

Package ‘CRFCSD’

April 4, 2019

Type Package

Title Mixture Cure Generalized Odds Ratio Frailty Models for Clustered Current Status Data

Version 1.0

Date 2019-03-24

Author Tong Wang, Kejun He, Wei Ma, Dipankar Bandyopadhyay, Samiran Sinha

Maintainer Tong Wang<tong@stat.tamu.edu>

Description A methodology to estimate the parameters for the cure rate frailty models with clustered current status data.

License GPL-2

Encoding UTF-8

Imports Rcpp (>= 0.12.18), numDeriv, splines2, orthopolynom

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.1.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2019-04-04 15:40:14 UTC

R topics documented:

CRFCSD-package	2
boot.CSD	2
boot.result	3
CSDfit	4
PD	5
Surv.CI	6
Index	8

CRFCSD-package	<i>Mixture Cure Generalized Odds Ratio Frailty Models for Clustered Current Status Data</i>
----------------	---

Description

A methodology to estimate the parameters for the cure rate frailty models with clustered current status data.

Author(s)

Tong Wang, Kejun He, Wei Ma, Dipankar Bandyopadhyay, Samiran Sinha
 Maintainer: Tong Wang<tong@stat.tamu.edu>

boot.CSD	<i>Bootstrap estimation for the parameters.</i>
----------	---

Description

For the purpose of the confidence band calculation, here we estimate the parametric and nonparametric components of the model- mixed cure GOR model based on the bootstrap resampling.

Usage

```
boot.CSD(Rawdata, n_subject.raw, n_within.raw, r, boot.rep, seed.begin,
         n_quad=30, lambda=0, Cauchy.pen=TRUE, tolerance=1e-2,
         knots.num=2, degree=2, scale.numr=TRUE, cure.reg = TRUE)
```

Arguments

Rawdata	This is a dataframe of the current status data. The first column should be the index of the subject (cluster). The second column is the inspection time. The next <code>n_subjec.raw</code> columns are the subject (cluster-specific) level covariates. Then the next <code>n_within.raw</code> columns are the within subject covariates. The last column is the indicator of the event where 1 or 0 indicate if the event has or has not happened by the inspection time, respectively. All the covariates are assumed to be either numerical or binary, and our program automatically detects if a covariate is a binary or numerical variable.
<code>n_subject.raw</code>	The number of subject (cluster-specific) level covariates.
<code>n_within.raw</code>	The number of within cluster covariates.
<code>r</code>	The index of the Generalized odds ratio (GOR) model. This index is a non-negative number and it must be specified by the user. Here <code>r=0</code> and <code>1</code> imply the proportional hazard and the proportional odds model, respectively.
<code>boot.rep</code>	The replications for the bootstrap (the number of bootstrap samples).

seed.begin	The random seed to start the bootstrap sampling. If the user wants to use parallel computation, just set different seed.begin for each nodes, then combine the results.
n_quad	The number of Gauss-Hermite quadrature nodes used in numerical integration. The default value is 30.
lambda	The tuning parameter of the roughness penalty used for estimating the non-parametric component of the GOR model. The default value is 0. One must use the roughness penalty when the number of basis functions in the non-parametric component of the GOR model is large.
Cauchy.pen	logical. If TRUE, then we use Cauchy penalty on the regression parameters to reduce the samll sample bias. The default is TRUE.
tolerance	This denotes the summation of the absolute values of the relative tolerance of all parameters in the model. It is used to define the convergence of the parameter estimates. The default value is 0.01.
knots.num	The number of equidistant interior knots for the integrated B-spline approximation of the nonparametric component of the GOR model. The default value is 2.
degree	The degree of integrated B-splines. The default value is 2.
scale.numr	logical. If TRUE, then all numeric covariates (cluster specifie and within cluster) are scaled with mean zero and standard deviation one. The default value is TRUE.
cure.reg	logical. TRUE and FALSE indicate modelling the cure rate part with covariates (including cluster and within cluster covariates) or not, respectively. The default value is TRUE.

Value

boot.CSD returns a dataframe, each row contains one parameter estimate based on one bootstrap sample.

Examples

```
data(PD)
boot.result=boot.CSD(PD,3,1,0,3,10,n_quad=5)
```

boot.result

A matrix of bootstrap estimates

Description

This is a matrix of bootstrap estimates, used to construct confidence interval for the predicated survival probability.

Usage

```
data(boot.result)
```

See Also

[Surv.CI](#)

CSDfit

Analysing clustered current status data

Description

CSDfit is used to analyze clustered current status data. The function provides parameter estimates, the maximum log likelihood value and the corresponding AIC value.

Usage

```
CSDfit(Rawdata, n_subject.raw, n_within.raw, r, n_quad = 30,
       lambda = 0, Cauchy.pen = TRUE, tolerance = 0.01, knots.num = 2,
       degree = 2, scale.numr = TRUE, cure.reg = TRUE)
```

Arguments

Rawdata	This is a dataframe of the current status data. The first column should be the index of the subject (cluster). The second column is the inspection time. The next <code>n_subjec.raw</code> columns are the subject (cluster-specific) level covariates. Then the next <code>n_within.raw</code> columns are the within subject covariates. The last column is the indicator of the event where 1 or 0 indicate if the event has or has not happened by the inspection time, respectively. All the covariates are assumed to be either numerical or binary, and our program automatically detects if a covariate is a binary or numerical variable.
<code>n_subject.raw</code>	The number of subject (cluster-specific) level covariates.
<code>n_within.raw</code>	The number of within cluster covariates.
<code>r</code>	The index of the Generalized odds ratio (GOR) model. This index is a non-negative number and it must be specified by the user. Here <code>r=0</code> and <code>1</code> imply the proportional hazard and the proportional odds model, respectively.
<code>n_quad</code>	The number of Gauss-Hermite quadrature nodes used in numerical integration. The default value is 30.
<code>lambda</code>	The tuning parameter of the roughness penalty used for estimating the non-parametric component of the GOR model. The default value is 0. One must use the roughness penalty when the number of basis functions in the non-parametric component of the GOR model is large.
<code>Cauchy.pen</code>	logical. If TRUE, then we use Cauchy penalty on the regression parameters to reduce the samll sample bias. The default is TRUE.

tolerance	This denotes the summation of the absolute values of the relative tolerance of all parameters in the model. It is used to define the convergence of the parameter estimates. The default value is 0.01.
knots.num	The number of equidistant interior knots for the integrated B-spline approximation of the nonparametric component of the GOR model. The default value is 2.
degree	The degree of integrated B-splines. The default value is 2.
scale.numr	logical. If TRUE, then all numeric covariates (cluster specific and within cluster) are scaled with mean zero and standard deviation one. The default value is TRUE.
cure.reg	logical. TRUE and FALSE indicate modelling the cure rate part with covariates (including cluster and within cluster covariates) or not, respectively. The default value is TRUE.

Value

Function CSDfit returns a list containing the following components:

parameter.est	It is a matrix. The column "par.est" contains the estimate of the regression parameters. The column "SE" contains the standard error of these estimators. The columns "Z" and "p-value" are the corresponding Z-statistic and p-value. The last two columns are the lower and upper bound of the 95% Wald's confidence interval for the parameters.
log_likelihoood	The maximum log likelihood value. This includes the logarithm of the roughness penalty and the Cauchy penalty if Cauchy.pen=TRUE.
AICvalue	The AIC value.
coefs	The estimates of the coefficients of the integrated B-spline basis functions.

Examples

```
data(PD)
PDfit=CSDfit(PD,3,1,0,n_quad=10)
```

PD

A clustered current status dataset.

Description

A clustered current status dataset arises from a periodontal disease (PD) study where tooth level data are clustered within subjects. The first and second columns are the index for patients and the inspection time for each tooth, respectively. The next 3 columns are the three subject level covariates (gender, smoking and HbA1c). After that, it is the tooth level covariate (jaw). The last column is the indicator for the event.

Usage

```
data(PD)
```

See Also[CSDfit](#)

Surv.CI	<i>Predicating the survival probability and the corresponding confidence interval.</i>
---------	--

Description

Surv.CI can predicate the survival probability and the corresponding confidence interval for a given set of time points and a given set of covariate values based on the bootstrap method. This uses the output of the function boot.CSD as an input.

Usage

```
Surv.CI(t,x,boot.par,Rawdata,n_subject,n_within,r,
CI.lev=0.95,n_quad=30,knots.num=2,degree=2,
lambda=0,Cauchy.pen=TRUE,tolerance=1e-2,scale.numr=TRUE,cure.reg=TRUE)
```

Arguments

t	The given set of time points. All time points should be between the minimum and maximum of the inspection time.
x	The given set of covariate vector. The first n_subject.raw covariates are subject (cluster) specific covariates. The rest of it are the within cluster covariates.
boot.par	The bootstrap estimation dataframe provided by the function boot.CSD
Rawdata	This is a dataframe of the current status data. The first column should be the index of the subject (cluster). The second column is the inspection time. The next n_subjec.raw columns are the subject (cluster-specific) level covariates. Then the next n_within.raw columns are the within subject covariates. The last column is the indicator of the event where 1 or 0 indicate if the event has or has not happened by the inspection time, respectively. All the covariates are assumed to be either numerical or binary, and our program automatically detects if a covariate is a binary or numerical variable.
n_subject	The number of subject (cluster-specific) level covariates.
n_within	The number of within cluster covariates.
r	The index of the Generalized odds ratio (GOR) model. This index is a non-negative number and it must be specified by the user. Here r=0 and 1 imply the proportional hazard and the proportional odds model, respectively.
CI.lev	The confidence level. The default value is 0.95.
n_quad	The number of Gauss-Hermite quadrature nodes used in numerical integration. The default value is 30.

lambda	The tuning parameter of the roughness penalty used for estimating the non-parametric component of the GOR model. The default value is 0. One must use the roughness penalty when the number of basis functions in the non-parametric component of the GOR model is large.
Cauchy.pen	logical. If TRUE, then we use Cauchy penalty on the regression parameters to reduce the samll sample bias. The default is TRUE.
tolerance	This denotes the summation of the absolute values of the relative tolerance of all parameters in the model. It is used to define the convergence of the parameter estimates. The default value is 0.01.
knots.num	The number of equidistant interior knots for the integrated B-spline approximation of the nonparametric component of the GOR model. The default value is 2.
degree	The degree of integrated B-splines. The default value is 2.
scale.numr	logical. If TRUE, then all numeric covariates (cluster specifie and within cluster) are scaled with mean zero and standard deviation one. The default value is TRUE.
cure.reg	logical. TRUE and FALSE indicate modelling the cure rate part with covariates (including cluster and within cluster covariates) or not, respectively. The default value is TRUE.

Value

Surv.band returns the predicated survival probability and the corresponding confidence interval.

Examples

```
data(PD)
t=c(45,50)
x=rbind(c(1,0,1,0),c(1,1,0,0))
data(boot.result)
CI_t=Surv.CI(t,x,boot.result,PD,3,1,0,n_quad=5)
```

Index

boot.CSD, [2](#)
boot.result, [3](#)

CRFCSD (CRFCSD-package), [2](#)
CRFCSD-package, [2](#)
CSDfit, [4, 6](#)

PD, [5](#)

Surv.CI, [4, 6](#)