Package 'CompetingRisk'

March 2, 2017

Type Package

Title The Semi-Parametric Cumulative Incidence Function

Version 1.0

Date 2017-02-25

Author Jiayi Hou,
Ronghui Xu

Maintainer Jiayi Hou <jhou12@icloud.com>

Description Computing the point estimator and pointwise confidence interval of the cumulative incidence function from the cause-specific hazards model.

Depends survival, MASS, wesanderson, reshape2, stats, graphics, methods, Matrix

License GPL-3

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2017-03-02 08:09:10

R topics documented:

`CompetingRisk-package` *The Semi-Parametric Cumulative Incidence Function*

Description

Computing the point estimator and pointwise confidence interval of the cumulative incidence function from the cause-specific hazards model. This package contains generic methods (`plot`,`summary`) that can be invoked for an object fitted using `est.CIF()`.

Details

Package:	CompetingRisk
Type:	Package
Title:	The Semi-Parametric Cumulative Incidence Function
Version:	1.0
Date:	2017-02-25
Author:	Jiayi Hou, Ronghui Xu
Maintainer:	Jiayi Hou <jhou12@icloud.com>
Description:	Computing the point estimator and pointwise confidence interval of the cumulative incidence function from t
Depends:	survival, MASS, wesanderson, reshape2, stats, graphics, methods, Matrix
License:	GPL-3
RoxygenNote:	6.0.1

Author(s)

Jiayi Hou and Ronghui Xu

Maintainer: Jiayi Hou <jhou12@icloud.com>

References

Cheng, S. C., Jason P. Fine, and L. J. Wei. "Prediction of cumulative incidence function under the proportional hazards model." Biometrics (1998): 219-228. DOI: 10.2307/2534009

Examples

```
## Not run:
data("train.data")
data("test.data2")

#1.Example: no group in training and test #
formulas <- list(as.formula("Surv(dtime, prostate) ~ AG + WT + PF + HX +HG + SZ + SG + RX"),
                 as.formula("Surv(dtime, cardio) ~ AG + WT + PF + HX +HG + SZ + SG + RX"),
               as.formula("Surv(dtime, Other_causes) ~ AG + WT + PF + HX +HG + SZ + SG + RX"))
```

```
CIF.output <- est.CIF(formulas, data = train.data, newdata = test.data2,
                  group=NULL, compute.CI=T,save=F)
plot(CIF.output)

## End(Not run)
```

`Baseline_Lambda_k` *Baseline cumulative hazards*

Description

Estimate the baseline cumulative hazards for the k^th type of failure.

Usage

```
Baseline_Lambda_k(est, x, y)
```

Arguments

`est`	The model fitting details from the k^th cause-specific hazards model.
`x`	The design matrix associated with the k^th type of failure, of dimension n observations and p_k covariates.
`y`	The survival response associated with the k^th type of failure.

Details

This function estimates the baseline cumulative hazards for the k^th type of failure. The estimated baseline cumulative hazard is a Breslow type estimator.

References

Breslow, Norman E. "Analysis of survival data under the proportional hazards model." International Statistical Review/Revue Internationale de Statistique (1975): 45-57.

`CIF_k` *The k^th cumulative incidence function.*

Description

This function aggregates the hazards table for all type of failures and computes the point estimator of the cumulative incidence function.

Usage

```
CIF_k(Hazard.table, group)
```

Arguments

Hazard.table	An output from Lambda_k(est, x, y, group,...).
group	The name of the group covariates (if any). If specified, the cumulative incidence function will be computed for each group seperately.

Details

This function estimates the cumulative incidence function for the k^th type of failure based on the observed data. When group is specified, the cumulative hazards for the k^th type of failure will be computed for each group seperately.

Value

It returns a data frame contains time of events, cumulative incidence, cumulative hazards for each type of failure, and group indicators (if any).

Compute_Omega *Omega_k*

Description

This function computes Omega_k.

Usage

```
Compute_Omega(est, x, y, kk, event, group, save, group.in.train)
```

Arguments

est	The model fitting details from the k^th cause-specific hazards model.
x	The design matrix associated with the k^th type of failure, of dimension n observations * p_k covariates.
y	The survival response associated with the k^th type of failure.
kk	The k^th type of failure.
event	This is an internal binary indicator to specify if any type of failure occurs at a given time.
group	The name of the group covariates (if any). If specified, the cumulative hazards will be estimated for each group seperately. Default is group=NULL.
save	An option to save the computed S0 and S1. It is highly recommended for large-scale dataset to improve the computational efficiency. Default is save=FALSE.
group.in.train	This argument is valid only when the group argument is specified. If group is presented in both data and newdata, use group.in.train=T; If group is presented in only newdata but not data, use group.in.train=F.

Details

This function computes

$$\hat{\Omega}_k = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{S^{(2)}(\hat{\beta}_k, \tilde{T}_i)}{S^{(0)}(\hat{\beta}_k, \tilde{T}_i)} - \bar{Z}(\hat{\beta}_k, \tilde{T}_i)^{\otimes 2}\right)\delta_{ki}$$

References

Cheng, S. C., Jason P. Fine, and L. J. Wei. "Prediction of cumulative incidence function under the proportional hazards model." Biometrics (1998): 219-228.

`Compute_S01` *S0 and S1*

Description

This function computes S0 and S1 in counting process.

Usage

```
Compute_S01(est, x, y, kk, event, group, save, group.in.train)
```

Arguments

`est`	The model fitting details from the k^th cause-specific hazards model.
`x`	The design matrix associated with the k^th type of failure, of dimension n observations * p_k covariates.
`y`	The survival response associated with the k^th type of failure.
`kk`	The k^th type of failure.
`event`	This is an internal binary indicator to specify if any type of failure occurs at a given time.
`group`	The name of the group covariates (if any). If specified, the cumulative hazards will be estimated for each group seperately. Default is `group=NULL`.
`save`	An option to save the computed S0 and S1. It is highly recommended for large-scale dataset to improve the computational efficiency. Default is `save=FALSE`.
`group.in.train`	This argument is valid only when the group argument is specified. If group is presented in both data and newdata, use `group.in.train=T`. If group is presented in only newdata but not data, use `group.in.train=F`.

Details

This function computes

$$S^{(0,1)}(\hat{\beta}, t; z_0) = \frac{1}{n}\sum_{i'=1}^{n} I(\tilde{T}_{i'} \geq t)\exp(\hat{\beta}^T Z_j) Z_j^{\otimes 0,1}$$

References

Andersen, Per Kragh, et al. Statistical models based on counting processes. Springer Science & Business Media, 2012.

`est.CIF` *Cumulative incidence function.*

Description

This is the main function to obtain the point estimator and pointwise confidence interval for the cumulative incidence function.

Usage

```
est.CIF(formulas, data, group, group.in.train, newdata,
        compute.CI, alpha = 0.05, transform = "log-log", save)
```

Arguments

`formulas`	A list of length K contains formula objects, where K is the number of failures. Each element is a formula object, with the response on the left of a ~ operator, and the predictors on the right. The response must be a survival object as returned by the `Surv` function.
`data`	A data frame with n observations and p number of covariates.
`group`	The name of the group covariates (if any). If specified, the cumulative incidence function will be estimated for each group separately. Default is `group=NULL`.
`group.in.train`	This argument is valid only when the group argument is specified. If group is presented in both data and newdata, use `group.in.train=T`; If group is presented in only newdata but not data, use `group.in.train=F`.
`newdata`	A data frame or a matrix used for prediction. If not specified, the original data will be used instead.
`compute.CI`	A logic operator for whether to compute the pointwise confidence interval. The default is `compute.CI=FALSE`.
`alpha`	The significance level. The default is 0.05, which computes the `95%` pointwise confidence interval.
`transform`	An indicator if transformation is used for the confidence interval. Only log-log transformation is implemented currently.
`save`	An option to save the computed S0 and S1. It is highly recommended for large-scale dataset to improve the computational efficiency. Default is `save=FALSE`.

Details

This is the main function to compute the point estimator and pointwise confidence interval for the cumulative incidence function. We implemented a flexible function to allow prediction of cumulative incidence function under the cause-specific proportion hazards model in scenarios: 1) when group segmentation is not specified in training and test data; 2) when group segmentation is specified in both training and test data; 3) group segmentation is presented in the test data only; 4) the type of failure is associated with a different setts of covariates. To predict, user can input the new values of all covariates as a vector. If the `newdata` is specified, the mean of all observations with corresponding covariates will be used in the prediction. In addition, the generic functions such as `plot()`, `summary()` are available for the newly defined `CIF` class.

Value

When `group=NULL`, `est.CIF` returns a list of length K. For each sublist, it contains:

`model.fit`	Details of model fit, such as parameter estimates, standard error, etc.
`call`	The survival object used for a the k^th type of failure.
`Estimate`	A data frame contains events.time, overall survival, cumulative incidence, upper and lower bounds (if `compute.CI = TRUE`).

When `group` is specified, `est.CIF` returns a list of length that equals to the number of levels in `group` . For each sublist, it contains a list of length K.

References

Cheng, S. C., Jason P. Fine, and L. J. Wei. "Prediction of cumulative incidence function under the proportional hazards model." Biometrics (1998): 219-228.

Examples

```
## Not run:
data("train.data")
data("test.data2")

#1.Example: no group in training and test #
formulas <- list(as.formula("Surv(dtime, prostate) ~ AG + WT + PF + HX +HG + SZ + SG + RX"),
                 as.formula("Surv(dtime, cardio) ~ AG + WT + PF + HX +HG + SZ + SG + RX"),
               as.formula("Surv(dtime, Other_causes) ~ AG + WT + PF + HX +HG + SZ + SG + RX"))
CIF.output <- est.CIF(formulas, data = train.data, newdata = test.data2,
                  group=NULL, compute.CI=T,save=F)
plot(CIF.output)

#2. Example: group in training data and test data #
formulas <- list(as.formula("Surv(dtime, prostate) ~ AG + WT + PF + HX +HG + SZ + SG + RX"),
                 as.formula("Surv(dtime, cardio) ~ AG + WT + PF + HX +HG + SZ + SG + RX"),
               as.formula("Surv(dtime, Other_causes) ~ AG + WT + PF + HX +HG + SZ + SG + RX"))
CIF.output <- est.CIF(formulas, data=train.data, group="RX", group.in.train=T,
                   newdata=test.data2, compute.CI=T, alpha=0.05, transform="log-log", save=F)
plot(CIF.output, group="RX")
```

```
#3. Example: group not in training data and but in test data #
formulas <- list(as.formula("Surv(dtime, prostate) ~ AG + WT + PF + HX +HG + SZ + SG"),
                 as.formula("Surv(dtime, cardio) ~ AG + WT + PF + HX +HG + SZ + SG "),
                as.formula("Surv(dtime, Other_causes) ~ AG + WT + PF + HX +HG + SZ + SG"))
CIF.output <- est.CIF(formulas, data=train.data, group="RX", group.in.train=F, newdata=test.data2,
                      compute.CI=T, alpha=0.05, transform="log-log", save=F)
plot(CIF.output, group="RX")


#4. Example: 1)group not in training data and but in test data
#            2)different risk factors associated with cause of failure #
formulas <- list(as.formula("Surv(dtime, prostate) ~ AG + WT + PF + HX +HG"),
                 as.formula("Surv(dtime, cardio) ~  PF + HX +HG "),
                 as.formula("Surv(dtime, Other_causes) ~ AG + WT + PF + HX +HG + SZ"))
CIF.output <- est.CIF(formulas, data=train.data, group="RX", group.in.train=F,
                  newdata=test.data2,compute.CI=T, alpha=0.05, transform="log-log", save=T)
plot(CIF.output, group="RX")

#5. Example: different risk factors associated with cause of failure #
formulas <- list(as.formula("Surv(dtime, prostate) ~ AG + WT + PF + HX +HG"),
                 as.formula("Surv(dtime, cardio) ~  PF + HX +HG + SZ"),
                as.formula("Surv(dtime, Other_causes) ~ AG + WT + PF + HX +HG + SZ + SG"))
CIF.output <- est.CIF(formulas, data=train.data, group=NULL, group.in.train=F, newdata=test.data2,
                      compute.CI=T, alpha=0.05, transform="log-log", save=F)
plot(CIF.output)

## End(Not run)
```

`Lambda_k` *The cumulative hazards for the k^th type of failure.*

Description

Estimate the cumulative hazards for the k^th type of failure.

Usage

```
Lambda_k(est, x, y, group, data, newdata, group.in.train)
```

Arguments

`est`	The model fitting details from the k^th cause-specific hazards model.
`x`	The design matrix associated with the k^th type of failure, of dimension n observation * p_k covariates.
`y`	The survival response associated with the k^th type of failure.
`group`	The name of the group covariates (if any). If specified, the cumulative hazards will be estimated for each group seperately. Default is `group=NULL`.
`data`	A data frame with n observations and p number of covariates.

newdata	A data frame or a matrix used for prediction. If not specified, the original data will be used instead.
group.in.train	This argument is valid only when the group argument is specified. If group is presented in both data and newdata, use group.in.train=T. If group is presented in only newdata but not data, use group.in.train=F.

Details

This function estimates the cumulative hazards for the k^th type of failure based on the observed data. It also introduces a flexible structure to predict the cumulative hazards for the k^th type of failure for a given new observation. When newdata has more than 2 observations, the mean values of the covariates will be used for prediction. When group is specified, the cumulative hazards for the k^th type of failure will be computed for each group seperately.

Value

It returns a data frame contains time of events, cumulative hazards for each type of failure, and group indicators (if any).

---

plot.CIF *Plots the cumulative incidence function and the pointwise confidence interval (if any).*

Description

Plots the cumulative incidence function and lower and upper pointwise confidence interval (if avaiable) for each group and each type of failure from the CIF class.

Usage

```
## S3 method for class 'CIF'
plot(x, type=NULL,...)
```

Arguments

x	The CIF object generated by est.CIF.
type	The name of the group covariates (if any).
...	other arguments

Psi_par

ψ

Description

This function computes psi_kl(t;z_0).

Usage

```
Psi_par(formulas, kk, ll, CIFest, data, newdata, group, event, save, group.in.train)
```

Arguments

formulas	A list of length K contains formula objects, where K is the number of types of failures. Each element is a formula object, with the response on the left of a ~ operator, and the terms on the right. The response must be a survival object as returned by the Surv function.
kk	The k^th type of failure.
ll	The l^th type of failure.
CIFest	The point estimator of the cumulative incidence function.
data	A data frame with n observations and p number of covariates.
newdata	A data frame or a matrix used for prediction. If not specified, the original data will be used instead.
group	The name of the group covariates (if any). If specified, the cumulative hazards will be estimated for each group seperately. Default is group=NULL.
event	This is an internal binary indicator to specify if any type of failure occurs at a given time.
save	An option to save the computed S0 and S1. It is highly recommended for large-scale dataset to improve the computational efficiency. Default is save=FALSE.
group.in.train	This argument is valid only when the group argument is specified. If group is presented in both data and newdata, use group.in.train=T; If group is presented in only newdata but not data, use group.in.train=F.

Details

This function computes

$$\hat{\psi}_{kl}(t;z_0) = \frac{1}{n}\sum_{i=1}^{n}(\hat{F}_k(t;z_0) - \hat{F}_k(\tilde{T}_i;z_0))(z_0 - \bar{Z}(\hat{\beta}_l,\tilde{T}_i))\frac{\exp(\hat{\beta}_l^T z_0)\delta_{li}I(\tilde{T}_i \le t)}{S^{(0)}(\hat{\beta}_l,\tilde{T}_i)}$$

References

Cheng, S. C., Jason P. Fine, and L. J. Wei. "Prediction of cumulative incidence function under the proportional hazards model." Biometrics (1998): 219-228.

RawVar.CIF *Variance of cumulative incidence function*

Description

This function computes the estimate of variance of the cumulative incidence function.

Usage

```
RawVar.CIF(formulas, CIFest, data, newdata, group, event, save, group.in.train)
```

Arguments

formulas	A list of length K contains formula objects, where K is the number of failures. Each element is a formula object, with the response on the left of a ~ operator, and the predictors on the right. The response must be a survival object as returned by the Surv function.
CIFest	The point estimator of the cumulative incidence function.
data	A data frame with n observations and p number of covariates.
newdata	A data frame or a matrix used for prediction. If not specified, the original data will be used instead.
group	The name of the group covariates (if any). If specified, the cumulative hazards will be estimated for each group seperately. Default is group=NULL.
event	This is an internal binary indicator to specify if any type of failure occurs at a given time.
save	An option to save the computed S0 and S1. It is highly recommended for large-scale dataset to improve the computational efficiency. Default is save=FALSE.
group.in.train	This argument is valid only when the group argument is specified. If group is presented in both data and newdata, use group.in.train=T; If group is presented in only newdata but not data, use group.in.train=F.

Details

This function computes:

$$\hat{\xi}(s,t;z_0) = \frac{1}{n}\sum_{i=1}^{n}[\hat{S}(\tilde{T}_i;z_0) - (\hat{F}_j(t;z_0) - \hat{F}_k(\tilde{T}_i;z_0))]^2 \frac{\exp(2\hat{\beta}_j^T z_0)\delta_{ji} I(\tilde{T}_i \le t)}{(S^{(0)}(\hat{\beta}_j, \tilde{T}_i))^2}$$

References

Cheng, S. C., Jason P. Fine, and L. J. Wei. "Prediction of cumulative incidence function under the proportional hazards model." Biometrics (1998): 219-228.

summary.CIF *Summary for the cumulative incidence function and the pointwise confidence interval.*

Description

Summarize the cumulative incidence function and lower and upper pointwise confidence interval (if avaiable) for each group and each type of failure from the CIF class.

Usage

```
## S3 method for class 'CIF'
summary(object, group, ...)
```

Arguments

object	The CIF object generated by est.CIF.
group	The name of the group covariates (if any).
...	other arguments

Details

For each group and each type of failure, this function returns the cumulative incidence function and lower and upper pointwise confidence interval (if avaiable) for the top 5 and bottom 5 observations.

test.data2 *test data*

Description

This is a subset of data from a prostate cancer trial described in Byar and Green (1980).

References

Byar, D. P., and S. B. Green. "The choice of treatment for cancer patients based on covariate information." Bulletin du cancer 67.4 (1979): 477-490.

Examples

```
data(test.data2)
```

train.data *train data*

Description

This is a subset of data from a prostate cancer trial described in Byar and Green (1980).

References

Byar, D. P., and S. B. Green. "The choice of treatment for cancer patients based on covariate information." Bulletin du cancer 67.4 (1979): 477-490.

Varphi_par φ

Description

This function computes varphi_k(t;z_0).

Usage

```
Varphi_par(est, x, y, kk, CIFest, data, newdata, group, event, save, group.in.train)
```

Arguments

est	The model fitting details from the k^th cause-specific hazards model.
x	The design matrix associated with the k^th type of failure, of dimension n observations * p_k covariates.
y	The survival response associated with the k^th type of failure.
kk	The k^th type of failure.
CIFest	(Internal) The point estimator of the cumulative incidence function.
data	A data frame with n observations and p number of covariates.
newdata	A data frame or a matrix used for prediction. If not specified, the original data will be used instead.
group	The name of the group covariates (if any). If specified, the cumulative hazards will be estimated for each group seperately. Default is group=NULL.
event	This is an internal binary indicator to specify if any type of failure occurs at a given time.
save	An option to save the computed S0 and S1. It is highly recommended for large-scale dataset to improve the computational efficiency. Default is save=FALSE.
group.in.train	This argument is valid only when the group argument is specified. If group is presented in both data and newdata, use group.in.train=T. If group is presented in only newdata but not data, use group.in.train=F.

Details

This function computes

$$\hat{\varphi}_k(t; z_0) = \frac{1}{n} \sum_{i=1}^{n} \hat{S}(\tilde{T}_i; z_0)(z_0 - \bar{Z}(\hat{\beta}_k, \tilde{T}_i)) \frac{\exp(\hat{\beta}_k^T z_0)\delta_{ji} I(\tilde{T}_i \leq t)}{S^{(0)}(\hat{\beta}_k, \tilde{T}_i)}$$

References

Cheng, S. C., Jason P. Fine, and L. J. Wei. "Prediction of cumulative incidence function under the proportional hazards model." Biometrics (1998): 219-228.

Index