Package ‘survivalMPLdc’

October 13, 2020

Type Package

Title Penalised Likelihood for Survival Analysis with Dependent Censoring

Version 0.1.0

Description Fitting Cox proportional hazard model under dependent right censoring using copula and maximum penalised likelihood methods.

Depends R (>= 3.5)

License GPL-3

Imports copula, stats, methods, splines2, survival, graphics, matrixcalc

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Suggests testthat, knitr, rmarkdown

Language en-GB

VignetteBuilder knitr

NeedsCompilation no

Author Jing Xu [aut, cre],
Jun Ma [aut],
Thomas Fung [aut] (<https://orcid.org/0000-0003-2601-0728>)

Maintainer Jing Xu <kenny.xu@duke-nus.edu.sg>

Repository CRAN

Date/Publication 2020-10-13 15:10:02 UTC

R topics documented:

`survivalMPLdc-package` *Penalised Likelihood for Survival Analysis with Dependent Censoring*

Description

Penalised Likelihood for Survival Analysis with Dependent Censoring

References

Ma, J. (2010). *"Positively constrained multiplicative iterative algorithm for maximum penalised likelihood tomographic reconstruction"*. IEEE Transactions On Signal Processing 57, 181-192.

Brodaty H, Woodward M, Boundy K, Ames D, Balshaw R. (2011). *"Patients in Australian memory clinics: baseline characteristics and predictors of decline at six months"*. Int Psychogeriatr 23, 1086-1096.

Ma, J. and Heritier, S. and Lo, S. (2014). *"On the Maximum Penalised Likelihood Approach for Proportional Hazard Models with Right Censored Survival Data"*. Computational Statistics and Data Analysis 74, 142-156.

Brodaty H, Connors M, Xu J, Woodward M, Ames D. (2014). *"Predictors of institutionalization in dementia: a three year longitudinal study"*. Journal of Alzheimers Disease 40, 221-226.

Xu J, Ma J, Connors MH, Brodaty H. (2018). *"Proportional hazard model estimation under dependent censoring using copulas and penalized likelihood"*. Statistics in Medicine 37, 2238–2251.

`coef.coxph_mpl_dc` *Extract regression coefficients of a coxph_mpl_dc Object*

Description

Extract the matrix of regression coefficients with their corresponding standard errors, z-statistics and p-values of the model part of interest of a `coxph_mpl_dc` object

Usage

```
## S3 method for class 'coxph_mpl_dc'
coef(object, parameter, ...)
```

Arguments

`object`	an object inheriting from class `coxph_mpl_dc`
`parameter`	the set of parameters of interest. Indicate `parameters="beta"` for the regression parameter of beta and `parameters="phi"` for the regression parameter of phi
`...`	other arguments

Details

When the input is of class `coxph_mpl_dc` and `parameters=="beta"`, the matrix of beta estimates with corresponding standar errors, z-statistics and p-values are reported. When the input is of class `coxph_mpl_dc` and `parameters=="phi"`, the matrix of phi estimates with corresponding standar errors, z-statistics and p-values are reported.

Value

`est` a matrix of coefficients with standard errors, z-statistics and corresponding p-values

Author(s)

Jing Xu, Jun Ma, Thomas Fung

References

Brodaty H, Connors M, Xu J, Woodward M, Ames D. (2014). *"Predictors of institutionalization in dementia: a three year longitudinal study"*. Journal of Alzheimers Disease 40, 221-226.

Xu J, Ma J, Connors MH, Brodaty H. (2018). *"Proportional hazard model estimation under dependent censoring using copulas and penalized likelihood"*. Statistics in Medicine 37, 2238–2251.

See Also

`plot.coxph_mpl_dc`, `coxph_mpl_dc.control`, `coxph_mpl_dc`

Examples

```
 ##-- Copula types
 copula3 <- 'frank'

##-- A real example
##-- One dataset from Prospective Research in Memory Clinics (PRIME) study
##-- Refer to article Brodaty et al (2014),
##   the predictors of institutionalization of dementia patients over 3-year study period

data(PRIME)

surv<-as.matrix(PRIME[,1:3]) #time, event and dependent censoring indicators
cova<-as.matrix(PRIME[, -c(1:3)]) #covariates
colMeans(surv[,2:3])  #the proportions of event and dependent censoring

n<-dim(PRIME)[1];print(n)
p<-dim(PRIME)[2]-3;print(p)
names(PRIME)

##--MPL estimate Cox proportional hazard model for institutionalization under independent censoring
control <- coxph_mpl_dc.control(ordSp = 4,
                                binCount = 200, tie = 'Yes',
```

```
                              tau = 0.5, copula = copula3,
                              pent = 'penalty_mspl', smpart = 'REML',
                              penc = 'penalty_mspl', smparc = 'REML',
                              cat.smpar = 'No' )

coxMPLests_tau <- coxph_mpl_dc(surv=surv, cova=cova, control=control, )
MPL_beta<-coef(object = coxMPLests_tau, parameter = "beta",)
MPL_phi<-coef(object = coxMPLests_tau, parameter = "phi",)
```

---

coxph_mpl_dc *Fit Cox Proportional Hazard Regression Model under dependent right censoring via MPL and Archimedean Copulas*

Description

Simultaneously estimate the regression coefficients and the baseline hazard function of proportional hazard Cox models under dependent right censoring using maximum penalised likelihood (MPL) and Archimedean Copulas

Usage

```
coxph_mpl_dc(surv, cova, control,...)
```

Arguments

surv	the outcome of survival data, with the first column X (observed time), second column del (event indicator) and third column eta (dependent right censoring indicator).
cova	the covariate matrix, with dimension of n rows and p columns, where 'n' is the sample size and 'p' is the number of covariates Default is `cova=matrix(0,n,1)`, which the covariates are not considered.
control	object of class `coxph_mpl_dc.control` specifying control options like basis choice, refer to `coxph_mpl_dc.control` to see the defaults.
...	other arguments. In `coxph_mpl_dc`, these elements, will be passed to `coxph_mpl_dc.control`.

Details

coxph_mpl_dc allows to simultaneously estimate the regression coefficients and baseline hazard function of Cox proportional hazard models, with dependent and independent right censored data, by maximizing a penalized likelihood, in which a penalty function is used to smooth the baseline hazard estimates. Note that the dependence between event and censoring times is modelled by an Archimedean copula.

Optimization is achieved using an iterative algorithm, which combines Newton's method and the multiplicative iterative algorithm proposed by Ma (2010), and respects the non-negativity constraints on the baseline hazard estimate (refer to Ma et al (2014) and Xu et al (2018)).

The centered covariate matrix Z is used in the optimization process to get a better shaped (penalized) log-likelihood. Baseline hazard parameter estimates and covariance matrix are then respectively corrected using a correction factor and the delta method.

The estimates of zero are possible for baseline hazard parameters (e.g., when number of knots is relatively large to sample size) and will correspond to active constraints as defined by Moore and Sadler (2008). Inference, as described by Ma et al (2014) or Xu et al (2018), is then corrected accordingly (refer to Moore and Sadler (2008)) by adequately cutting the corresponding covariance matrix.

There are currently three ways to perform inference on model parameters: Let H denote the negative of Hessian matrix of the non-penalized likelihood, Q denote the product of the first order derivative of the penalized likelihood by its transpose, and M_2 denote the negative of the second order derivative of the penalized likelihood. Then the three estimated covariance matrices for the MPL estimates are M_2^{-1}, $M_2^{-1}HM_2^{-1}$ and $M_2^{-1}QM_2^{-1}$.

Simulation studies the coverage levels of confidence intervals for the regression parameters seem to indicate $M_2^{-1}HM_2^{-1}$ performs better when using the piecewise constant basis, and that $M_2^{-1}QM_2^{-1}$ performs better when using other bases.

Value

`mpl_theta`	MPL estimates of the regression coefficient for the basis functions of the baseline hazard of T, i.e. theta
`mpl_gamma`	MPL estimates of the regression coefficient for the basis functions of the baseline hazard of C, i.e. gamma
`mpl_h0t`	MPL estimates of the baseline hazard for T, i.e. h_0T(x_i)
`mpl_h0c`	MPL estimates of the baseline hazard for C, i.e. h_0C(x_i)
`mpl_H0t`	MPL estimates of the baseline cumulative hazard for T, i.e. H_0T(x_i)
`mpl_H0c`	MPL estimates of the baseline cumulative hazard for C, i.e. H_0c(x_i)
`mpl_S0t`	MPL estimates of the baseline survival for T, i.e. S_0T(x_i)
`mpl_S0c`	MPL estimates of the baseline survival for C, i.e. S_0C(x_i)
`mpl_beta`	MPL estimates of beta
`mpl_phi`	MPL estimates of phi
`penloglik`	the penalized log-likelihood function given the MPL estimates
`mpl_Ubeta`	the first derivative of penalized log-likelihood function with respect to beta given the MPL estimates
`mpl_Uphi`	the first derivative of penalized log-likelihood unction with respect to phi given the MPL estimates
`mpl_Utheta`	the first derivative of penalized log-likelihood function with respect to theta given the MPL estimates
`mpl_Ugamma`	the first derivative of penalized log-likelihood function with respect to gamma given the MPL estimates
`iteration`	a vector of length 3 indicating the number of iterations used to estimate the smoothing parameter (first value, equal to 1 when the user specified a chosen value), the beta, phi, theta and gamma parameters during the entire process (second value), and beta, phi, theta and gamma parameters during the last smoothing parameter iteration (third value)

mpl_cvl the cross validation value given the MPL estimates

mpl_aic the AIC value given the MPL estimates

mpl_beta_sd the asymptotic standard deviation of the MPL estimated beta

mpl_phi_sd the asymptotic standard deviation of the MPL estimated phi

mpl_h0t_sd the asymptotic standard deviation of the MPL estimates for the baseline hazard coefficient of T, i.e. theta

mpl_h0c_sd the asymptotic standard deviation of the MPL estimates for the baseline hazard coefficient of C, i.e. gamma

mpl_ht0_sd the asymptotic standard deviation of the MPL estimates for the baseline hazard of T

mpl_hc0_sd the asymptotic standard deviation of the MPL estimates for the baseline hazard of C

mpl_Ht0_sd the asymptotic standard deviation of the MPL estimates for the cumulative baseline hazard of T

mpl_Hc0_sd the asymptotic standard deviation of the MPL estimates for the cumulative baseline hazard of C

mpl_St0_sd the asymptotic standard deviation of the MPL estimates for the baseline survival of T

mpl_Sc0_sd the asymptotic standard deviation of the MPL estimates for the baseline survival of C

mpl_est_cov the asymptotic covariance matrix of the MPL estimates

mpl_beta_phi_zp
the MPL estimates for regression coefficient with their corresponding standard deviations, z scores and p-values

binwv the width of each discretized bin of the observed times when piecewise constant approximation applied to the baseline hazards

ID the bin ID for each individual of the sample when piecewise constant approximation applied to the baseline hazards

binedg the edge for each discretized bin among the observed time vector X, which are the internal knots and boundaries

psix basis function matrix psi(x_i) with dimension of n by m for baseline hazard, where m=number of internal knots+ordSp

Psix basis function matrix Psi(x_i) with dimension of n by m for baseline cumulative hazard

Inputs defined in coxph_mpl_dc.control

Author(s)

Jing Xu, Jun Ma, Thomas Fung

References

Ma, J. (2010). *"Positively constrained multiplicative iterative algorithm for maximum penalised likelihood tomographic reconstruction"*. IEEE Transactions On Signal Processing 57, 181-192.

Ma, J. and Heritier, S. and Lo, S. (2014). *"On the Maximum Penalised Likelihood Approach forProportional Hazard Models with Right Censored Survival Data"*. Computational Statistics and Data Analysis 74, 142-156.

Xu J, Ma J, Connors MH, Brodaty H. (2018). *"Proportional hazard model estimation under dependent censoring using copulas and penalized likelihood"*. Statistics in Medicine 37, 2238–2251.

See Also

plot.coxph_mpl_dc, coxph_mpl_dc.control, coef.coxph_mpl_dc

Examples

```
##-- Copula types
copula3 <- 'frank'

##-- Marginal distribution for T, C, and A
a <- 2
lambda <- 2
cons7 <- 0.2
cons9 <- 10
tau <- 0.8
betas <- c(-0.5, 0.1)
phis <- c(0.3, 0.2)
distr.ev <- 'weibull'
distr.ce <- 'exponential'

##-- Sample size
n <- 200

##-- One sample Monte Carlo dataset
cova <- cbind(rbinom(n, 1, 0.5), runif(n, min=-10, max=10))
surv <- surv_data_dc(n, a, cova, lambda, betas, phis, cons7, cons9, tau, copula3,
                    distr.ev, distr.ce)
n <- nrow(cova)
p <- ncol(cova)
##-- event and dependent censoring proportions
colSums(surv)[c(2,3)]/n
X <- surv[,1] # Observed time
del<-surv[,2] # failure status
eta<-surv[,3] # dependent censoring status

##-- control inputs for the coxph_mpl_dc function
control <- coxph_mpl_dc.control(ordSp = 4,
                            binCount = 100,
                            tau = 0.8, copula = copula3,
                            pent = 'penalty_mspl', smpart = 'REML',
                            penc = 'penalty_mspl', smparc = 'REML',
```

```
                            cat.smpar = 'No' )

##-- Fitting cox ph hazard model for T using MPL and an correct copula
#with REML smoothing parameters
coxMPLests5 <- coxph_mpl_dc(surv, cova, control, )
mpl_beta_phi_zp5 <- coxMPLests5$mpl_beta_phi_zp
mpl_h0t5 <- coxMPLests5$mpl_h0t
mpl_h0Ti5 <- approx( X, mpl_h0t5, xout = seq(0, 5.4, 0.01),
                   method="constant", rule = 2, ties = mean)$y

##-- Real marginal baseline hazard for T
ht0b <- a * (seq(0, 5.4, 0.01) ^ (a - 1)) / (lambda ^ a)


##-- Fitting cox ph hazard model for T using MPL and an correct copula
#with zero smoothing parameters
coxMPLests3 <- coxph_mpl_dc(surv, cova,
                         ordSp=4, binCount=100,
                         tau=0.8, copula=copula3,
                         pent='penalty_mspl', smpart=0, penc='penalty_mspl', smparc=0,
                         cat.smpar = 'No')
mpl_beta_phi_zp3 <- coxMPLests3$mpl_beta_phi_zp
mpl_h0t3 <- coxMPLests3$mpl_h0t
mpl_h0Ti3 <- approx( X, mpl_h0t3, xout = seq(0, 5.4, 0.01),
method="constant", rule = 2, ties = mean)$y

##-- Plot the true and estimated baseline hazards for T
t_up <- 3.5
y_uplim <- 2
Ti<-seq(0, 5.4, 0.01)[seq(0, 5.4, 0.01)<=t_up]
h0Ti<-ht0b[seq(0, 5.4, 0.01)<=t_up]
h0Ti5<-mpl_h0Ti5[seq(0, 5.4, 0.01)<=t_up]
h0Ti3<-mpl_h0Ti3[seq(0, 5.4, 0.01)<=t_up]

plot( x = Ti, y = h0Ti5,
     type="l", col="grey", lty=4, lwd=3, cex.axis=1.6, cex.lab=1.6, ylim=c(0, y_uplim),
     xlab='Time', ylab='Hazard')
lines(x = Ti, y = h0Ti,
     col="green",
     lty=1, lwd=3, cex.axis=1.6, cex.lab=1.6, ylim=c(0, y_uplim)
     )
lines(x = Ti, y = h0Ti3,
     col="blue",
     lty=4, lwd=3, cex.axis=1.6, cex.lab=1.6, ylim=c(0, y_uplim)
     )
```

---

coxph_mpl_dc.control *Ancillary arguments for controlling the outputs of coxph_mpl_dc*

Description

This is used to set various numeric parameters controlling a Cox model fit using coxph_mpl_dc. Typically it would only be used in a call to coxph_mpl_dc.

Usage

```
coxph_mpl_dc.control(ordSp,
                        binCount, tie,
                        tau, copula,
                        pent, smpart, penc, smparc,
                        maxit2, maxit,
                        mid, asy, ac, cv,
                        ac.theta, ac.gamma, ac.Utheta, ac.Ugamma,
                        min.theta, min.gamma,
                        min.ht, min.hc, min.St, min.Sc, min.C, min.dC,
                        eps, tol.thga, tol.bph, cat.smpar, tol.smpar
                        )
```

Arguments

`ordSp`	the order of spline for the basis function for baseline hazard for both T and C, can be 'piecewise constant' if `ordSp=1`, cubic 'm-spline' if `ordSp=4`, etc. Default is `ordSp=1`.
`binCount`	the number of subjects in each discretized bin, can be selected either by trial and error or AIC method Default is `binCount=1`.
`tie`	tie='No' if tied observations are not existed, otherwise tied observations existed. Default is `tie='No'`.
`tau`	the kendall's correlation coefficient between T and C. Default is `tau=0`.
`copula`	Archimedean copula type, i.e. 'independent', 'clayton', 'gumbel' and 'frank'. Default is `copula='independent'`.
`pent`	penalty function type for T, i.e. mat1 (first order difference) or mat2 (second order difference) for piecewise constant basis, penalty_mspl for m-spline basis Default is `pent='mat1'`.
`smpart`	value of smoothing parameter for T, can be selected by either trial and error or cross validation method. Note that smpart can be also estimated by restricted maximum likelihood (i.e. `smpart='REML'`). Default is `smpart=0`.
`penc`	penalty function type for C, i.e. mat1 (first order difference) or mat2 (second order difference) for piecewise constant basis, penalty_mspl for m-spline basis Default is `pent='mat1'`.

smparc
: value of smoothing parameter for C, can be selected by either trial and error or cross validation method. Note that smparc can be also estimated by restricted maximum likelihood (i.e. `smparc='REML'`). Default is `smparc=0`.

maxit2
: maximum number of iterations for smpart and smparc. Defualt is `maxit2=50`.

maxit
: maximum number of iteration for updating beta, phi, theta and gamma given fixed smpart and smparc. Default is `maxit=5000`.

mid
: the middle matrix selection for the sandwich formula that used to computed the asymptotic covariance matrix, i.e. `mid=1` (negative of the hessian matrix with zeros smoothing parameters, i.e. `smpart=smparc=0`, or negative of the matrix with second derivatives of the MPL estimates with respect to the log-likelihood), 2 (the matrix created by the vector of first derivative of the penalized log-likelihood with respect to the MPL estimates times its transpose) and otherwise (negative of the hessian matrix or negative of the matrix with second derivatives of the MPL estimates with respect to the penalized log-likelihood). Default is `mid=1`.

asy
: `asy=1` if asymptotic standard deviation of the MPL estimates are computed and 0 if not computed. Default is `asy=1`.

ac
: `ac=1` if aic value is calculated 0 if not. Default is `ac=0`.

cv
: `cv=0` if cv value is calculated 0 if not. Default is `cv=0`.

ac.theta
: the minimum value of theta for active contraints. Default is `ac.theta=1e-5`.

ac.gamma
: the minimum value of gamma for active contraints. Default is `ac.gamma=1e-5`.

ac.Utheta
: the minimum value of Utheta (the first derivative of the penalized log-likelihood with respect to theta) for active contraints. Default is `ac.Utheta=1e-2`.

ac.Ugamma
: the minimum value of Ugamma (the first derivative of the penalized log-likelihood with respect to gamma) for active contraints. Default is `ac.Ugamma=1e-2`.

min.theta
: a value indicating the minimal baseline hazard parameter value theta updated at each iteration. Baseline hazard parameter theta estimates at each iteration lower than min.theta will be considered as min.theta. Default is `min.theta=1e-7`.

min.gamma
: a value indicating the minimal baseline hazard parameter value gamma updated at each iteration. Baseline hazard parameter gamma estimates at each iteration lower than min.gamma will be considered as min.gamma. Default is `min.gamma=1e-7`.

min.ht
: a value indicating the minimal baseline hazard of T updated at each iteration. Baseline hazard estimates of T at each iteration lower than min.ht will be considered as min.ht. Default is `min.ht=1e-7`.

min.hc
: a value indicating the minimal baseline hazard of C updated at each iteration. Baseline hazard estimates of C at each iteration lower than min.hc will be considered as min.hc. Default is `min.hc=1e-7`.

min.St
: a value indicating the minimal baseline survival of T updated at each iteration. Baseline survival estimates of T at each iteration lower than min.St will be considered as min.St. Default is `min.St=1e-7`.

min.Sc
: a value indicating the minimal baseline survival of C updated at each iteration. Baseline survival estimates of C at each iteration lower than min.Sc will be considered as min.Sc. Default is `min.Sc=1e-7`.

`min.C`	a value indicating the minimal copula $K(u,v)$ at each iteration, lower than min.C will be considered as min.C. Default is `min.C=1e-7`.
`min.dC`	a value indicating the minimal first i.e. $dK(u,v)/du$ and $dK(u,v)/dv$ and second i.e. $d^2K(u,v)/dudv$ derivatives of copula $K(u,v)$ at each iteration, lower than min.dC will be considered as min.dC. Default is `min.dC=1e-7`.
`eps`	a small positive value added to the diagonal of a square matrix. Default value is `eps=1e-5`.
`tol.thga`	the convergence tolerence value for both theta and gamma. Convergence is achieved when the maximum absolute difference between the parameter estimates at iteration k and iteration k-1 is smaller than tol.thga. Default is `tol.thga=1e-5`.
`tol.bph`	the convergence tolerence value for both beta and phi. Convergence is achieved when the maximum absolute difference between the parameter estimates at iteration k and iteration k-1 is smaller than tol.bph. Default is `tol.bph=1e-5`.
`cat.smpar`	cat.smpar='Yes' to display the smoothing parameters estimation process, otherwise not to display. Default is `cat.smpar='Yes'`.
`tol.smpar`	the convergence tolerence value for both smpart and smparc. Convergence is achieved when the maximum absolute difference between the parameter estimates at iteration k and iteration k-1 is smaller than tol.smpar. Default is `tol.smpar=1e-2`.

Value

A list containing the values of each of the above arguments for most of the inputs of Coxph_mpl_dc.

Author(s)

Jing Xu, Jun Ma, Thomas Fung

References

Ma, J. and Heritier, S. and Lo, S. (2014). *"On the Maximum Penalised Likelihood Approach forProportional Hazard Models with Right Censored Survival Data"*. Computational Statistics and Data Analysis 74, 142-156.

Xu J, Ma J, Connors MH, Brodaty H. (2018). *"Proportional hazard model estimation under dependent censoring using copulas and penalized likelihood"*. Statistics in Medicine 37, 2238–2251.

See Also

`plot.coxph_mpl_dc`, `coxph_mpl_dc`, `coef.coxph_mpl_dc`

Examples

```
control <- coxph_mpl_dc.control(ordSp=4,
                             binCount=40,
                             tau=0.8, copula='frank',
                  pent='penalty_mspl', smpart='REML', penc='penalty_mspl', smparc='REML',
                             cat.smpar='No'
                             )
```

`plot.coxph_mpl_dc` *Plot a baseline hazard estimates from coxph_mpl_dc Object*

Description

Plot the baseline hazard with the confidence interval estimates

Usage

```
## S3 method for class 'coxph_mpl_dc'
plot(
  x,
  parameter = "theta",
  funtype = "hazard",
  xout,
  se = TRUE,
  ltys,
  cols,
  ...
)
```

Arguments

`x`	an object inheriting from class `coxph_mpl_dc`
`parameter`	the set of parameters of interest. Indicate `parameters="theta"` for the baseline hazard estimated by $theta$ and `parameters="gamma"` for the baseline hazard estimated by $gamma$
`funtype`	the type of function for ploting, i.e. `funtype="hazard"` for baseline hazard, `funtype="cumhazard"` for baseline cumulative hazard and `funtype="survival"` for baseline survival function
`xout`	the time values for the baseline hazard plot
`se`	se=TRUE gives both the MPL baseline estimates and 95% confidence interval plots while se=FALSE gives only the MPL baseline estimate plot.
`ltys`	a line type vector with two components, the first component is the line type of the baseline hazard while the second component is the line type of the 95% confidence interval
`cols`	a colour vector with two components, the first component is the colour of the baseline hazard while the second component is the colour the 95% confidence interval
`...`	other arguments

Details

When the input is of class `coxph_mpl_dc` and `parameters=="theta"`, the baseline estimates base on θ and xout (with the corresponding 95% confidence interval if se=TRUE) are ploted. When the input is of class `coxph_mpl_dc` and `parameters=="gamma"`, the baseline hazard estimates based on γ and xout (with the corresponding 95% confidence interval if se=TRUE) are ploted.

Value

the baseline hazard plot

Author(s)

Jing Xu, Jun Ma, Thomas Fung

References

Brodaty H, Connors M, Xu J, Woodward M, Ames D. (2014). *"Predictors of institutionalization in dementia: a three year longitudinal study".* Journal of Alzheimers Disease 40, 221-226.

Xu J, Ma J, Connors MH, Brodaty H. (2018). *"Proportional hazard model estimation under dependent censoring using copulas and penalized likelihood".* Statistics in Medicine 37, 2238–2251.

See Also

coef.coxph_mpl_dc, coxph_mpl_dc.control, coxph_mpl_dc

Examples

```
 ##-- Copula types
 copula3 <- 'frank'

##-- A real example
##-- One dataset from Prospective Research in Memory Clinics (PRIME) study
##-- Refer to article Brodaty et al (2014),
##   the predictors of institutionalization of dementia patients over 3-year study period

data(PRIME)

surv<-as.matrix(PRIME[,1:3]) #time, event and dependent censoring indicators
cova<-as.matrix(PRIME[, -c(1:3)]) #covariates
colMeans(surv[,2:3])  #the proportions of event and dependent censoring

n<-dim(PRIME)[1];print(n)
p<-dim(PRIME)[2]-3;print(p)
names(PRIME)

##--MPL estimate Cox proportional hazard model for institutionalization under medium positive
##--dependent censoring
control <- coxph_mpl_dc.control(ordSp = 4,
                                binCount = 200, tie = 'Yes',
                                tau = 0.5, copula = copula3,
                                pent = 'penalty_mspl', smpart = 'REML',
                                penc = 'penalty_mspl', smparc = 'REML',
                                cat.smpar = 'No' )

coxMPLests_tau <- coxph_mpl_dc(surv=surv, cova=cova, control=control, )

plot(x = coxMPLests_tau, parameter = "theta", funtype="hazard",
```

```
     xout = seq(0, 36, 0.01), se = TRUE,
     cols=c("blue", "red"), ltys=c(1, 2), type="l", lwd=1, cex=1, cex.axis=1, cex.lab=1,
     xlab="Time (Month)", ylab="Hazard",
     xlim=c(0, 36), ylim=c(0, 0.05)
     )
     title("MPL Hazard", cex.main=1)
     legend( 'topleft',legend = c( expression(tau==0.5), "95% Confidence Interval"),
     col = c("blue", "red"),
     lty = c(1, 2),
     cex = 1)

plot(x = coxMPLests_tau, parameter = "theta", funtype="cumhazard",
    xout = seq(0, 36, 0.01), se = TRUE,
    cols=c("blue", "red"), ltys=c(1, 2),
    type="l", lwd=1, cex=1, cex.axis=1, cex.lab=1,
    xlab="Time (Month)", ylab="Hazard",
    xlim=c(0, 36), ylim=c(0, 1.2)
)
title("MPL Cumulative Hazard", cex.main=1)
legend( 'topleft',
       legend = c( expression(tau==0.5), "95% Confidence Interval"),
       col = c("blue", "red"),
       lty = c(1, 2),
       cex = 1
)

plot(x = coxMPLests_tau, parameter = "theta", funtype="survival",
    xout = seq(0, 36, 0.01), se = TRUE,
    cols=c("blue", "red"), ltys=c(1, 2),
    type="l", lwd=1, cex=1, cex.axis=1, cex.lab=1,
    xlab="Time (Month)", ylab="Hazard",
    xlim=c(0, 36), ylim=c(0, 1)
)
title("MPL Survival", cex.main=1)
legend( 'bottomleft',
       legend = c( expression(tau==0.5), "95% Confidence Interval"),
       col = c("blue", "red"),
       lty = c(1, 2),
       cex = 1
)
```

PRIME *PRIME data set*

Description

This data set is from a longitudinal study called "Prospective Research in Memory Clinics" (PRIME), see Brodaty et al (2011), with a period of 3-year. The data set includes 583 dementia patients. The

outcome is time to institutionalised. The predicators are age, sex, educational level, living status, dementia type, baseline cognitive ability (MMSE), baseline functional ability (SMAF), baseline neuropsychiatric symptoms (total NPI), baseline dementia severity (CDR), baseline caregiver burden (ZBI), medication types, change in cognitive ability at 3 months, change in functional ability at 3 months, and change in neuropsychiatric symptoms at 3 months. Note that this data set is complete and was analyzed by Brodaty et al (2014).

Usage

```
data(PRIME)
```

Format

A data frame with 583 observations on 18 variables.

Time observed time in term of months

Event event or institutionalisation, 1=Yes and 0=No

Depcen dependent right censoring or withdrawal, 1=Yes and 0=No

Age age at baseline, 1=80 years or above and 0=80 year below

Gender gender, 1=Female and 0=Male

HighEdu education level at baseline, 1=high school above and 0=high school or below

Alzheimer Alzheimer disease, 1=Yes and 0=No

CDR_base dementia severity at baseline

MMSE_base cognitive ability at baseline

SMAF_base functional ability at baseline

ZBI_base caregiver burden at baseline

NPI_base neuropsychiatric symptoms at baseline

Benzon benzodiazepines taking, 1=Yes and 0=No

Antiphsy anti-psychotics taking, 1=Yes and 0=No

LivingAlone living alone, 1=Yes and 0=No

MMSE_change_3m cognitive ability change at 3-month from baseline

SMAF_change_3m functional ability change at 3-month from baseline

NPI_change_3m neuropsychiatric symptoms change at 3-month from baseline

References

Brodaty H, Woodward M, Boundy K, Ames D, Balshaw R. (2011). *"Patients in Australian memory clinics: baseline characteristics and predictors of decline at six months"*. Int Psychogeriatr 23, 1086-1096.

Brodaty H, Connors M, Xu J, Woodward M, Ames D. (2014). *"Predictors of institutionalization in dementia: a three year longitudinal study"*. Journal of Alzheimers Disease 40, 221-226.

`surv_data_dc` *Generate a sample of time to event dataset with dependent right censoring under an Archimedean copula*

Description

Generate a sample of time to event dataset with, dependent right censoring based on one of the Archimedean copulas the given Kendall's tau, sample size n and covariates matrix Z.

Usage

```
surv_data_dc(n, a, Z, lambda, betas, phis, cons7, cons9, tau, copula, distr.ev, distr.ce)
```

Arguments

`n`	the sample size, or the number of the subjects in a sample.
`a`	the shape parameter of baseline hazard for the event time T.
`Z`	the covariate matrix with dimension of n by p, where p is the number of covariates.
`lambda`	the scale parameter of baseline hazard for event time T.
`betas`	the regression coefficient vector of proportional hazard model for the event time T with dimenion of p by 1.
`phis`	the regression coefficient vector of proportional hazard model for dependent censoring time C with dimenion of p by 1.
`cons7`	the parameter of baseline hazard for the dependent censoring time C if assuming an exponential distribution.
`cons9`	the upper limit parameter of uniform distribution for the independent censoring time A, i.e. A~U(0, cons9).
`tau`	the Kendall's correlation coefficient between T and C.
`copula`	the Archemedean copula that captures the dependence between T and C, a characteristc value, i.e. 'independent', 'clayton', 'gumbel' or 'frank'.
`distr.ev`	the distribution of the event time, a characteristc value, i.e. 'weibull' or 'log logit'.
`distr.ce`	the distribution of the dependent censoring time, a characteristc value, i.e. 'exponential' or 'weibull'.

Details

surv_data_dc allows to generate a survival dataset under dependent right censoring, at sample size `n`, based on one of the Archimedean `copula`, Kendall's `tau`, and covariates matrix `Z` with dimension of n by p. For example, at `p=2`, we have `Z=cbind(Z1,Z2)`, where `Z1` is treatment generated by distribution of bernoulli(0.5), i.e. 1 represents treatment group and 0 represents control group; `Z2` is the age generated by distribution of U(-10, 10).

The generated dataset includes three varaibles, which are X_i, δ_i and η_i, i.e. $X_i = min(T_i, C_i, A_i)$, $\delta_i = I(X_i = T_i)$ and $\eta_i = I(X_i = C_i)$, for $i = 1, \ldots, n$. 'T' represents the event time, whose hazard function is

$$h_T(x) = h_{0T}(x)exp(Z^\top \beta)$$

, where the baseline hazard can take weibull form, i.e. $h_{0T}(x) = ax^{a-1}/\lambda^a$, or log logistic form, i.e.

$$h_{0T}(x) = \frac{\frac{1}{aexp(\lambda)}(\frac{x}{exp(\lambda)})^{1/a-1}}{1 + (\frac{x}{exp(\lambda)})^{1/a}}$$

. 'C' represents the dependent censoring time, whose hazard function is $h_C(x) = h_{0C}(x)exp(Z^\top \phi)$, where the baseline hazard can take exponential form, i.e. $h_{0C}(x) = cons7$, or weibull form, i.e. $h_{0C}(x) = ax^{a-1}/\lambda^a$.'A' represents the administrative or independent censoring time, where A~U(0, cons9).

Value

A sample of time to event dataset under dependent right censoring, which includes observed time X, event indicator δ and dependent censoring indicator η.

Author(s)

Jing Xu, Jun Ma, Thomas Fung

References

Xu J, Ma J, Connors MH, Brodaty H. (2018). *"Proportional hazard model estimation under dependent censoring using copulas and penalized likelihood"*. Statistics in Medicine 37, 2238–2251.

See Also

coxph_mpl_dc

Examples

```
##-- Copula types
copula3 <- 'frank'

##-- Marginal distribution for T, C, and A
a <- 2
lambda <- 2
cons7 <- 0.2
cons9 <- 10
tau <- 0.8
betas <- c(-0.5, 0.1)
phis <- c(0.3, 0.2)
distr.ev <- 'weibull'
distr.ce <- 'exponential'

##-- Sample size
n <- 200
```

```
##-- One sample Monte Carlo dataset
cova <- cbind(rbinom(n, 1, 0.5), runif(n, min=-10, max=10))
surv <- surv_data_dc(n, a, cova, lambda, betas, phis, cons7, cons9,
                    tau, copula3, distr.ev, distr.ce)
n <- nrow(cova)
p <- ncol(cova)
##-- event and dependent censoring proportions
colSums(surv)[c(2,3)]/n
X <- surv[,1] # Observed time
del<-surv[,2] # failure status
eta<-surv[,3] # dependent censoring status
```

Index