(time, status) ~ 1")
vary.coef <- as.formula("~ age")

pen.obj <- pencoxfrail(fix=fix.form,vary.coef=vary.coef, rnd = list(inst=~1),
              data=lung, xi=10,control=list(print.iter=TRUE))

# show fit
plot(pen.obj)

# predict survival curve of new subject, institution 1 and up to time 500
pred.obj <- predict(pen.obj,newdata=data.frame(inst=1,time=NA,status=NA,age=26),
              time.grid=seq(0,500,by=1))

# plot predicted hazard function
plot(pred.obj$time.grid,pred.obj$haz,type="l",xlab="time",ylab="hazard")

# plot predicted survival function
plot(pred.obj$time.grid,pred.obj$survival,type="l",xlab="time",ylab="survival")

# see also demo("pencoxfrail-pbc")

## End(Not run)
```

`pencoxfrailControl` *Control Values for* `pencoxfrail` *fit*

Description

The values supplied in the function call replace the defaults and a list with all possible arguments is returned. The returned list is used as the `control` argument to the `pencoxfrail` function.

Usage

```
pencoxfrailControl(start = NULL, q_start = NULL, conv.eps = 1e-4,
                             standardize = FALSE, center = FALSE,
                             smooth=list(nbasis = 6, penal = 0.1),
                             ridge.pen = 1e-4, print.iter = FALSE,
                             max.iter = 100, c.app = 1e-6, zeta = 0.5,
                             exact = 1e-2, xr = NULL, ...)
```

Arguments

`start`	a vector of suitable length containing starting values for the spline-coefficients of the baseline hazard and the time-varying effects, followed by the fixed and random effects. The correct ordering is important. Default is a vector full of zeros.
`q_start`	a scalar or matrix of suitable dimension, specifying starting values for the random-effects variance-covariance matrix. Default is a scalar 0.1 or diagonal matrix with 0.1 in the diagonal, depending on the dimension of the random effects.
`conv.eps`	controls the speed of convergence. Default is 1e-4.
`center`	logical. If true, the covariates corresponding to the time-varying effects will be centered. Default is FALSE (and centering is only recommended if really necessary; it can also have a strong effect on the baseline hazard, in particular, if a strong penalty is selected).
`standardize`	logical. If true, the the covariates corresponding to the time-varying effects will be scaled to a variance equal to one (*after* possible centering). Default is FALSE.
`smooth`	a list specifying the number of basis functions `nbasis` (used for the baseline hazard and all time-varying effects) and the smoothness penalty parameter `penal`, which is only applied to the baseline hazard. All time-varying effects are penalized by the specific double-penalty $\xi \cdot J(\zeta,\alpha)$ (see `pencoxfrail`), which is based on the overall penalty parameter ξ (specified in the main function `pencoxfrail`) and on the weighting between the two penalty parts ζ. The degree of the B-splines is fixed to be three (i.e. cubic splines).
`ridge.pen`	On all time-varying effects (except for the baseline hazard) a slight ridge penalty is applied on the second order differences of the corresponding spline coefficients to stabilize estimation. Default is 1e-4.
`print.iter`	logical. Should the number of iterations be printed? Default is FALSE.
`max.iter`	the number of iterations for the final Fisher scoring re-estimation procedure. Default is 200.
`c.app`	The parameter controlling the exactness of the quadratic approximations of the penalties. Default is 1e-6.
`zeta`	The parameter controlling the weighting between the two penalty parts in the specific double-penalty $\xi \cdot J(\zeta,\alpha)$ (see `pencoxfrail`). Default is 0.5.
`exact`	controls the exactness of the (Riemann) integral approximations. Default is 1e-2.
`xr`	maximal time point that is regarded. Default is NULL and the maximal event or censoring time point in the data is used.
`...`	Futher arguments to be passed.

Value

a list with components for each of the possible arguments.

Author(s)

Andreas Groll <groll@math.lmu.de>

See Also

`pencoxfrail`

Examples

```
# Use different weighting of the two penalty parts
# and lighten the convergence criterion
pencoxfrailControl(zeta=0.3, conv.eps=1e-3)
```

Index