

# Package ‘pec’

September 5, 2014

**Title** Prediction Error Curves for Survival Models

**Version** 2.3.7

**Author** Thomas A. Gerds

**Description** Validation of predicted survival probabilities using inverse weighting and resampling.

**Depends** R (>= 1.9.1)

**Imports** prodlim, foreach, rms, survival

**Suggests** timereg, randomForestSRC, party, cmprsk, rpart, doMC, CoxBoost, crrstep, riskRegression, lava, Hmisc

**Maintainer** Thomas A. Gerds <tag@biostat.ku.dk>

**License** GPL (>= 2)

**NeedsCompilation** yes

**Repository** CRAN

**Date/Publication** 2014-09-05 17:51:01

## R topics documented:

|                                |    |
|--------------------------------|----|
| calPlot . . . . .              | 2  |
| cindex . . . . .               | 5  |
| cost . . . . .                 | 9  |
| coxboost . . . . .             | 10 |
| crps . . . . .                 | 11 |
| GBSG2 . . . . .                | 12 |
| ipcw . . . . .                 | 13 |
| Pbc3 . . . . .                 | 15 |
| pec . . . . .                  | 17 |
| pecCforest . . . . .           | 24 |
| plot.pec . . . . .             | 25 |
| plotPredictEventProb . . . . . | 27 |

|                                |    |
|--------------------------------|----|
| plotPredictSurvProb . . . . .  | 29 |
| predictEventProb . . . . .     | 30 |
| predictLifeYearsLost . . . . . | 32 |
| predictSurvProb . . . . .      | 33 |
| print.pec . . . . .            | 35 |
| R2 . . . . .                   | 36 |
| resolvesplitMethod . . . . .   | 38 |
| selectCox . . . . .            | 39 |
| selectFGR . . . . .            | 40 |
| simCost . . . . .              | 41 |
| Special . . . . .              | 42 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>44</b> |
|--------------|-----------|

---

|         |  |
|---------|--|
| calPlot | <i>Calibration plots for right censored data</i> |
|---------|--|

---

## Description

Calibration plots for risk prediction models in right censored survival and competing risks data

## Usage

```
calPlot(object, time, formula, data, splitMethod = "none", B = 1, M,
  outcome = c("pseudo", "prodlim"), showPseudo, pseudo.col = NULL,
  pseudo.pch = NULL, method = "nne", round = TRUE, bandwidth = NULL,
  q = 10, jack.density = 55, add = FALSE, diag = !add, legend = !add,
  axes = !add, xlim, ylim, xlab = "Predicted event probability",
  ylab = "Pseudo-observed event status", col, lwd, lty, pch, cause = 1,
  percent = TRUE, giveToModel = NULL, na.action = na.fail, cores = 1,
  verbose = FALSE, ...)
```

## Arguments

|         |  |
|---------|--|
| object  | A named list of prediction models, where allowed entries are (1) R-objects for which a <a href="#">predictSurvProb</a> method exists (see details), (2) a call that evaluates to such an R-object (see examples), (3) a matrix with predicted probabilities having as many rows as data and as many columns as times. For cross-validation all objects in this list must include their call. |
| time    | The evaluation time point at predicted event probabilities are plotted against pseudo-observed event status.   |
| formula | A survival or event history formula. The left hand side is used to compute the expected event status. If formula is missing, try to extract a formula from the first element in object.  |
| data    | A data frame in which to validate the prediction models and to fit the censoring model. If data is missing, try to extract a data set from the first element in object.  |

|              |  |
|--------------|--|
| splitMethod  | <p>Defines the internal validation design:</p> <p>none/noPlan: Assess the models in the give data, usually either in the same data where they are fitted, or in independent test data.</p> <p>BootCv: Bootstrap cross validation. The prediction models are trained on B bootstrap samples, that are either drawn with replacement of the same size as the original data or without replacement from data of the size M. The models are assessed in the observations that are NOT in the bootstrap sample.</p> |
| B            | The number of cross-validation steps.  |
| M            | The size of the subsamples for cross-validation.   |
| outcome      | <p>The method for estimating expected event status:</p> <p>"pseudo": Use average pseudo-values. "prodlim": Use the product-limit estimate, i.e., apply the Kaplan-Meier method for right censored survival and the Aalen-Johansen method for right censored competing risks data.</p>  |
| showPseudo   | If TRUE and outcome=="pseudo" the pseudo-values are shown as dots on the plot.   |
| pseudo.col   | Colour for pseudo-values.  |
| pseudo.pch   | Dot type (see par) for pseudo-values.  |
| method       | <p>The method for estimating the calibration curve(s):</p> <p>"nne": The expected event status is obtained in the nearest neighborhood around the predicted event probabilities.</p> <p>"quantile": The expected event status is obtained in groups defined by quantiles of the predicted event probabilities.</p>   |
| round        | If TRUE predicted probabilities are rounded to two digits before smoothing. This may have a considerable effect on computing efficiency in large data sets.  |
| bandwidth    | The bandwidth for method="nne"   |
| q            | The number of quantiles for method="quantile".   |
| jack.density | Gray scale for pseudo-observations.  |
| add          | If TRUE the line(s) are added to an existing plot.   |
| diag         | If FALSE no diagonal line is drawn.  |
| legend       | If FALSE no legend is drawn.   |
| axes         | If FALSE no axes are drawn.  |
| xlim         | Limits of x-axis.  |
| ylim         | Limits of y-axis.  |
| xlab         | Label for y-axis.  |
| ylab         | Label for x-axis.  |
| col          | Vector with colors, one for each element of object. Passed to <a href="#">lines</a> .  |
| lwd          | Vector with line widths, one for each element of object. Passed to <a href="#">lines</a> .   |
| lty          | lwd Vector with line style, one for each element of object. Passed to <a href="#">lines</a> .  |
| pch          | Passed to <a href="#">points</a> .   |
| cause        | For competing risks models, the cause of failure or event of interest  |

|             |  |
|-------------|--|
| percent     | If TRUE axes labels are multiplied by 100 and thus interpretable on a percent scale.   |
| giveToModel | List of with exactly one entry for each entry in object. Each entry names parts of the value of the fitted models that should be extracted and added to the value. |
| na.action   | Passed to <code>model.frame</code>   |
| cores       | Number of cores for parallel computing. Passed as value of argument <code>mc.cores</code> to <code>mclapply</code> .   |
| verbose     | if TRUE report details of the progress, e.g. count the steps in cross-validation.  |
| ...         | Used to control the subroutines: plot, axis, lines, legend. See <code>SmartControl</code> .  |

### Details

For method "nne" the optimal bandwidth with respect to is obtained with the function `dpik` from the package `KernSmooth` for a box kernel function.

### Value

list with elements: `time`, `pseudoFrame` and `bandwidth` (NULL for method `quantile`).

### Author(s)

Thomas Alexander Gerds <tag@biostat.ku.dk>

### Examples

```
library(prodlm)
library(lava)
library(riskRegression)
library(survival)
set.seed(13)
m <- crModel()
regression(m, from = "X1", to = "eventtime1") <- 1
regression(m, from = "X2", to = "eventtime1") <- 1
m <- addvar(m,c("X3","X4","X5"))
distribution(m, "X1") <- binomial.lvm()
distribution(m, "X4") <- binomial.lvm()
d1 <- sim(m,100)
d2 <- sim(m,100)
csc <- CSC(Hist(time,event)~X1+X2+X3+X4+X5,data=d1)
fgr <- FGR(Hist(time,event)~X1+X2+X3+X4+X5,data=d1,cause=1)
predict.crr <- cmprsk:::predict.crr
par(mar=c(5,5,5,5),cex=1.3)
calPlot(list(csc,fgr),
         time=5,
         legend.x=-0.3,
         legend.y=1.35,
         ylab="Observed event status",
         legend.legend=c("Cause-specific Cox regression","Fine-Gray regression"),
         legend.xpd=NA)
```

---

|        |  |
|--------|--|
| cindex | <i>Concordance index for right censored survival time data</i> |
|--------|--|

---

## Description

In survival analysis, a pair of patients is called concordant if the risk of the event predicted by a model is lower for the patient who experiences the event at a later timepoint. The concordance probability (C-index) is the frequency of concordant pairs among all pairs of subjects. It can be used to measure and compare the discriminative power of a risk prediction models. The function provides an inverse of the probability of censoring weighed estimate of the concordance probability to adjust for right censoring. Cross-validation based on bootstrap resampling or bootstrap subsampling can be applied to assess and compare the discriminative power of various regression modelling strategies on the same set of data.

## Usage

```
cindex(object, formula, data, eval.times, pred.times, cause, lyl = FALSE,
  cens.model = "marginal", ipcw.refit = FALSE, ipcw.args = NULL,
  ipcw.limit, tiedPredictionsIn = TRUE, tiedOutcomeIn = TRUE,
  tiedMatchIn = TRUE, splitMethod = "noPlan", B, M, model.args = NULL,
  model.parms = NULL, keep.models = "Call", keep.residuals = FALSE,
  keep.pvalues = FALSE, keep.weights = FALSE, keep.index = FALSE,
  keep.matrix = FALSE, multiSplitTest = FALSE, testTimes, confInt = FALSE,
  confLevel = 0.95, verbose = TRUE, savePath = NULL, slaveseed = NULL,
  na.action = na.fail, ...)
```

## Arguments

- |         |  |
|---------|--|
| object  | A named list of prediction models, where allowed entries are (1) R-objects for which a <a href="#">predictSurvProb</a> method exists (see details), (2) a call that evaluates to such an R-object (see examples), (3) a matrix with predicted probabilities having as many rows as data and as many columns as times. For cross-validation all objects in this list must include their call.   |
| formula | A survival formula. The left hand side is used to find the status response variable in data. For right censored data, the right hand side of the formula is used to specify conditional censoring models. For example, set <code>Surv(time, status)~x1+x2</code> and <code>cens.model="cox"</code> . Then the weights are based on a Cox regression model for the censoring times with predictors <code>x1</code> and <code>x2</code> . Note that the usual coding is assumed: <code>status=0</code> for censored times and that each variable name that appears in <code>formula</code> must be the column name in <code>data</code> . If there are no covariates, i.e. <code>formula=Surv(time,status)~1</code> the <code>cens.model</code> is coerced to "marginal" and the Kaplan-Meier estimator for the censoring times is used to calculate the weights. If <code>formula</code> is missing, try to extract a formula from the first element in <code>object</code> . |
| data    | A data frame in which to validate the prediction models and to fit the censoring model. If <code>data</code> is missing, try to extract a data set from the first element in <code>object</code> .   |

|                                |  |
|--------------------------------|--|
| <code>eval.times</code>        | A vector of timepoints for evaluating the discriminative ability. At each timepoint the c-index is computed using only those pairs where one of the event times is known to be earlier than this timepoint. If <code>eval.times</code> is missing or <code>Inf</code> then the largest uncensored event time is used.  |
| <code>pred.times</code>        | A vector of timepoints for evaluating the prediction models. This should either be exactly one timepoint used for all <code>eval.times</code> , or be as long as <code>eval.times</code> , in which case the predicted order of risk for the <code>j</code> th entry of <code>eval.times</code> is based on the <code>j</code> th entry of <code>pred.times</code> corresponding   |
| <code>cause</code>             | For competing risks, the event of interest. Defaults to the first state of the response, which is obtained by evaluating the left hand side of formula in <code>data</code> .  |
| <code>lyl</code>               | If TRUE rank subjects according to predicted life-years-lost (See Andersen due to this cause instead of predicted risk.  |
| <code>cens.model</code>        | Method for estimating inverse probability of censoring weights:<br><code>cox</code> : A semi-parametric Cox proportional hazard model is fitted to the censoring times<br><code>marginal</code> : The Kaplan-Meier estimator for the censoring times<br><code>nonpar</code> : Nonparametric extension of the Kaplan-Meier for the censoring times using symmetric nearest neighborhoods – available for arbitrary many strata variables on the right hand side of argument formula but at most one continuous variable. See the documentation of the functions <code>prodlim</code> and <code>neighborhood</code> from the <code>prodlim</code> package.<br><code>aalen</code> : The nonparametric Aalen additive model fitted to the censoring times. Requires the <code>timereg</code> package maintained by Thomas Scheike. |
| <code>ipcw.refit</code>        | If TRUE the inverse probability of censoring weights are estimated separately in each training set during cross-validation.  |
| <code>ipcw.args</code>         | List of arguments passed to function specified by argument <code>cens.model</code> .   |
| <code>ipcw.limit</code>        | Value between 0 and 1 (but no equal to 0!) used to cut for small weights in order to stabilize the estimate at late times where few individuals are observed.  |
| <code>tiedPredictionsIn</code> | If FALSE pairs with identical predictions are excluded, unless also the event times are identical and uncensored and <code>tiedMatchIn</code> is set to TRUE.  |
| <code>tiedOutcomeIn</code>     | If TRUE pairs with identical and uncensored event times are excluded, unless also the predictions are identical and <code>tiedMatchIn</code> is set to TRUE.   |
| <code>tiedMatchIn</code>       | If TRUE then pairs with identical predictions and identical and uncensored event times are counted as concordant pairs.  |
| <code>splitMethod</code>       | Defines the internal validation design:<br><code>none/noPlan</code> : Assess the models in the give data, usually either in the same data where they are fitted, or in independent test data.<br><code>BootCv</code> : Bootstrap cross validation. The prediction models are trained on <code>B</code> bootstrap samples, that are either drawn with replacement of the same size as the original data or without replacement from data of the size <code>M</code> . The models are assessed in the observations that are NOT in the bootstrap sample.<br><code>Boot632</code> : Linear combination of <code>AppCindex</code> and <code>OutOfBagCindex</code> using the constant weight <code>.632</code> .  |

|                             |  |
|-----------------------------|--|
| B                           | Number of bootstrap samples. The default depends on argument <code>splitMethod</code> . When <code>splitMethod</code> in <code>c("BootCv", "Boot632")</code> the default is 100. For <code>splitMethod="none"</code> B is the number of bootstrap simulations e.g. to obtain bootstrap confidence limits – default is 0. |
| M                           | The size of the bootstrap samples for resampling without replacement. Ignored for resampling with replacement.   |
| <code>model.args</code>     | List of extra arguments that can be passed to the <code>predictSurvProb</code> methods. The list must have an entry for each entry in <code>object</code> .  |
| <code>model.parms</code>    | Experimental. List of with exactly one entry for each entry in <code>object</code> . Each entry names parts of the value of the fitted models that should be extracted and added to the value.   |
| <code>keep.index</code>     | Logical. If FALSE remove the bootstrap or cross-validation index from the output list which otherwise is included in the method part of the output list.   |
| <code>keep.matrix</code>    | Logical. If TRUE add all B prediction error curves from bootstrapping or cross-validation to the output.   |
| <code>keep.models</code>    | Logical. If TRUE keep the models in <code>object</code> . If "Call" keep only the call of these models.  |
| <code>keep.residuals</code> | Experimental.  |
| <code>keep.pvalues</code>   | Experimental.  |
| <code>keep.weights</code>   | Experimental.  |
| <code>multiSplitTest</code> | Experimental.  |
| <code>testTimes</code>      | A vector of time points for testing differences between models in the time-point specific Brier scores.  |
| <code>confInt</code>        | Experimental.  |
| <code>confLevel</code>      | Experimental.  |
| <code>verbose</code>        | if TRUE report details of the progress, e.g. count the steps in cross-validation.  |
| <code>savePath</code>       | Place in your filesystem (directory) where training models fitted during cross-validation are saved. If missing training models are not saved.   |
| <code>slaveseed</code>      | Vector of seeds, as long as B, to be given to the slaves in parallel computing.  |
| <code>na.action</code>      | Passed immediately to <code>model.frame</code> . Defaults to <code>na.fail</code> . If set otherwise most prediction models will not work.   |
| ...                         | Not used.  |

### Details

Pairs with identical observed times, where one is uncensored and one is censored, are always considered usable (independent of the value of `tiedOutcomeIn`), as it can be assumed that the event occurs at a later timepoint for the censored observation.

For uncensored response the result equals the one obtained with the functions `rcorr.cens` and `rcorrcens` from the `Hmisc` package (see examples).

### Value

Estimates of the C-index.

**Author(s)**

Thomas A Gerds <tag@biostat.ku.dk>

**References**

TA Gerds, MW Kattan, M Schumacher, and C Yu. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in Medicine*, Ahead of print:to appear, 2013. DOI = 10.1002/sim.5681

Wolbers, M and Koller, MT and Wittelman, JCM and Gerds, TA (2013) Concordance for prognostic models with competing risks Research report 13/3. Department of Biostatistics, University of Copenhagen

Andersen, PK (2012) A note on the decomposition of number of life years lost according to causes of death Research report 12/2. Department of Biostatistics, University of Copenhagen

**Examples**

```
# simulate data based on Weibull regression
library(prodlim)
set.seed(13)
dat <- SimSurv(100)
# fit three different Cox models and a random survival forest
# note: low number of trees for the purpose of illustration
library(survival)
library(randomForestSRC)
cox12 <- coxph(Surv(time,status)~X1+X2,data=dat)
cox1 <- coxph(Surv(time,status)~X1,data=dat)
cox2 <- coxph(Surv(time,status)~X2,data=dat)
rsf1 <- rfsrc(Surv(time,status)~X1+X2,data=dat,ntree=15,forest=TRUE)
#
# compute the apparent estimate of the C-index at different time points
#
ApparrentCindex <- cindex(list("Cox X1"=cox1,
                             "Cox X2"=cox2,
                             "Cox X1+X2"=cox12,
                             "RSF"=rsf1),
                          formula=Surv(time,status)~X1+X2,
                          data=dat,
                          eval.times=seq(5,500,50))
print(ApparrentCindex)
plot(ApparrentCindex)
#
# compute the bootstrap-crossvalidation estimate of
# the C-index at different time points
#
set.seed(142)
bcvCindex <- cindex(list("Cox X1"=cox1,
                        "Cox X2"=cox2,
                        "Cox X1+X2"=cox12,
                        "RSF"=rsf1),
                    formula=Surv(time,status)~X1+X2,
                    data=dat,
```



```

                                splitMethod="bootcv",
                                B=5,
                                eval.times=seq(5,500,50))
                                print(bcvCindex)
                                plot(bcvCindex)
                                # for uncensored data the results are the same
                                # as those obtained with the function rcorr.cens from Hmisc
                                library(Hmisc)
                                set.seed(16)
                                dat <- SimSurv(30,cens=FALSE)
                                fit12 <- coxph(Surv(time,status)~X1+X2,data=dat)
                                fit1 <- coxph(Surv(time,status)~X1,data=dat)
                                fit2 <- coxph(Surv(time,status)~X2,data=dat)
                                Cpec <- cindex(list("Cox X1+X2"=fit12,"Cox X1"=fit1,"Cox X2"=fit2),
                                                formula=Surv(time,status)~1,
                                                data=dat,
                                                eval.times=Inf)
                                p1 <- predictSurvProb(fit1,newdata=dat,times=100)
                                p2 <- predictSurvProb(fit2,newdata=dat,times=100)
                                p12 <- predictSurvProb(fit12,newdata=dat,times=100)
                                harrelC1 <- rcorr.cens(p1,with(dat,Surv(time,status)))
                                harrelC2 <- rcorr.cens(p2,with(dat,Surv(time,status)))
                                harrelC12 <- rcorr.cens(p12,with(dat,Surv(time,status)))
                                harrelC1[["C Index"]]==Cpec$AppCindex[["Cox.X1"]]
                                harrelC2[["C Index"]]==Cpec$AppCindex[["Cox.X2"]]
                                harrelC12[["C Index"]]==Cpec$AppCindex[["Cox.X1.X2"]]
                                #
                                # competing risks
                                #
                                library(riskRegression)
                                library(prodlim)
                                set.seed(30)
                                dcr.learn <- SimCompRisk(30)
                                dcr.val <- SimCompRisk(30)
                                cindex(CSC(Hist(time,event)~X1+X2,data=dcr.learn),data=dcr.val)

```

---

 cost

*Copenhagen Stroke Study*


---

## Description

This data set contains a subset of the data from the Copenhagen stroke study.

## Format

This data frame contains the observations of 518 stroke patients :

**age** Age of the patients in years.

**sex** A factor with two levels female and male.

**hypTen** Hypertension, a factor with two levels no and yes.

- ihd** History of ischemic heart disease at admission, a factor with two levels no and yes.
- prevStroke** History of previous strokes before admission, a factor with two levels no and yes.
- othDisease** History of other disabling diseases (e.g. severe dementia), a factor with two levels no and yes.
- alcohol** Daily alcohol consumption, a factor with two levels no and yes.
- diabetes** Diabetes mellitus status indicating if the glucose level was higher than 11 mmol/L, a factor with two levels no and yes.
- smoke** Daily smoking status, a factor with two levels no and yes.
- atrialFib** Atrial fibrillation, a factor with two levels no and yes.
- hemor** Hemorrhage (stroke subtype), a factor with two levels no (infarction) and yes (hemorrhage).
- strokeScore** Scandinavian stroke score at admission to the hospital. Ranges from 0 (worst) to 58 (best).
- cholest** Cholesterol level
- time** Survival time (in days).
- status** Status (0: censored, 1: event).

## References

- Joergensen HS, Nakayama H, Reith J, Raaschou HO, and Olsen TS. Acute stroke with atrial fibrillation. The Copenhagen Stroke Study. *Stroke*, 27(10):1765-9, 1996.
- Mogensen UB, Ishwaran H, and Gerds TA. Evaluating random forests for survival analysis using prediction error curves. Technical Report 8, University of Copenhagen, Department of Biostatistics, 2010.

---

coxboost

*Formula interface for function CoxBoost of package CoxBoost.*

---

## Description

Formula interface for function CoxBoost of package CoxBoost.

## Usage

```
coxboost(formula, data, cv = TRUE, cause = 1, penalty, ...)
```

## Arguments

- |         |  |
|---------|--|
| formula | An event-history formula for competing risks of the form <code>Hist(time, status)~sex+age</code> where <code>status</code> defines competing events and right censored data. The code for right censored can be controlled with argument <code>cens.code</code> , see man page the function <a href="#">Hist</a> . |
| data    | A <code>data.frame</code> in which the variables of formula are defined.   |

|         |  |
|---------|--|
| cv      | If TRUE perform cross-validation to optimize the parameter <code>stepno</code> . This calls the function <code>cv.CoxBoost</code> whose arguments are prefix controlled, that is <code>cv.K=7</code> sets the argument <code>K</code> of <code>cv.CoxBoost</code> to 7. If FALSE use <code>stepno</code> . |
| cause   | The cause of interest in competing risk models.  |
| penalty | See <code>CoxBoost</code> .  |
| ...     | Arguments passed to either <code>CoxBoost</code> via <code>CoxBoost.arg</code> or to <code>cv.CoxBoost</code> via <code>cv.CoxBoost.arg</code> .   |

**Details**

See `CoxBoost`.

**Value**

See `CoxBoost`.

**Author(s)**

Thomas Alexander Gerds <tag@biostat.ku.dk>

**References**

See `CoxBoost`.

**See Also**

See `CoxBoost`.

---

crps

*Summarizing prediction error curves*


---

**Description**

Computes the cumulative prediction error curves, aka integrated Brier scores, in ranges of time.

**Usage**

```
crps(object, models, what, times, start)
```

**Arguments**

|        |  |
|--------|--|
| object | An object with estimated prediction error curves obtained with the function <code>pec</code>   |
| models | Which models in <code>object\$models</code> should be considered.  |
| what   | The name of the entry in <code>x</code> to be cumulated. Defaults to <code>PredErr</code> Other choices are <code>AppErr</code> , <code>BootCvErr</code> , <code>Boot632</code> , <code>Boot632plus</code> . |
| times  | Time points at which the integration of the prediction error curve stops.  |
| start  | The time point at which the integration of the prediction error curve is started.  |

**Details**

The cumulative prediction error (continuous ranked probability score) is defined as the area under the prediction error curve, hence the alias name, `ibs`, which is short for integrated Brier score.

**Value**

A matrix with a column for the `crps` (`ibs`) at every requested time point and a row for each model

**Author(s)**

Thomas A. Gerds <tag@biostat.ku.dk>

**References**

E. Graf et al. (1999), Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, vol 18, pp= 2529–2545.

Gerds TA, Cai T & Schumacher M (2008) The performance of risk prediction models *Biometrical Journal*, 50(4), 457–479

**See Also**

[pec](#)

**Examples**

```
set.seed(18713)
library(prodlim)
library(survival)
dat=SimSurv(100)
pmodel=coxph(Surv(time,status)~X1+X2,data=dat)
perror=pec(list(Cox=pmodel),Hist(time,status)~1,data=dat)

## cumulative prediction error
crps(perror,times=1) # between min time and 1
## same thing:
ibs(perror,times=1) # between min time and 1
crps(perror,times=1,start=0) # between 0 and 1
crps(perror,times=seq(0,1,.2),start=0) # between 0 and seq(0,1,.2)
```

**Description**

A data frame containing the observations from the GBSG2 study.

**Format**

This data frame contains the observations of 686 women:

**horTh** hormonal therapy, a factor at two levels no and yes.

**age** of the patients in years.

**menostat** menopausal status, a factor at two levels pre (premenopausal) and post (postmenopausal).

**tsize** tumor size (in mm).

**tgrade** tumor grade, a ordered factor at levels I < II < III.

**pnodes** number of positive nodes.

**progrec** progesterone receptor (in fmol).

**estrec** estrogen receptor (in fmol).

**time** recurrence free survival time (in days).

**cens** censoring indicator (0- censored, 1- event).

**Source**

<http://www.blackwellpublishers.com/rss/Volumes/A162p1.htm>

**References**

M. Schumacher, G. Basert, H. Bojar, K. Huebner, M. Olschewski, W. Sauerbrei, C. Schmoor, C. Beyerle, R.L.A. Neumann and H.F. Rauschecker for the German Breast Cancer Study Group (1994), Randomized  $2 \times 2$  trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. *Journal of Clinical Oncology*, **12**, 2086–2093.

---

ipcw

*Estimation of censoring probabilities*

---

**Description**

This function is used internally by the function pec to obtain inverse of the probability of censoring weights.

**Usage**

```
ipcw(formula, data, method, args, times, subjectTimes, subjectTimesLag = 1,
      what)
```

**Arguments**

|                 |  |
|-----------------|--|
| formula         | A survival formula like, $\text{Surv}(\text{time}, \text{status}) \sim 1$ , where as usual $\text{status}=0$ means censored. The status variable is internally reversed for estimation of censoring rather than survival probabilities. Some of the available models (see argument <code>model</code> ) will use predictors on the right hand side of the formula. |
| data            | The data used for fitting the censoring model  |
| method          | Censoring model used for estimation of the (conditional) censoring distribution.   |
| args            | A list of arguments which is passed to <code>method</code>   |
| times           | For <code>what="IPCW.times"</code> a vector of times at which to compute the probabilities of not being censored.  |
| subjectTimes    | For <code>what="IPCW.subjectTimes"</code> a vector of individual times at which the probabilities of not being censored are computed.  |
| subjectTimesLag | If equal to 1 then obtain $G(T_i -  X_i)$ , if equal to 0 estimate the conditional censoring distribution at the <code>subjectTimes</code> , i.e. $(G(T_i   X_i))$ .   |
| what            | Decide about what to do: If equal to <code>"IPCW.times"</code> then weights are estimated at given times. If equal to <code>"IPCW.subjectTimes"</code> then weights are estimated at individual <code>subjectTimes</code> . If missing then produce both.  |

**Details**

Inverse of the probability of censoring weights (IPCW) usually refer to the probabilities of not being censored at certain time points. These probabilities are also the values of the conditional survival function of the censoring time given covariates. The function `ipcw` estimates the conditional survival function of the censoring times and derives the weights.

**IMPORTANT:** the data set should be ordered, `order(time, -status)` in order to get the values `IPCW.subjectTimes` in the right order for some choices of `method`.

**Value**

|                   |   |
|-------------------|---|
| times             | The times at which weights are estimated                              |
| IPCW.times        | Estimated weights at times  |
| IPCW.subjectTimes | Estimated weights at individual time values <code>subjectTimes</code> |
| fit               | The fitted censoring model  |
| method            | The method for modelling the censoring distribution                   |
| call              | The call  |

**Author(s)**

Thomas A. Gerds <tag@biostat.ku.dk>

**See Also**

[pec](#)

**Examples**

```

library(problim)
dat=SimSurv(30)

dat <- dat[order(dat$time),]

# using the marginal Kaplan-Meier for the censoring times

WKM=ipcw(Hist(time,status)~X2,
  data=dat,
  method="marginal",
  times=sort(unique(dat$time)),
  subjectTimes=dat$time)
plot(WKM$fit)
WKM$fit

# using the Cox model for the censoring times given X2
library(survival)
WCox=ipcw(Hist(time=time,event=status)~X2,
  data=dat,
  method="cox",
  times=sort(unique(dat$time)),
  subjectTimes=dat$time)
WCox$fit

plot(WKM$fit)
lines(sort(unique(dat$time)),
  1-WCox$IPCW.times[1,],
  type="l",
  col=2,
  lty=3,
  lwd=3)
lines(sort(unique(dat$time)),
  1-WCox$IPCW.times[5,],
  type="l",
  col=3,
  lty=3,
  lwd=3)

# using the stratified Kaplan-Meier
# for the censoring times given X2

WKM2=ipcw(Hist(time,status)~X2,
  data=dat,
  method="nonpar",
  times=sort(unique(dat$time)),
  subjectTimes=dat$time)
plot(WKM2$fit,add=FALSE)

```

## Description

PBC3 was a multi-centre randomized clinical trial conducted in six European hospitals. Between 1 Jan. 1983 and 1 Jan. 1987, 349 patients with the liver disease primary biliary cirrhosis (PBC) were randomized to either treatment with Cyclosporin A (CyA, 176 patients) or placebo (173 patients). The purpose of the trial was to study the effect of treatment on the survival time. However, during the course of the trial an increased use of liver transplantation for patients with this disease made the investigators redefine the main response variable to be time to “failure of medical treatment” defined as either death or liver transplantation. Patients were then followed from randomization until treatment failure, drop-out or 1 Jan, 1989; 61 patients died (CyA: 30, placebo: 31), another 29 were transplanted (CyA: 14, placebo: 15) and 4 patients were lost to follow-up before 1 Jan. 1989. At entry a number of clinical, biochemical and histological variables, including serum bilirubin, serum albumin, sex, age were recorded.

## Format

A data frame with 349 observations on the following 15 variables.

**ptno** patient identification  
**unit** hospital (1: Hvidovre, 2: London, 3: Copenhagen, 4: Barcelona, 5: Munich, 6: Lyon)  
**tment** treatment (0: placebo, 1: CyA)  
**sex** (1: males, 0: females)  
**age** age in years  
**stage** histological stage (1, 2, 3, 4)  
**gibleed** previous gastrointestinal bleeding (1: yes, 0: no)  
**crea** creatinine (micromoles/L)  
**alb** albumin (g/L)  
**bili** bilirubin (micromoles/L)  
**alkph** alkaline phosphatase (IU/L)  
**asptr** aspartate transaminase (IU/L)  
**weight** body weight (kg)  
**days** observation time (days)  
**status** status at observation time (0: censored, 1: liver transplantation, 2 : dead)

## Source

<http://192.38.117.59/~linearpredictors/?page=datasets&dataset=Pbc3>

## References

Andersen and Skovgaard. Regression with linear predictors. Springer, 2010.

## Examples

```
data(Pbc3)
```



**Description**

Evaluating the performance of risk prediction models in survival analysis. The Brier score is a weighted average of the squared distances between the observed survival status and the predicted survival probability of a model. Roughly the weights correspond to the probabilities of not being censored. The weights can be estimated depend on covariates. Prediction error curves are obtained when the Brier score is followed over time. Cross-validation based on bootstrap resampling or bootstrap subsampling can be applied to assess and compare the predictive power of various regression modelling strategies on the same set of data.

**Usage**

```
pec(object, formula, data, times, cause, start, maxtime, exact = TRUE,
    exactness = 100, fillChar = NA, cens.model = "cox",
    ipcw.refit = FALSE, ipcw.args = NULL, splitMethod = "none", B, M,
    reference = TRUE, model.args = NULL, model.parms = NULL,
    keep.index = FALSE, keep.matrix = FALSE, keep.models = "Call",
    keep.residuals = FALSE, keep.pvalues = FALSE, noinf.permute = FALSE,
    multiSplitTest = FALSE, testIBS, testTimes, confInt = FALSE,
    confLevel = 0.95, verbose = TRUE, savePath = NULL, slaveseed = NULL,
    na.action = na.fail, ...)
```

**Arguments**

- |         |  |
|---------|--|
| object  | A named list of prediction models, where allowed entries are (1) R-objects for which a <code>predictSurvProb</code> method exists (see details), (2) a <code>call</code> that evaluates to such an R-object (see examples), (3) a matrix with predicted probabilities having as many rows as data and as many columns as times. For cross-validation all objects in this list must include their <code>call</code> .   |
| formula | A survival formula. The left hand side is used to find the status response variable in data. For right censored data, the right hand side of the formula is used to specify conditional censoring models. For example, set <code>Surv(time, status)~x1+x2</code> and <code>cens.model="cox"</code> . Then the weights are based on a Cox regression model for the censoring times with predictors <code>x1</code> and <code>x2</code> . Note that the usual coding is assumed: <code>status=0</code> for censored times and that each variable name that appears in formula must be the column name in data. If there are no covariates, i.e. <code>formula=Surv(time,status)~1</code> the <code>cens.model</code> is coerced to "marginal" and the Kaplan-Meier estimator for the censoring times is used to calculate the weights. If formula is missing, try to extract a formula from the first element in object. |
| data    | A data frame in which to validate the prediction models and to fit the censoring model. If data is missing, try to extract a data set from the first element in object.  |

|                          |  |
|--------------------------|--|
| <code>times</code>       | A vector of time points. At each time point the prediction error curves are estimated. If <code>exact==TRUE</code> the times are merged with all the unique values of the response variable. If <code>times</code> is missing and <code>exact==TRUE</code> all the unique values of the response variable are used. If missing and <code>exact==FALSE</code> use an equidistant grid of values between <code>start</code> and <code>maxtime</code> . The distance is determined by <code>exactness</code> .  |
| <code>cause</code>       | For competing risks, the event of interest. Defaults to the first state of the response, which is obtained by evaluating the left hand side of <code>formula</code> in <code>data</code> .   |
| <code>start</code>       | Minimal time for estimating the prediction error curves. If missing and <code>formula</code> defines a <code>Surv</code> or <code>Hist</code> object then <code>start</code> defaults to 0, otherwise to the smallest observed value of the response variable. <code>start</code> is ignored if <code>times</code> are given.  |
| <code>maxtime</code>     | Maximal time for estimating the prediction error curves. If missing the largest value of the response variable is used.  |
| <code>exact</code>       | Logical. If <code>TRUE</code> estimate the prediction error curves at all the unique values of the response variable. If <code>times</code> are given and <code>exact=TRUE</code> then the <code>times</code> are merged with the unique values of the response variable.  |
| <code>exactness</code>   | An integer that determines how many equidistant gridpoints are used between <code>start</code> and <code>maxtime</code> . The default is 100.  |
| <code>fillChar</code>    | Symbol used to fill-in places where the values of the prediction error curves are not available. The default is NA.  |
| <code>cens.model</code>  | Method for estimating inverse probability of censoring weights:<br><code>cox</code> : A semi-parametric Cox proportional hazard model is fitted to the censoring times<br><code>marginal</code> : The Kaplan-Meier estimator for the censoring times<br><code>nonpar</code> : Nonparametric extension of the Kaplan-Meier for the censoring times using symmetric nearest neighborhoods – available for arbitrary many strata variables on the right hand side of argument <code>formula</code> but at most one continuous variable. See the documentation of the functions <code>prodlim</code> and <code>neighborhood</code> from the <code>prodlim</code> package.<br><code>aalen</code> : The nonparametric Aalen additive model fitted to the censoring times. Requires the <code>timereg</code> package.                       |
| <code>ipcw.refit</code>  | If <code>TRUE</code> the inverse probability of censoring weights are estimated separately in each training set during cross-validation.   |
| <code>ipcw.args</code>   | List of arguments passed to function specified by argument <code>cens.model</code> .   |
| <code>splitMethod</code> | SplitMethod for estimating the prediction error curves.<br><code>none/noPlan</code> : Assess the models in the same data where they are fitted. <code>boot</code> : DEPRECATED.<br><code>cvK</code> : K-fold cross-validation, i.e. <code>cv10</code> for 10-fold cross-validation. After splitting the data in K subsets, the prediction models (ie those specified in <code>object</code> ) are evaluated on the data omitting the Kth subset (training step). The prediction error is estimated with the Kth subset (validation step).<br>The random splitting is repeated B times and the estimated prediction error curves are obtained by averaging.<br><code>BootCv</code> : Bootstrap cross validation. The prediction models are trained on B bootstrap samples, that are either drawn with replacement of the same size as |

the original data or without replacement from data of the size  $M$ . The models are assessed in the observations that are NOT in the bootstrap sample.

Boot632: Linear combination of AppErr and BootCvErr using the constant weight .632.

Boot632plus: Linear combination of AppErr and BootCv using weights dependent on how the models perform in permuted data.

loocv: Leave one out cross-validation.

NoInf: Assess the models in permuted data.

|                |  |
|----------------|--|
| B              | Number of bootstrap samples. The default depends on argument <code>splitMethod</code> . When <code>splitMethod</code> in <code>c("BootCv", "Boot632", "Boot632plus")</code> the default is 100. For <code>splitMethod="cvK"</code> B is the number of cross-validation cycles, and – default is 1. For <code>splitMethod="none"</code> B is the number of bootstrap simulations e.g. to obtain bootstrap confidence limits – default is 0. |
| M              | The size of the bootstrap samples for resampling without replacement. Ignored for resampling with replacement.   |
| reference      | Logical. If TRUE add the marginal Kaplan-Meier prediction model as a reference to the list of models.  |
| model.args     | List of extra arguments that can be passed to the <code>predictSurvProb</code> methods. The list must have an entry for each entry in object.  |
| model.parms    | Experimental. List of with exactly one entry for each entry in object. Each entry names parts of the value of the fitted models that should be extracted and added to the value.   |
| keep.index     | Logical. If FALSE remove the bootstrap or cross-validation index from the output list which otherwise is included in the <code>splitMethod</code> part of the output list.   |
| keep.matrix    | Logical. If TRUE add all B prediction error curves from bootstrapping or cross-validation to the output.   |
| keep.models    | Logical. If TRUE keep the models in object. If "Call" keep only the call of these models.  |
| keep.residuals | Logical. If TRUE keep the patient individual residuals at <code>testTimes</code> .   |
| keep.pvalues   | For <code>multiSplitTest</code> . If TRUE keep the pvalues from the single splits.   |
| noinf.permute  | If TRUE the noinformation error is approximated using permutation.   |
| multiSplitTest | If TRUE the test proposed by van de Wiel et al. (2009) is applied. Requires subsampling bootstrap cross-validation, i.e. that <code>splitMethod</code> equals <code>bootcv</code> and that M is specified.   |
| testIBS        | A range of time points for testing differences between models in the integrated Brier scores.  |
| testTimes      | A vector of time points for testing differences between models in the time-point specific Brier scores.  |
| confInt        | Experimental.  |
| confLevel      | Experimental.  |
| verbose        | if TRUE report details of the progress, e.g. count the steps in cross-validation.  |

|                        |  |
|------------------------|--|
| <code>savePath</code>  | Place in your file system (i.e., a directory on your computer) where training models fitted during cross-validation are saved. If missing training models are not saved. |
| <code>slaveseed</code> | Vector of seeds, as long as <code>B</code> , to be given to the slaves in parallel computing.  |
| <code>na.action</code> | Passed immediately to <code>model.frame</code> . Defaults to <code>na.fail</code> . If set otherwise most prediction models will not work.                               |
| <code>...</code>       | Not used.  |

## Details

Missing data in the response or in the input matrix cause a failure.

The status of the continuous response variable at cutpoints (`times`), ie `status=1` if the response value exceeds the cutpoint and `status=0` otherwise, is compared to predicted event status probabilities which are provided by the prediction models on the basis of covariates. The comparison is done with the Brier score: the quadratic difference between 0-1 response status and predicted probability.

There are two different sources for bias when estimating prediction error in right censored survival problems: censoring and high flexibility of the prediction model. The first is controlled by inverse probability of censoring weighting, the second can be controlled by special Monte Carlo simulation. In each step, the resampling procedures reevaluate the prediction model. Technically this is done by replacing the argument `object$call$data` by the current subset or bootstrap sample of the full data.

For each prediction model there must be a `predictSurvProb` method: for example, to assess a prediction model which evaluates to a `myclass` object one defines a function called `predictSurvProb.myclass` with arguments `object, newdata, cutpoints, train.data, ...`

Such a function takes the object which was fitted with `train.data` and derives a matrix with predicted event status probabilities for each subject in `newdata` (rows) and each cutpoint (column) of the response variable that defines an event status.

Currently, `predictSurvProb` methods are available for the following R-objects:

```
matrix
aalen, cox.aalen from library(timereg)
mfp from library(mfp)
phnnet, survnnet from library(survnnet)
rpart (from library(rpart))
coxph, survfit from library(survival)
cph, psm from library(rms)
prodlm from library(prodlm)
glm from library(stats)
```

## Value

A `pec` object. See also the help pages of the corresponding `print`, `summary`, and `plot` methods. The object includes the following components:

|             |  |
|-------------|--|
| PredErr     | The estimated prediction error according to the <code>splitMethod</code> . A matrix where each column represents the estimated prediction error of a fit at the time points in time.   |
| AppErr      | The training error or apparent error obtained when the model(s) are evaluated in the same data where they were trained. Only if <code>splitMethod</code> is one of "NoInf", "cvK", "BootCv", "Boot632" or "Boot632plus".   |
| BootCvErr   | The prediction error when the model(s) are trained in the bootstrap sample and evaluated in the data that are not in the bootstrap sample. Only if <code>splitMethod</code> is one of "Boot632" or "Boot632plus". When <code>splitMethod="BootCv"</code> then the <code>BootCvErr</code> is stored in the component <code>PredErr</code> . |
| NoInfErr    | The prediction error when the model(s) are evaluated in the permuted data. Only if <code>splitMethod</code> is one of "BootCv", "Boot632", or "Boot632plus". For <code>splitMethod="NoInf"</code> the <code>NoInfErr</code> is stored in the component <code>PredErr</code> .  |
| weight      | The weight used to linear combine the <code>AppErr</code> and the <code>BootCvErr</code> Only if <code>splitMethod</code> is one of "Boot632", or "Boot632plus".   |
| overfit     | Estimated overfit of the model(s). See Efron & Tibshirani (1997, Journal of the American Statistical Association) and Gerds & Schumacher (2007, Biometrics). Only if <code>splitMethod</code> is one of "Boot632", or "Boot632plus".   |
| call        | The call that produced the object  |
| time        | The time points at which the prediction error curves change.   |
| ipcw.fit    | The fitted censoring model that was used for re-weighting the Brier score residuals. See Gerds & Schumacher (2006, Biometrical Journal)  |
| n.risk      | The number of subjects at risk for all time points.  |
| models      | The prediction models fitted in their own data.  |
| cens.model  | The censoring models.  |
| maxtime     | The latest timepoint where the prediction error curves are estimated.  |
| start       | The earliest timepoint where the prediction error curves are estimated.  |
| exact       | TRUE if the prediction error curves are estimated at all unique values of the response in the full data.   |
| splitMethod | The <code>splitMethod</code> used for estimation of the overfitting bias.  |

### Author(s)

Thomas Alexander Gerds <tag@biostat.ku.dk>

### References

- Ulla B. Mogensen, Hemant Ishwaran, Thomas A. Gerds (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of Statistical Software*, 50(11), 1-23. URL <http://www.jstatsoft.org/v50/i11/>.
- E. Graf et al. (1999), Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, vol 18, pp= 2529–2545.
- Efron, Tibshirani (1997) *Journal of the American Statistical Association* 92, 548–560 Improvement On Cross-Validation: The .632+ Bootstrap Method.

Gerds, Schumacher (2006), Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal*, vol 48, 1029–1040.

Thomas A. Gerds, Martin Schumacher (2007) Efron-Type Measures of Prediction Error for Survival Analysis *Biometrics*, 63(4), 1283–1287 doi:10.1111/j.1541-0420.2007.00832.x

Martin Schumacher, Harald Binder, and Thomas Gerds. Assessment of survival prediction models based on microarray data. *Bioinformatics*, 23(14):1768-74, 2007.

Mark A. van de Wiel, Johannes Berkhof, and Wessel N. van Wieringen Testing the prediction error difference between 2 predictors *Biostatistics* (2009) 10(3): 550-560 doi:10.1093/biostatistics/kxp011

## See Also

[plot.pec](#), [summary.pec](#), [R2](#), [crps](#)

## Examples

```
# simulate an artificial data frame
# with survival response and two predictors

set.seed(130971)
library(prodlim)
library(survival)
dat <- SimSurv(100)

# fit some candidate Cox models and compute the Kaplan-Meier estimate

Models <- list("Cox.X1"=coxph(Surv(time,status)~X1,data=dat,y=TRUE),
              "Cox.X2"=coxph(Surv(time,status)~X2,data=dat,y=TRUE),
              "Cox.X1.X2"=coxph(Surv(time,status)~X1+X2,data=dat,y=TRUE))

# compute the apparent prediction error
PredError <- pec(object=Models,
                 formula=Surv(time,status)~X1+X2,
                 data=dat,
                 exact=TRUE,
                 cens.model="marginal",
                 splitMethod="none",
                 B=0,
                 verbose=TRUE)

print(PredError,times=seq(5,30,5))
summary(PredError)
plot(PredError,xlim=c(0,30))

# Comparison of Weibull model and Cox model
library(survival)
library(rms)
library(pec)
data(pbc)
pbc <- pbc[sample(1:NROW(pbc),size=100),]
f1 <- psm(Surv(time,status!=0)~edema+log(bili)+age+sex+albumin,data=pbc)
f2 <- coxph(Surv(time,status!=0)~edema+log(bili)+age+sex+albumin,data=pbc)
```

```

f3 <- cph(Surv(time,status!=0)~edema+log(bili)+age+sex+albumin,data=pec, surv=TRUE)
brier <- pec(list("Weibull"=f1,"CoxPH"=f2,"CPH"=f3),data=pec,formula=Surv(time,status!=0)~1)
print(brier)
plot(brier)

# compute the .632+ estimate of the generalization error
set.seed(130971)
library(prodlm)
library(survival)
dat <- SimSurv(100)
set.seed(17100)
PredError.632plus <- pec(object=Models,
                        formula=Surv(time,status)~X1+X2,
                        data=dat,
                        exact=TRUE,
                        cens.model="marginal",
                        splitMethod="Boot632plus",
                        B=3,
                        verbose=TRUE)

print(PredError.632plus,times=seq(4,12,4))
summary(PredError.632plus)
plot(PredError.632plus,xlim=c(0,30))
# do the same again but now in parallel
## Not run: set.seed(17100)
library(doMC)
registerDoMC()
PredError.632plus <- pec(object=Models,
                        formula=Surv(time,status)~X1+X2,
                        data=dat,
                        exact=TRUE,
                        cens.model="marginal",
                        splitMethod="Boot632plus",
                        B=3,
                        verbose=TRUE)

## End(Not run)
# assessing parametric survival models in learn/validation setting
learndat <- SimSurv(50)
testdat <- SimSurv(30)
library(rms)
f1 <- psm(Surv(time,status)~X1+X2,data=learndat)
f2 <- psm(Surv(time,status)~X1,data=learndat)
pf <- pec(list(f1,f2),formula=Surv(time,status)~1,data=testdat,maxtime=200)
plot(pf)
summary(pf)

# ----- competing risks -----

library(survival)
library(riskRegression)
library(cmprsk)
library(pec)

```

```
cdat <- SimCompRisk(100)
data(cdat)
f1 <- CSC(Hist(time,event)~X1+X2, cause=2, data=cdat)
f1a <- CSC(Hist(time,event)~X1+X2, cause=2, data=cdat, survtype="surv")
f2 <- CSC(Hist(time,event)~X1, data=cdat, cause=2)
require(cmprsk)
## predict.crr <- cmprsk::predict.crr
f3 <- FGR(Hist(time,event)~X1+X2, cause=2, data=cdat)
f4 <- FGR(Hist(time,event)~X1+X2, cause=2, data=cdat)
p1 <- pec(list(f1, f1a, f2, f3, f4), formula=Hist(time,event)~1, data=cdat, cause=2)
```

---

pecCforest

*S3-wrapper function for cforest from the party package*

---

## Description

S3-wrapper function for cforest from the party package

## Usage

```
pecCforest(formula, data, ...)
```

## Arguments

|         |                                  |
|---------|----------------------------------|
| formula | See cforest of the party package |
| data    | See cforest of the party package |
| ...     | See cforest of the party package |

## Details

See cforest of the party package.

## References

Ulla B. Mogensen, Hemant Ishwaran, Thomas A. Gerds (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of Statistical Software*, 50(11), 1-23. URL <http://www.jstatsoft.org/v50/i11/>.



plot.pec

*Plotting prediction error curves***Description**

Plotting prediction error curves for one or more prediction models.

**Usage**

```
## S3 method for class 'pec'
plot(x, what, models, xlim = c(x$start, x$minmaxtime),
     ylim = c(0, 0.3), xlab = "Time", ylab, axes = TRUE, col, lty, lwd, type,
     smooth = FALSE, add.refline = FALSE, add = FALSE, legend = ifelse(add,
     FALSE, TRUE), special = FALSE, ...)
```

**Arguments**

|             |  |
|-------------|--|
| x           | Object of class pec obtained with function <a href="#">pec</a> .   |
| what        | The name of the entry in x. Defaults to PredErr Other choices are AppErr, BootCvErr, Boot632, Boot632plus.   |
| models      | Specifies models in x\$models for which the prediction error curves are drawn. Defaults to all models.   |
| xlim        | Plotting range on the x-axis.  |
| ylim        | Plotting range on the y-axis.  |
| xlab        | Label given to the x-axis.   |
| ylab        | Label given to the y-axis.   |
| axes        | Logical. If FALSE no axes are drawn.   |
| col         | Vector of colors given to the curves of models in the order determined by models.  |
| lty         | Vector of lty's given to the curves of models in the order determined by models.   |
| lwd         | Vector of lwd's given to the curves of models in the order determined by models.   |
| type        | Plotting type: either "l" or "s", see <a href="#">lines</a> .  |
| smooth      | Logical. If TRUE the plotting values are smoothed with the function <a href="#">smooth</a> kind="3R".  |
| add.refline | Logical. If TRUE a dotted horizontal line is drawn as a symbol for the naive rule that predicts probability .5 at all cutpoints (i.e. time points in survival analysis).   |
| add         | Logical. If TRUE only lines are added to an existing device  |
| legend      | if TRUE a legend is plotted by calling the function <a href="#">legend</a> . Optional arguments of the function <a href="#">legend</a> can be given in the form legend.x=val where x is the name of the argument and val the desired value. See also <a href="#">Details</a> .                                       |
| special     | Logical. If TRUE the bootstrap curves of models are plotted together with predErr of models by invoking the function <a href="#">Special</a> . Optional arguments of the function <a href="#">Special</a> can be given in the form special.x=val as with <a href="#">legend</a> . See also <a href="#">Details</a> . |
| ...         | Extra arguments that are passed to <a href="#">plot</a> .  |

## Details

From version 2.0.1 on the arguments `legend.text`, `legend.args`, `lines.type`, `lwd.lines`, `specials` are obsolete and only available for backward compatibility. Instead arguments for the invoked functions `legend`, `axis`, `Special` are simply specified as `legend.lty=2`. The specification is not case sensitive, thus `Legend.lty=2` or `LEGEND.lty=2` will have the same effect. The function `axis` is called twice, and arguments of the form `axis1.labels`, `axis1.at` are used for the time axis whereas `axis2.pos`, `axis1.labels`, etc. are used for the y-axis.

These arguments are processed via `...{}` of `plot.pec` and inside by using the function `resolveSmartArgs`. Documentation of these arguments can be found in the help pages of the corresponding functions.

## Value

The (invisible) object.

## Author(s)

Ulla B. Mogensen <ulmo@biostat.ku.dk>, Thomas A. Gerds <tag@biostat.ku.dk>

## See Also

[pecsummary.pec](#)[Specialprodlim](#)

## Examples

```
# simulate data
# with a survival response and two predictors
library(prodlim)
library(survival)
set.seed(280180)
dat <- SimSurv(100)

# fit some candidate Cox models and
# compute the Kaplan-Meier estimate

Models <- list("Kaplan.Meier"=survfit(Surv(time,status)~1,data=dat),
              "Cox.X1"=coxph(Surv(time,status)~X1,data=dat),
              "Cox.X2"=coxph(Surv(time,status)~X2,data=dat),
              "Cox.X1.X2"=coxph(Surv(time,status)~X1+X2,data=dat))
Models <- list("Cox.X1"=coxph(Surv(time,status)~X1,data=dat),
              "Cox.X2"=coxph(Surv(time,status)~X2,data=dat),
              "Cox.X1.X2"=coxph(Surv(time,status)~X1+X2,data=dat))

# compute the .632+ estimate of the generalization error
set.seed(17100)
PredError.632plus <- pec(object=Models,
                        formula=Surv(time,status)~X1+X2,
                        data=dat,
                        exact=TRUE,
                        cens.model="marginal",
                        splitMethod="boot632plus",
```

```

                                B=5,
                                keep.matrix=TRUE,
                                verbose=TRUE)

# plot the .632+ estimates of the generalization error
plot(PredError.632plus,xlim=c(0,30))

# plot the bootstrapped curves, .632+ estimates of the generalization error
# and Apparent error for the Cox model 'Cox.X1' with the 'Cox.X2' model
# as benchmark
plot(PredError.632plus,
     xlim=c(0,30),
     models="Cox.X1",
     special=TRUE,
     special.bench="Cox.X2",
     special.benchcol=2,
     special.addprederr="AppErr")

```

---

plotPredictEventProb *Plotting predicted survival curves.*

---

## Description

Plotting time-dependent event risk predictions.

## Usage

```
plotPredictEventProb(x, newdata, times, cause = 1, xlim, ylim, xlab, ylab,
  axes = TRUE, col, density, lty, lwd, add = FALSE, legend = TRUE,
  percent = FALSE, ...)
```

## Arguments

|         |   |
|---------|---|
| x       | Object specifying an event risk prediction model.                                     |
| newdata | A data frame with the same variable names as those that were used to fit the model x. |
| times   | Vector of times at which to return the estimated probabilities.                       |
| cause   | Show predicted risk of events of this cause   |
| xlim    | Plotting range on the x-axis.   |
| ylim    | Plotting range on the y-axis.   |
| xlab    | Label given to the x-axis.  |
| ylab    | Label given to the y-axis.  |
| axes    | Logical. If FALSE no axes are drawn.  |
| col     | Vector of colors given to the survival curve.   |
| density | Density of the color – useful for showing many (overlapping) curves.                  |

|         |   |
|---------|---|
| lty     | Vector of lty's given to the survival curve.  |
| lwd     | Vector of lwd's given to the survival curve.  |
| add     | Logical. If TRUE only lines are added to an existing device   |
| legend  | Logical. If TRUE a legend is plotted by calling the function legend. Optional arguments of the function legend can be given in the form legend.x=val where x is the name of the argument and val the desired value. See also Details. |
| percent | Logical. If TRUE the y-axis is labeled in percent.  |
| ...     | Parameters that are filtered by <code>SmartControl</code> and then passed to the functions: <code>plot</code> , <code>axis</code> , <code>legend</code> .   |

### Details

Arguments for the invoked functions `legend` and `axis` are simply specified as `legend.lty=2`. The specification is not case sensitive, thus `Legend.lty=2` or `LEGEND.lty=2` will have the same effect. The function `axis` is called twice, and arguments of the form `axis1.labels`, `axis1.at` are used for the time axis whereas `axis2.pos`, `axis1.labels`, etc. are used for the y-axis.

These arguments are processed via `...{}` of `plotPredictEventProb` and inside by using the function `SmartControl`.

### Value

The (invisible) object.

### Author(s)

Ulla B. Mogensen <ulmo@biostat.ku.dk>, Thomas A. Gerds <>tag@biostat.ku.dk>

### References

Ulla B. Mogensen, Hemant Ishwaran, Thomas A. Gerds (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of Statistical Software*, 50(11), 1-23. URL <http://www.jstatsoft.org/v50/i11/>.

### See Also

[predictEventProbprodlm](#)

### Examples

```
# generate some competing risk data
```

---

plotPredictSurvProb *Plotting predicted survival curves.*

---

### Description

Plotting prediction survival curves for one prediction model using predictSurvProb .

### Usage

```
plotPredictSurvProb(x, newdata, times, xlim, ylim, xlab, ylab, axes = TRUE,
  col, density, lty, lwd, add = FALSE, legend = TRUE, percent = FALSE,
  ...)
```

### Arguments

|         |   |
|---------|---|
| x       | A survival prediction model including call and formula object.  |
| newdata | A data frame with the same variable names as those that were used to fit the model x.   |
| times   | Vector of times at which to return the estimated probabilities.   |
| xlim    | Plotting range on the x-axis.   |
| ylim    | Plotting range on the y-axis.   |
| xlab    | Label given to the x-axis.  |
| ylab    | Label given to the y-axis.  |
| axes    | Logical. If FALSE no axes are drawn.  |
| col     | Vector of colors given to the survival curve.   |
| density | Density of the color – useful for showing many (overlapping) curves.  |
| lty     | Vector of lty's given to the survival curve.  |
| lwd     | Vector of lwd's given to the survival curve.  |
| add     | Logical. If TRUE only lines are added to an existing device   |
| legend  | Logical. If TRUE a legend is plotted by calling the function legend. Optional arguments of the function legend can be given in the form legend.x=val where x is the name of the argument and val the desired value. See also Details. |
| percent | Logical. If TRUE the y-axis is labeled in percent.  |
| ...     | Parameters that are filtered by <a href="#">SmartControl</a> and then passed to the functions: <a href="#">plot</a> , <a href="#">axis</a> , <a href="#">legend</a> .   |

### Details

Arguments for the invoked functions legend and axis are simply specified as legend.lty=2. The specification is not case sensitive, thus Legend.lty=2 or LEGEND.lty=2 will have the same effect. The function axis is called twice, and arguments of the form axis1.labels, axis1.at are used for the time axis whereas axis2.pos, axis1.labels, etc. are used for the y-axis.

These arguments are processed via ...{} of plotPredictSurvProb and inside by using the function SmartControl.

**Value**

The (invisible) object.

**Author(s)**

Ulla B. Mogensen <ulmo@biostat.ku.dk>, Thomas A. Gerds <tag@biostat.ku.dk>

**References**

Ulla B. Mogensen, Hemant Ishwaran, Thomas A. Gerds (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of Statistical Software*, 50(11), 1-23. URL <http://www.jstatsoft.org/v50/i11/>.

**See Also**

[predictSurvProbprodlm](#)

**Examples**

```
# generate some survival data
library(prodlm)
d <- SimSurv(100)
# then fit a Cox model
library(rms)
coxmodel <- cph(Surv(time,status)~X1+X2,data=d,surv=TRUE)
# plot predicted survival probabilities for all time points
ttt <- sort(unique(d$time))
# and for selected predictor values:
ndat <- data.frame(X1=c(0.25,0.25,-0.05,0.05),X2=c(0,1,0,1))
plotPredictSurvProb(coxmodel,newdata=ndat,times=ttt)

# the same can be done e.g. for a randomSurvivalForest model
library(randomForestSRC)
rsfmodel <- rfsrc(Surv(time,status)~X1+X2,data=d)
plotPredictSurvProb(rsfmodel,newdata=ndat,times=ttt)
```

---

|                  |   |
|------------------|---|
| predictEventProb | <i>Predicting event probabilities (cumulative incidences) in competing risk models.</i> |
|------------------|---|

---

**Description**

Function to extract event probability predictions from various modeling approaches. The most prominent one is the combination of cause-specific Cox regression models which can be fitted with the function `cumincCox` from the package `compRisk`.

**Usage**

```
predictEventProb(object, newdata, times, cause, ...)
```

**Arguments**

|         |  |
|---------|--|
| object  | A fitted model from which to extract predicted event probabilities   |
| newdata | A data frame containing predictor variable combinations for which to compute predicted event probabilities.                    |
| times   | A vector of times in the range of the response variable, for which the cumulative incidences event probabilities are computed. |
| cause   | Identifies the cause of interest among the competing events.   |
| ...     | Additional arguments that are passed on to the current method.   |

**Details**

The function `predictEventProb` is a generic function that means it invokes specifically designed functions depending on the 'class' of the first argument.

See [predictSurvProb](#).

**Value**

A matrix with as many rows as `NROW(newdata)` and as many columns as `length(times)`. Each entry should be a probability and in rows the values should be increasing.

**Author(s)**

Thomas A. Gerds <tag@biostat.ku.dk>

**See Also**

See [predictSurvProb](#).

**Examples**

```
library(pec)
library(CoxBoost)
library(survival)
library(riskRegression)
library(prodlim)
train <- SimCompRisk(100)
test <- SimCompRisk(10)
cox.fit <- CSC(Hist(time, cause)~X1+X2, data=train)
predictEventProb(cox.fit, newdata=test, times=seq(1:10), cause=1)
## cb.fit <- coxboost(Hist(time, cause)~X1+X2, cause=1, data=train, stepno=10)
## predictEventProb(cb.fit, newdata=test, times=seq(1:10), cause=1)
```

---

predictLifeYearsLost *Predicting life years lost (cumulative cumulative incidences) in competing risk models.*

---

### Description

Function to extract predicted life years lost from various modeling approaches. The most prominent one is the combination of cause-specific Cox regression models which can be fitted with the function `cumincCox` from the package `compRisk`.

### Usage

```
predictLifeYearsLost(object, newdata, times, cause, ...)
```

### Arguments

|                      |  |
|----------------------|--|
| <code>object</code>  | A fitted model from which to extract predicted event probabilities   |
| <code>newdata</code> | A data frame containing predictor variable combinations for which to compute predicted event probabilities.                    |
| <code>times</code>   | A vector of times in the range of the response variable, for which the cumulative incidences event probabilities are computed. |
| <code>cause</code>   | Identifies the cause of interest among the competing events.   |
| <code>...</code>     | Additional arguments that are passed on to the current method.   |

### Details

The function `predictLifeYearsLost` is a generic function that means it invokes specifically designed functions depending on the 'class' of the first argument.

See [predictSurvProb](#).

### Value

A matrix with as many rows as `NROW(newdata)` and as many columns as `length(times)`. Each entry should be a positive value and in rows the values should be increasing.

### Author(s)

Thomas A. Gerds <[tag@biostat.ku.dk](mailto:tag@biostat.ku.dk)>

### See Also

[predictSurvProb](#), [predictEventProb](#).



**Examples**

```

library(pec)
library(riskRegression)
library(survival)
library(prodlim)
train <- SimCompRisk(100)
test <- SimCompRisk(10)
fit <- CSC(Hist(time,cause)~X1+X2,data=train,cause=1)
predictLifeYearsLost(fit,newdata=test,times=seq(1:10),cv=FALSE,cause=1)

```

---

predictSurvProb      *Predicting survival probabilities*

---

**Description**

Function to extract survival probability predictions from various modeling approaches. The most prominent one is the Cox regression model which can be fitted for example with ‘coxph’ and with ‘cph’.

**Usage**

```

## S3 method for class 'aalen'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'riskRegression'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'cox.aalen'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'cph'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'coxph'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'mfp'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'matrix'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'selectCox'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'pecCforest'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'prodlim'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'psm'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'survfit'
predictSurvProb(object,newdata,times,...)
## S3 method for class 'survnnet'
predictSurvProb(object,newdata,times,train.data,...)

```

```
## S3 method for class 'rpart'
predictSurvProb(object,newdata,times,train.data,...)
## S3 method for class 'phnnet'
predictSurvProb(object,newdata,times,train.data,...)
```

### Arguments

|                         |  |
|-------------------------|--|
| <code>object</code>     | A fitted model from which to extract predicted survival probabilities  |
| <code>newdata</code>    | A data frame containing predictor variable combinations for which to compute predicted survival probabilities.   |
| <code>times</code>      | A vector of times in the range of the response variable, e.g. times when the response is a survival object, at which to return the survival probabilities. |
| <code>train.data</code> | An optional data frame which contains the response and predictor variable combinations in which the prediction model was trained                           |
| <code>...</code>        | Additional arguments that are passed on to the current method.   |

### Details

The function `predictSurvProb` is a generic function that means it invokes specifically designed functions depending on the 'class' of the first argument.

The function `pec` requires survival probabilities for each row in `newdata` at requested times. These probabilities are extracted from a fitted model of class `CLASS` with the function `predictSurvProb.CLASS`.

Currently there are `predictSurvProb` methods for objects of class `cph` (library `rms`), `coxph` (library `survival`), `aalen` (library `timereg`), `cox.aalen` (library `timereg`), `mfp` (library `mfp`), `phnnet` (library `survnnet`), `survnnet` (library `survnnet`), `rpart` (library `rpart`), `product.limit` (library `prodlim`), `survfit` (library `survival`), `psm` (library `rms`)

### Value

A matrix with as many rows as `NROW(newdata)` and as many columns as `length(times)`. Each entry should be a probability and in rows the values should be decreasing.

### Note

In order to assess the predictive performance of a new survival model a specific `predictSurvProb` S3 method has to be written. For examples, see the bodies of the existing methods.

The performance of the assessment procedure, in particular for resampling where the model is repeatedly evaluated, will be improved by suppressing in the call to the model all the computations that are not needed for probability prediction. For example, `se.fit=FALSE` can be set in the call to `cph`.

### Author(s)

Thomas A. Gerds <tag@biostat.ku.dk>

## References

Ulla B. Mogensen, Hemant Ishwaran, Thomas A. Gerds (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of Statistical Software*, 50(11), 1-23. URL <http://www.jstatsoft.org/v50/i11/>.

## See Also

[predict,survfit](#)

## Examples

```
# generate some survival data
library(proplim)
d <- SimSurv(100)
# then fit a Cox model
library(rms)
coxmodel <- cph(Surv(time,status)~X1+X2,data=d,surv=TRUE)

# predicted survival probabilities can be extracted
# at selected time-points:
ttt <- quantile(d$time)
# for selected predictor values:
ndat <- data.frame(X1=c(0.25,0.25,-0.05,0.05),X2=c(0,1,0,1))
# as follows
predictSurvProb(coxmodel,newdata=ndat,times=ttt)

## simulate some learning and some validation data
learndat <- SimSurv(100)
valdat <- SimSurv(100)
## use the learning data to fit a Cox model
library(survival)
fitCox <- coxph(Surv(time,status)~X1+X2,data=learndat)
## suppose we want to predict the survival probabilities for all patients
## in the validation data at the following time points:
## 0, 12, 24, 36, 48, 60
psurv <- predictSurvProb(fitCox,newdata=valdat,times=seq(0,60,12))
## This is a matrix with survival probabilities
## one column for each of the 5 time points
## one row for each validation set individual

# the same can be done e.g. for a randomSurvivalForest model
library(randomForestSRC)
rsfmodel <- rfsrc(Surv(time,status)~X1+X2,data=d)
predictSurvProb(rsfmodel,newdata=ndat,times=ttt)
```

**Description**

Print the important arguments of call and the prediction error values at selected time points.

**Usage**

```
## S3 method for class 'pec'
print(x, times, digits = 3, what = NULL, ...)
```

**Arguments**

|        |   |
|--------|---|
| x      | Object of class pec   |
| times  | Time points at which to show the values of the prediction error curve(s)  |
| what   | What estimate of the prediction error curve to show. Should be a string matching an element of x. The default is determined by splitMethod. |
| digits | Number of decimals used in tables.  |
| print  | Set to FALSE to suppress printing.  |
| ...    | Not used  |

**Value**

The first argument in the invisible cloak.

**Author(s)**

Thomas A. Gerds <tag@biostat.ku.dk>

**See Also**

[pec](#)

---

R2

*Explained variation for survival models*

---

**Description**

This function computes a time-dependent  $R^2$  like measure of the variation explained by a survival prediction model, by dividing the mean squared error (Brier score) of the model by the mean squared error (Brier score) of a reference model which ignores all the covariates.

**Usage**

```
R2(object, models, what, times, reference = 1)
```

**Arguments**

|           |   |
|-----------|---|
| object    | An object with estimated prediction error curves obtained with the function <a href="#">pec</a>   |
| models    | For which of the models in <code>object\$models</code> should we compute $R^2(t)$ . By default all models are used except for the reference model.  |
| what      | The name of the entry in <code>x</code> to be used. Defaults to <code>PredErr</code> . Other choices are <code>AppErr</code> , <code>BootCvErr</code> , <code>Boot632</code> , <code>Boot632plus</code> . |
| times     | Time points at which the summaries are shown.   |
| reference | Position of the model whose prediction error is used as the reference in the denominator when constructing $R^2$  |

**Details**

In survival analysis the prediction error of the Kaplan-Meier estimator plays a similar role as the total sum of squares in linear regression. Hence, it is a sensible reference model for  $R^2$ .

**Value**

A matrix where the first column holds the times and the following columns are the corresponding  $R^2$  values for the requested prediction models.

**Author(s)**

Thomas A. Gerds <tag@biostat.ku.dk>

**References**

E. Graf et al. (1999), Assessment and comparison of prognostic classification schemes for survival data. *Statistics in Medicine*, vol 18, pp= 2529–2545.

Gerds TA, Cai T & Schumacher M (2008) The performance of risk prediction models *Biometrical Journal*, 50(4), 457–479

**See Also**

[pec](#)

**Examples**

```
set.seed(18713)
library(proplim)
library(survival)
dat=SimSurv(100)
nullmodel=proplim(Hist(time,status)~1,data=dat)
pmodel1=coxph(Surv(time,status)~X1+X2,data=dat)
pmodel2=coxph(Surv(time,status)~X2,data=dat)
perror=pec(list(Cox1=pmodel1,Cox2=pmodel2),Hist(time,status)~1,data=dat,reference=TRUE)
R2(perror,times=seq(0,1,.1),reference=1)
```

---

resolvesplitMethod      *Resolve the splitMethod for estimation of prediction performance*

---

### Description

The function computes a matrix of random indices obtained by drawing from the row numbers of a data set either with or without replacement. The matrix can be used to repeatedly set up independent training and validation sets.

### Usage

```
resolvesplitMethod(splitMethod, B, N, M)
```

### Arguments

|             |  |
|-------------|--|
| splitMethod | String that determines the splitMethod to use. Available splitMethods are none/noPlan (no splitting), bootcv or outofbag (bootstrap cross-validation), cvK (K-fold cross-validation, e.g. cv10 gives 10-fold), boot632, boot632plus or boot632+, loocv (leave-one-out) |
| B           | The number of repetitions.   |
| N           | The sample size  |
| M           | For subsampling bootstrap the size of the subsample. Note $M < N$ .  |

### Value

A list with the following components

|               |   |
|---------------|---|
| name          | the official name of the splitMethod  |
| internal.name | the internal name of the splitMethod  |
| index         | a matrix of indices with B columns and either N or M rows, dependent on splitMethod |
| B             | the value of the argument B   |
| N             | the value of the argument N   |
| M             | the value of the argument M   |

### Author(s)

Thomas Alexander Gerds <tag@biostat.ku.dk>

**Examples**

```
# BootstrapCrossValidation: Sampling with replacement
  resolvesplitMethod("BootCv",N=10,B=10)

# 10-fold cross-validation: repeated 2 times
  resolvesplitMethod("cv10",N=10,B=2)

# leave-one-out cross-validation
  resolvesplitMethod("loocv",N=10)

  resolvesplitMethod("bootcv632plus",N=10,B=2)
```

---

 selectCox

*Backward variable selection in the Cox regression model*


---

**Description**

This is a wrapper function which first selects variables in the Cox regression model using `fastbw` from the `rms` package and then returns a fitted Cox regression model with the selected variables.

**Usage**

```
selectCox(formula, data, rule = "aic")
```

**Arguments**

|         |  |
|---------|--|
| formula | A formula object with a <code>Surv</code> object on the left-hand side and all the variables on the right-hand side. |
| data    | Name of an data frame containing all needed variables.   |
| rule    | The method for selecting variables. See <a href="#">fastbw</a> for details.  |

**Details**

This function first calls `cph` then `fastbw` and finally `cph` again.

**References**

Ulla B. Mogensen, Hemant Ishwaran, Thomas A. Gerds (2012). Evaluating Random Forests for Survival Analysis Using Prediction Error Curves. *Journal of Statistical Software*, 50(11), 1-23. URL <http://www.jstatsoft.org/v50/i11/>.

**Examples**

```
data(GBSG2)
library(survival)
f <- selectCox(Surv(time,cens)~horTh+age+menostat+tsize+tgrade+pnodes+progrec+estrec ,
  data=GBSG2)
```

---

|           |   |
|-----------|---|
| selectFGR | <i>Stepwise variable selection in the Fine &amp; Gray regression competing risk model</i> |
|-----------|---|

---

### Description

This is a wrapper function which first selects variables in the Fine & Gray regression model using `crrstep` from the `crrstep` package and then returns a fitted Fine & Gray regression model with the selected variables.

### Usage

```
selectFGR(formula, data, cause = 1, rule = "AIC", direction = "backward",
  ...)
```

### Arguments

|           |   |
|-----------|---|
| formula   | A formula whose left hand side is a <code>Hist</code> object – see <a href="#">Hist</a> . The right hand side specifies (a linear combination of) the covariates. See examples below. |
| data      | A <code>data.frame</code> in which all the variables of <code>formula</code> can be interpreted.  |
| cause     | The failure type of interest. Defaults to 1.  |
| rule      | Rule to pass on to <code>crrstep</code> ("AIC", "BIC" or "BICcr"), also see <code>crrstep</code>  |
| direction | see <code>crrstep</code>  |
| ...       | Further arguments passed to <code>crrstep</code> .  |

### Author(s)

Rob C.M. van Kruijsdijk <R.C.M.vanKruijsdijk@umcutrecht.nl>

Thomas Alexander Gerds <tag@biostat.ku.dk>

### Examples

```
## Not run:
library(riskRegression)
library(prodlm)
library(lava)
library(cmprsk)

m <- crModel()
m <- addvar(m,c('X1','X2','X3','X4','X5','X6','X7','X8','X9','X10'))
distribution(m,c("X2","X7","X9")) <- binomial.lvm()
regression(m,eventtime1~X1+X2+X5+X9) <- c(-1,1,0.5,0.8)
set.seed(100)
d <- sim(m,100)
## full formula
ff <- Hist(time, event) ~ X1 + X2 + X3 + X4 +X5 + X6 + X7+ X8 + X9 + X10
```



```

# Fit full model with FGR
fg <- FGR(ff,cause=1,data=d)

# Backward selection based on the AIC
sfgAIC <- selectFGR(ff, data=d, rule="AIC", direction="backward")

sfgAIC$fit # Final FGR-model with selected variables

# Risk reclassification plot at time = 4
plot(predictEventProb(fg,times=4,newdata=d),
      predictEventProb(sfgAIC,times=4,newdata=d))

# Backward selection based on the BIC, while forcing
# the last two variables (X9 and X10) in the model
sfgBIC <- selectFGR(ff, data=d, rule="BIC", direction="backward",
                    scope.min=~X9+X10)

## apparent performance
pec(list(full.model=fg,selectedAIC=sfgAIC,selectedBIC=sfgBIC),
     formula=Hist(time, event)~1,
     data=d)

## bootstrap cross-validation performance
set.seed(7)
pec(list(full.model=fg,selectedAIC=sfgAIC,selectedBIC=sfgBIC),
     formula=Hist(time, event)~1,
     data=d,
     B=5,
     splitMethod="bootcv")

## End(Not run)

```

---

simCost

*Simulate COST alike data*


---

## Description

Simulate data alike the data from the Copenhagen stroke study (COST)

## Usage

```
simCost(N)
```

## Arguments

N                      Sample size

## Details

This uses functionality of the lava package.

**Value**

Data frame

**Author(s)**

Thomas Alexander Gerds

---

|         |  |
|---------|--|
| Special | <i>Drawing bootstrapped cross-validation curves and the .632 or .632plus error of models. The prediction error for an optional benchmark model can be added together with bootstrapped cross-validation error and apparent errors.</i> |
|---------|--|

---

**Description**

This function is invoked and controlled by `plot.pec`.

**Usage**

```
Special(x, y, addprederr, models, bench, benchcol, times, maxboot, bootcol, col,
        lty, lwd)
```

**Arguments**

|                         |   |
|-------------------------|---|
| <code>x</code>          | an object of class 'pec' as returned by the <code>pec</code> function.  |
| <code>y</code>          | Prediction error values.  |
| <code>addprederr</code> | Additional prediction errors. The options are bootstrap cross-validation errors or apparent errors.                                 |
| <code>models</code>     | One model also specified in <code>pec</code> for which the <code>predErr</code> in <code>plot.pec</code> is to be drawn.            |
| <code>bench</code>      | A benchmark model (also specified in <code>pec</code> ) for which the <code>predErr</code> in <code>plot.pec</code> is to be drawn. |
| <code>benchcol</code>   | Color of the benchmark curve.   |
| <code>times</code>      | Time points at which the curves must be plotted.  |
| <code>maxboot</code>    | Maximum number of bootstrap curves to be added. Default is all.   |
| <code>bootcol</code>    | Color of the bootstrapped curves. Default is 'gray77'.  |
| <code>col</code>        | Color of the different error curves for models.   |
| <code>lty</code>        | Line type of the different error curves for models.   |
| <code>lwd</code>        | Line width of the different error curves for models.  |

**Details**

This function should not be called directly. The arguments can be specified as `Special.arg` in the call to `plot.pec`.

**Value**

Invisible object.

**See Also**

[plot.pec](#)

# Index

## \*Topic **datasets**

cost, 9  
GBSG2, 12  
Pbc3, 15

## \*Topic **prediction**

resolvesplitMethod, 38

## \*Topic **survival**

calPlot, 2  
cindex, 5  
coxboost, 10  
crps, 11  
ipcw, 13  
pec, 17  
pecCforest, 24  
plot.pec, 25  
plotPredictEventProb, 27  
plotPredictSurvProb, 29  
predictEventProb, 30  
predictLifeYearsLost, 32  
predictSurvProb, 33  
print.pec, 35  
R2, 36  
selectCox, 39  
selectFGR, 40

axis, 28, 29

calPlot, 2  
cindex, 5  
cost, 9  
coxboost, 10  
crps, 11, 22

dpik, 4

fastbw, 39

GBSG2, 12

Hist, 10, 40

ibs (crps), 11

ipcw, 13

legend, 28, 29

lines, 3

mclapply, 4

model.frame, 4

Pbc3, 15

pec, 11, 12, 14, 17, 25, 26, 36, 37

pecCforest, 24

plot, 25, 28, 29

plot.pec, 22, 25, 43

plotPredictEventProb, 27

plotPredictSurvProb, 29

points, 3

predict, 35

predictEventProb, 28, 30, 32

predictLifeYearsLost, 32

predictSurvProb, 2, 5, 17, 30–32, 33

print.pec, 35

prodlim, 26, 28, 30

R2, 22, 36

resolvesplitMethod, 38

selectCox, 39

selectFGR, 40

simCost, 41

SmartControl, 4, 28, 29

smooth, 25

Special, 26, 42

summary.pec, 22, 26

summary.pec (print.pec), 35

survfit, 35