

# Package ‘PRIMsrc’

November 10, 2015

**Type** Package

**Title** PRIM Survival Regression Classification

**Version** 0.6.3

**Date** 2015-11-10

**Author**

Jean-Eudes Dazard [aut, cre], Michael Choe [ctb], Michael LeBlanc [ctb], Alberto Santana [ctb]

**Maintainer** Jean-Eudes Dazard <jxd101@case.edu>

**Description** Performs a unified treatment of Bump Hunting by Patient Rule Induction Method (PRIM) in Survival, Regression and Classification settings (SRC). The current version is a development release that only implements the case of a survival response. New features will be added soon as they are available.

**Depends** R (>= 3.0.2), parallel, survival, Hmisc, glmnet, MASS

**Imports** graphics, grDevices, stats

**URL** <https://github.com/jedazard/PRIMsrc>

**Repository** PRIMsrc, GitHub, Inc.

**License** GPL (>= 3) | file LICENSE

**LazyLoad** yes

**LazyData** yes

**Archs** i386, x64

## R topics documented:

|                           |    |
|---------------------------|----|
| PRIMsrc-package . . . . . | 2  |
| plot.PRSP . . . . .       | 5  |
| plot_boxkm . . . . .      | 8  |
| plot_boxtrace . . . . .   | 10 |
| plot_boxtraj . . . . .    | 13 |
| plot_profile . . . . .    | 16 |
| predict.PRSP . . . . .    | 19 |
| PRIMsrc.news . . . . .    | 21 |
| print.PRSP . . . . .      | 22 |
| Real.1 . . . . .          | 23 |
| Real.2 . . . . .          | 24 |
| sbh . . . . .             | 26 |
| summary.PRSP . . . . .    | 32 |

|                        |    |
|------------------------|----|
| Synthetic.1 . . . . .  | 34 |
| Synthetic.1b . . . . . | 35 |
| Synthetic.2 . . . . .  | 36 |
| Synthetic.3 . . . . .  | 37 |
| Synthetic.4 . . . . .  | 38 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>40</b> |
|--------------|-----------|

---

|                 |  |
|-----------------|--|
| PRIMsrc-package | <i>Bump Hunting by Patient Rule Induction Method in Survival, Regression and Classification settings</i> |
|-----------------|--|

---

## Description

Performs a unified treatment of Bump Hunting by Patient Rule Induction Method (PRIM) in Survival, Regression and Classification settings (SRC). The method generates decision rules delineating a region in the predictor space, where the response is larger than its average over the entire space. The region is shaped as a hyperdimensional box or hyperrectangle that is not necessarily contiguous. Assumptions are that the multivariate input variables can be discrete or continuous and the univariate response variable can be discrete (Classification), continuous (Regression) or a time-to event, possibly censored (Survival). It is intended to handle low and high-dimensional multivariate datasets, including the situation where the number of covariates exceeds or dominates that of samples ( $p > n$  or  $p \gg n$  paradigm).

## Details

The current version is a development release that only implements the case of a survival response. At this point, survival bump hunting is also restricted to a directed peeling search of the first box covered by the recursive coverage (outer) loop of our Patient Recursive Survival Peeling (PRSP) algorithm. New features will be added soon.

The main function relies on an optional variable pre-selection procedure that is run before the PRSP algorithm. At this point, this is done by a cross-validated penalization of the partial likelihood using the R package **glmnet**.

The following describes the end-user functions that are needed to run a complete procedure. The other internal subroutines are not documented in the manual and are not to be called by the end-user at any time. For computational efficiency, some end-user functions offer a parallelization option that is done by passing a few parameters needed to configure a cluster. This is indicated by an asterisk (\* = optionally involving cluster usage). The R features are categorized as follows:

1. END-USER FUNCTION FOR PACKAGE NEWS  
[PRIMsrc.news](#) **Display the PRIMsrc Package News**  
 Function to display the log file NEWS of updates of the **PRIMsrc** package.
2. END-USER S3-GENERIC FUNCTIONS FOR SUMMARY, DISPLAY, PLOT AND PREDICTION  
[summary](#) **Summary Function**  
 S3-generic summary function to summarize the main parameters used to generate the PRSP object.  
  
[print](#) **Print Function**  
 S3-generic print function to display the cross-validated estimated values of the PRSP object.

**plot 2D Visualization of Data Scatter and Box Vertices**

S3-generic plotting function for two-dimensional visualization of original or predicted data scatter as well as cross-validated box vertices of a PRSP object. The scatter plot is for a given peeling step of the peeling sequence and a given plane, both specified by the user.

**predict Predict Function**

S3-generic predict function to predict the box membership and box vertices on an independent set.

**3. END-USER SURVIVAL BUMP HUNTING FUNCTION****sbh (\*) Cross-Validated Survival Bump Hunting**

Main end-user function for fitting a cross-validated Survival Bump Hunting (SBH) model. Returns a cross-validated PRSP object, as generated by our Patient Recursive Survival Peeling or PRSP algorithm, containing cross-validated estimates of end-points statistics of interest. The function relies on an internal variable pre-selection procedure before the PRSP algorithm is run. At this point, this is done by regular Cox-regression (from the R package **survival**) or a cross-validated Elasticnet Regularized Cox-Regression (from the R package **glmnet**), depending on whether the number of covariates is less ( $p \leq n$ ) or greater ( $p > n$ ) than the number of samples, respectively. At this point, the main function sbh performs the search of the *first* box of the recursive coverage (outer) loop of our Patient Recursive Survival Peeling (PRSP) algorithm. The PRSP object contains cross-validated estimates of all the decision-rules of pre-selected covariates and all other statistical quantities of interest at each iteration of the peeling sequence (inner loop of the PRSP algorithm). It enables the display of results graphically off/for model tuning/selection, all peeling trajectories, covariate traces, and survival distributions (see plotting functions below for more details). The function offers a few options such as the type of  $K$ -fold cross-validation desired ((replicated)-averaged or-combined), the peeling criterion for peeling the next box, the optimization criterion for model tuning and selection and a few more parameters for the PRSP algorithm. The function takes advantage of the R package **parallel** for efficient parallel execution. It allows users to create a cluster of workstations on a local and/or remote machine(s), enabling scaling-up to the number of specified CPU cores.

**4. END-USER PLOTTING FUNCTIONS FOR MODEL VALIDATION AND VISUALIZATION OF RESULTS****plot\_profile Visualization for Model Selection/Validation**

Function for plotting the cross-validated tuning profiles of a PRSP object. It uses the user's choice of statistics among the Log Hazard Ratio (LHR), Log-Rank Test (LRT) or Concordance Error Rate (CER) as a function of the model tuning parameter, that is, the optimal number of peeling steps of the peeling sequence (inner loop of our PRSP algorithm).

**plot\_boxtraj Visualization of Peeling Trajectories/Profiles**

Function for plotting the cross-validated peeling trajectories/profiles of a PRSP object. Applies to the user-specified covariates among the pre-selected ones and all other statistical quantities of interest at each iteration of the peeling sequence (inner loop of our PRSP algorithm).

**plot\_boxtrace Visualization of Covariates Traces**

Function for plotting the cross-validated covariates traces of a PRSP object. Plot the cross-validated modal trace curves of covariate importance and covariate usage of the user-specified covariates among the pre-selected ones at each iteration of the peeling sequence (inner loop of our PRSP algorithm).

**plot\_boxkm Visualization of Survival Distributions**

Function for plotting the cross-validated survival distributions of a PRSP object. Plot the cross-validated Kaplan-Meier estimates of survival distributions for the highest risk (inbox) versus lower-risk (outbox) groups of samples at each iteration of the peeling sequence (inner loop of

our PRSP algorithm).

## 5. END-USER DATASETS

### [Synthetic.1](#), [Synthetic.1b](#), [Synthetic.2](#), [Synthetic.3](#), [Synthetic.4](#) **Five Datasets From Simulated Regression Survival Models**

Five datasets from simulated regression survival models #1-4 as described in Dazard et al. (2015) representing low- and high-dimensional situations, and where regression parameters represent various types of relationship between survival times and covariates including saturated and noisy situations. In three datasets where non-informative noisy covariates were used, these covariates were not part of the design matrix (models #2-3 and #4). In one dataset, the signal is limited to a box-shaped region  $R$  of the predictor space (model #1b). In the last dataset, the signal is limited to 10% of the predictors in a  $p > n$  situation (model #4). Survival time was generated from an exponential model with rate parameter  $\lambda$  (and mean  $\frac{1}{\lambda}$ ) according to a Cox-PH model with hazard  $\exp(\eta)$ , where  $\eta(\cdot)$  is the regression function. Censoring indicator were generated from a uniform distribution on  $[0,3]$  (models #1-3) or  $[0,2]$  (model #4). In these synthetic datasets, all covariates are continuous, i.i.d. from a multivariate uniform distribution on  $[0,1]$  (models #1-3) or from a multivariate standard normal distribution (model #4).

### [Real.1](#) **Clinical Dataset**

Publicly available HIV clinical data from the Women's Interagency HIV cohort Study (WIHS). Inclusion criteria of the study were that women at enrolment were (i) alive, (ii) HIV-1 infected, and (iii) free of clinical AIDS symptoms. Women were followed until the first of the following occurred: (i) treatment initiation (HAART), (ii) AIDS diagnosis, (iii) death, or administrative censoring. The studied outcomes were the competing risks "AIDS/Death (before HAART)" and "Treatment Initiation (HAART)". However, here, for simplification purposes, only the first of the two competing events (i.e. the time to AIDS/Death), was used in this dataset example. Likewise, the entire study enrolled 1164 women, but only the complete cases were used in this dataset example for simplification. Variables included history of Injection Drug Use ("IDU") at enrollment, African American ethnicity ("Race"), age ("Age"), and baseline CD4 count ("CD4"). The question in this dataset example was whether it is possible to achieve a prognostication of patients for AIDS and HAART.

### [Real.2](#) **Genomic Dataset**

Publicly available lung cancer genomic data from the Chemores Cohort Study. This was an integrated study of mRNA, miRNA and clinical variables to characterize the molecular distinctions between squamous cell carcinoma (SCC) and adenocarcinoma (AC) in Non Small Cell Lung Cancer (NSCLC). Tissue samples were analysed from a cohort of 123 patients who underwent complete surgical resection at the Institut Mutualiste Montsouris (Paris, France) between 30 January 2002 and 26 June 2006. In this genomic dataset, only the expression levels of Agilent miRNA probes ( $p = 939$ ) were included from the  $n = 123$  samples of the Chemores cohort. It represents a situation where the number of covariates dominates the number of complete observations, or  $p \gg n$  case.

Known Bugs/Problems : None at this time.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>

- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

## See Also

- makeCluster (R package **parallel**)
- plot.survfit (R package **survival**)
- glmnet (R package **glmnet**)

---

plot.PRSP

*2D Visualization of Data Scatter and Box Vertices*

---

## Description

S3-generic plotting function for two-dimensional visualization of original or predicted data scatter as well as cross-validated box vertices of a PRSP object. The scatter plot is for a given peeling step of the peeling sequence and a given plane, both specified by the user.

## Usage

```
## S3 method for class 'PRSP'
plot(x,
      main = NULL,
      proj = c(1,2),
      splom = TRUE,
      boxes = FALSE,
      steps = x$cvfit$cv.nsteps,
      pch = 16,
      cex = 0.5,
      col = 2:(length(steps)+1),
      col.box = 2:(length(steps)+1),
      lty.box = rep(2,length(steps)),
      lwd.box = rep(1,length(steps)),
      add.legend = TRUE,
      device = NULL,
      file = "Scatter Plot",
```

```

path=getwd(),
horizontal = FALSE,
width = 5,
height = 5, ...)

```

## Arguments

|                         |  |
|-------------------------|--|
| <code>x</code>          | Object of class PRSP as generated by the main function <a href="#">sbh</a> .   |
| <code>main</code>       | Character vector. Main Title. Defaults to NULL.  |
| <code>proj</code>       | Integer vector of length two, specifying the two dimensions of the projection plane. Defaults to <code>c(1,2)</code> .   |
| <code>splom</code>      | Logical scalar. Shall the scatter plot of points inside the box(es) be plotted? Default to TRUE.   |
| <code>boxes</code>      | Logical scalar. Shall the box vertices be plotted or just the scatter of points? Default to FALSE.   |
| <code>steps</code>      | Integer vector. Vector of peeling steps at which to plot the in-box samples and box vertices. Defaults to the last peeling step of PRSP object <code>object</code> .                               |
| <code>pch</code>        | Integer scalar of symbol number for the scatter plot. Defaults to 16.  |
| <code>cex</code>        | Integer scalar of symbol expansion. Defaults to 0.5.   |
| <code>col</code>        | Integer vector specifying the symbol color for each step. Defaults to vector of colors of length the number of steps. The vector is reused cyclically if it is shorter than the number of steps.   |
| <code>col.box</code>    | Integer vector of line color of box vertices for each step. Defaults to vector of colors of length the number of steps. The vector is reused cyclically if it is shorter than the number of steps. |
| <code>lty.box</code>    | Integer vector of line type of box vertices for each step. Defaults to vector of 2's of length the number of steps. The vector is reused cyclically if it is shorter than the number of steps.     |
| <code>lwd.box</code>    | Integer vector of line width of box vertices for each step. Defaults to vector of 1's of length the number of steps. The vector is reused cyclically if it is shorter than the number of steps.    |
| <code>add.legend</code> | Logical scalar. Shall the legend of steps numbers be plotted? Defaults to TRUE.  |
| <code>device</code>     | Graphic display device in {NULL, "PS", "PDF"}. Defaults to NULL (standard output screen). Currently implemented graphic display devices are "PS" (Postscript) or "PDF" (Portable Document Format). |
| <code>file</code>       | File name for output graphic. Defaults to "Scatter Plot".  |
| <code>path</code>       | Absolute path (without final (back)slash separator). Defaults to working directory path.   |
| <code>horizontal</code> | Logical scalar. Orientation of the printed image. Defaults to FALSE, that is potrait orientation.  |
| <code>width</code>      | Numeric scalar. Width of the graphics region in inches. Defaults to 5.   |
| <code>height</code>     | Numeric scalar. Height of the graphics region in inches. Defaults to 5.  |
| <code>...</code>        | Generic arguments passed to other plotting functions.  |

## Details

The scatterplot is drawn on a graphical device with geometrically equal scales on the  $X$  and  $Y$  axes.

**Value**

Invisible. None. Displays the plot(s) on the specified device.

**Note**

End-user plotting function.

**Author(s)**

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

**References**

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods*." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods*." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting*." J. Comp Graph. Statistics, 19(4):900-92.

**Examples**

```
#####
# Loading the library and its dependencies
#####
library("PRIMsrc")

#####
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#####
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
  cvtype = "combined", cvcriterion = "lrt",
  B = 1, K = 5,
  vs = TRUE, cpv = FALSE,
  decimals = 2, probval = 0.5,
  arg = "beta=0.05,
        alpha=0.1,
        minn=10,
        L=NULL,
        peelcriterion=\"lr\"",
  parallel = FALSE, conf = NULL, seed = 123)
```

```
plot(x = CVCOMBREP.synt1,
     main = paste("Scatter plot for model #1", sep=""),
     proj = c(1,2), splom = TRUE, boxes = TRUE,
     steps = CVCOMBREP.synt1$cvfit$cv.nsteps,
     pch = 16, cex = 0.5, col = 2,
     col.box = 2, lty.box = 2, lwd.box = 1,
     add.legend = TRUE,
     device = NULL, file = "Scatter Plot", path=getwd(),
     horizontal = FALSE, width = 5.0, height = 5.0)
```

---

plot\_boxkm

---

*Visualization of Survival Distributions*


---

## Description

Function for plotting the cross-validated survival distributions of a PRSP object. Plot the cross-validated Kaplan-Meier estimates of survival distributions for the highest risk (inbox) versus lower-risk (outbox) groups of samples at each iteration of the peeling sequence (inner loop of our PRSP algorithm).

## Usage

```
plot_boxkm(object,
            main = NULL,
            xlab = "Time",
            ylab = "Probability",
            precision = 1e-3,
            mark = 3,
            col = 2,
            cex = 1,
            steps = 1:object$cvfit$cv.nsteps,
            nr = 3,
            nc = 4,
            device = NULL,
            file = "Survival Plots",
            path=getwd(),
            horizontal = TRUE,
            width = 11.5,
            height = 8.5, ...)
```

## Arguments

|           |   |
|-----------|---|
| object    | Object of class PRSP as generated by the main function <a href="#">sbh</a> .                                  |
| main      | Character vector. Main Title. Defaults to NULL.   |
| xlab      | Character vector. X axis label. Defaults to "Time".   |
| ylab      | Character vector. Y axis label. Defaults to "Probability".  |
| precision | Precision of cross-validated log-rank p-values of separation between two survival curves. Defaults to 1e-3.   |
| mark      | Integer scalar of mark parameter, which will be used to label the inbox and out-of-box curves. Defaults to 3. |



|            |  |
|------------|--|
| col        | Integer scalar specifying the color of the inbox curve. Defaults to 2.   |
| cex        | Numeric scalar specifying the size of the marks. Defaults to 1.  |
| steps      | Integer vector. Vector of peeling steps at which to plot the survival curves. Defaults to all the peeling steps of PRSP object object.   |
| nr         | Integer scalar of the number of rows in the plot. Defaults to 3.   |
| nc         | Integer scalar of the number of columns in the plot. Defaults to 4.  |
| device     | Graphic display device in {NULL, "PS", "PDF"}. Defaults to NULL (standard output screen). Currently implemented graphic display devices are "PS" (Postscript) or "PDF" (Portable Document Format). |
| file       | File name for output graphic. Defaults to "Survival Plots".  |
| path       | Absolute path (without final (back)slash separator). Defaults to the working directory path.   |
| horizontal | Logical scalar. Orientation of the printed image. Defaults to TRUE, that is potrait orientation.   |
| width      | Numeric scalar. Width of the graphics region in inches. Defaults to 11.5.  |
| height     | Numeric scalar. Height of the graphics region in inches. Defaults to 8.5.  |
| ...        | Generic arguments passed to other plotting functions, including <code>plot.survfit</code> (R package <b>survival</b> ).  |

## Details

Some of the plotting parameters are further defined in the function `plot.survfit` (R package **survival**). Step #0 always corresponds to the situation where the starting box covers the entire test-set data before peeling. Cross-validated LRT, LHR of inbox samples and log-rank p-values of separation are shown at the bottom of the plot with the corresponding peeling step. P-values are lower-bounded by the precision limit given by  $1/A$ , where  $A$  is the number of permutations.

## Value

Invisible. None. Displays the plot(s) on the specified device.

## Note

End-user plotting function.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

## See Also

- `plot.survfit` (R package **survival**)

## Examples

```
#####
# Loading the library and its dependencies
#####
library("PRIMsrc")

#####
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#####
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 1, K = 5,
                      vs = TRUE, cpv = FALSE,
                      decimals = 2, probval = 0.5,
                      arg = "beta=0.05,
                           alpha=0.1,
                           minn=10,
                           L=NULL,
                           peelcriterion=\"lr\"",
                      parallel = FALSE, conf = NULL, seed = 123)

plot_boxkm(object = CVCOMBREP.synt1,
           main = paste("RCCV probability curves for model #1", sep=""),
           xlab = "Time", ylab = "Probability",
           device = NULL, file = "Survival Plots", path=getwd(),
           horizontal = TRUE, width = 11.5, height = 8.5)
```

---

plot\_boxtrace

*Visualization of Covariates Traces*

---

## Description

Function for plotting the cross-validated covariates traces of a PRSP object. Plot the cross-validated modal trace curves of covariate importance and covariate usage of the pre-selected covariates specified by user at each iteration of the peeling sequence (inner loop of our PRSP algorithm).

**Usage**

```
plot_boxtrace(object,
              main = NULL,
              xlab = "Box Mass",
              ylab = "Covariate Range (centered)",
              topplot = object$cvfit$cv.used,
              center = TRUE,
              scale = FALSE,
              col.cov,
              lty.cov,
              lwd.cov,
              col = 1,
              lty = 1,
              lwd = 1,
              cex = 1,
              add.legend = FALSE,
              text.legend = NULL,
              device = NULL,
              file = "Covariate Trace Plots",
              path=getwd(),
              horizontal = FALSE,
              width = 8.5,
              height = 8.5, ...)
```

**Arguments**

|         |  |
|---------|--|
| object  | Object of class PRSP as generated by the main function <a href="#">sbh</a> .   |
| main    | Character vector. Main Title. Defaults to.   |
| xlab    | Character vector. X axis label. Defaults to "Box Mass". NULL   |
| ylab    | Character vector. Y axis label. Defaults to "Covariate Range (centered)".  |
| topplot | Numeric vector. Which of the pre-selected covariates to plot (in reference to the original index of covariates). Defaults to covariates used for peeling.  |
| center  | Logical scalar. Shall the data be centered?. Defaults to TRUE.   |
| scale   | Logical scalar. Shall the data be scaled? Defaults to FALSE.   |
| col.cov | Integer vector. Line color for the covariate importance curve of each selected covariate. Defaults to vector of colors of length the number of selected covariates. The vector is reused cyclically if it is shorter than the number of selected covariates. |
| lty.cov | Integer vector. Line type for the covariate importance curve of each selected covariate. Defaults to vector of 1's of length the number of selected covariates. The vector is reused cyclically if it is shorter than the number of selected covariates.     |
| lwd.cov | Integer vector. Line width for the covariate importance curve of each selected covariate. Defaults to vector of 1's of length the number of selected covariates. The vector is reused cyclically if it is shorter than the number of selected covariates.    |
| col     | Integer scalar. Line color for the covariate trace curve. Defaults to 1.   |
| lty     | Integer scalar. Line type for the covariate trace curve. Defaults to 1.  |
| lwd     | Integer scalar. Line width for the covariate trace curve. Defaults to 1.   |

|             |  |
|-------------|--|
| cex         | Integer scalar. Symbol expansion used for titles, legends, and axis labels. Defaults to 1.   |
| add.legend  | Logical scalar. Should the legend be added to the current open graphics device?. Defaults to FALSE.  |
| text.legend | Character vector of legend content. Defaults to NULL.  |
| device      | Graphic display device in {NULL, "PS", "PDF"}. Defaults to NULL (standard output screen). Currently implemented graphic display devices are "PS" (Postscript) or "PDF" (Portable Document Format). |
| file        | File name for output graphic. Defaults to "Covariate Trace Plots".   |
| path        | Absolute path (without final (back)slash separator). Defaults to working directory path.   |
| horizontal  | Logical scalar. Orientation of the printed image. Defaults to FALSE, that is potrait orientation.  |
| width       | Numeric scalar. Width of the graphics region in inches. Defaults to 8.5.   |
| height      | Numeric scalar. Height of the graphics region in inches. Defaults to 8.5.  |
| ...         | Generic arguments passed to other plotting functions.  |

### Details

The trace plots limit the display of traces to those only covariates that are used for peeling. If centered, an horizontal black dotted line about 0 is added to the plot.

Due to the variability induced by cross-validation and replication, it is possible that more than one covariate be used for peeling at a given step. So, for simplicity of the trace plots, only the modal or majority vote trace value (over the folds and replications of the cross-validation) is plotted.

The top plot shows the overlay of covariate importance curves for each covariate. The bottom plot shows the overlay of covariate usage curves for each covariate. It is a discretized view of covariate importance.

Both point to the magnitude and order with which covariates are used along the peeling sequence.

### Value

Invisible. None. Displays the plot(s) on the specified device.

### Note

End-user plotting function.

### Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "Local Sparse Bump Hunting." J. Comp Graph. Statistics, 19(4):900-92.

## Examples

```
#####
# Loading the library and its dependencies
#####
library("PRIMsrc")

#####
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#####
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 1, K = 5,
                      vs = TRUE, cpv = FALSE,
                      decimals = 2, probval = 0.5,
                      arg = "beta=0.05,
                           alpha=0.1,
                           minn=10,
                           L=NULL,
                           peelcriterion=\"lr\"",
                      parallel = FALSE, conf = NULL, seed = 123)

plot_boxtrace(object = CVCOMBREP.synt1,
              main = paste("RCCV trace plots for model #1", sep=""),
              xlab = "Box Mass", ylab = "Covariate Range (centered)",
              topplot = CVCOMBREP.synt1$cvfit$cv.used,
              center = TRUE, scale = FALSE,
              device = NULL, file = "Covariate Trace Plots", path=getwd(),
              horizontal = FALSE, width = 8.5, height = 8.5)
```

---

plot\_boxtraj

---

Visualization of Peeling Trajectories/Profiles

---

## Description

Function for plotting the cross-validated peeling trajectories/profiles of a PRSP object. Applies to the pre-selected covariates specified by user and all other statistical quantities of interest at each iteration of the peeling sequence (inner loop of our PRSP algorithm).

**Usage**

```
plot_boxtraj(object,
             main = NULL,
             xlab = "Box Mass",
             ylab = "Covariate Range",
             topplot = object$cvfit$cv.used,
             col.cov,
             lty.cov,
             lwd.cov,
             col = 1,
             lty = 1,
             lwd = 1,
             cex = 1,
             add.legend = FALSE,
             text.legend = NULL,
             nr = NULL,
             nc = NULL,
             device = NULL,
             file = "Covariate Trajectory Plots",
             path=getwd(),
             horizontal = FALSE,
             width = 8.5,
             height = 11.5, ...)
```

**Arguments**

|         |  |
|---------|--|
| object  | Object of class PRSP as generated by the main function <a href="#">sbh</a> .   |
| main    | Character vector. Main Title. Defaults to NULL.  |
| xlab    | Character vector. X axis label. Defaults to "Box Mass".  |
| ylab    | Character vector. Y axis label. Defaults to "Covariate Range".   |
| topplot | Numeric vector. Which of the pre-selected covariates to plot (in reference to the original index of covariates). Defaults to covariates used for peeling.  |
| col.cov | Integer vector. Line color for the covariate trajectory curve of each selected covariate. Defaults to vector of colors of length the number of selected covariates. The vector is reused cyclically if it is shorter than the number of selected covariates. |
| lty.cov | Integer vector. Line type for the covariate trajectory curve of each selected covariate. Defaults to vector of 1's of length the number of selected covariates. The vector is reused cyclically if it is shorter than the number of selected covariates.     |
| lwd.cov | Integer vector. Line width for the covariate trajectory curve of each selected covariate. Defaults to vector of 1's of length the number of selected covariates. The vector is reused cyclically if it is shorter than the number of selected covariates.    |
| col     | Integer scalar. Line color for the trajectory curve of each statistical quantity of interest. Defaults to 1.   |
| lty     | Integer scalar. Line type for the trajectory curve of each statistical quantity of interest. Defaults to 1.  |
| lwd     | Integer scalar. Line width for the trajectory curve of each statistical quantity of interest. Defaults to 1.   |

|             |  |
|-------------|--|
| cex         | Integer scalar. Symbol expansion used for titles, legends, and axis labels. Defaults to 1.   |
| add.legend  | Logical scalar. Should the legend be added to the current open graphics device? Defaults to FALSE.   |
| text.legend | Character vector of legend content. Defaults to NULL.  |
| nr          | Integer scalar of the number of rows in the plot. If NULL, defaults to 3.  |
| nc          | Integer scalar of the number of columns in the plot. If NULL, defaults to 3.   |
| device      | Graphic display device in {NULL, "PS", "PDF"}. Defaults to NULL (standard output screen). Currently implemented graphic display devices are "PS" (Postscript) or "PDF" (Portable Document Format). |
| file        | File name for output graphic. Defaults to "Covariate Trajectory Plots".  |
| path        | Absolute path (without final (back)slash separator). Defaults to working directory path.   |
| horizontal  | Logical scalar. Orientation of the printed image. Defaults to FALSE, that is potrait orientation.  |
| width       | Numeric scalar. Width of the graphics region in inches. Defaults to 8.5.   |
| height      | Numeric scalar. Height of the graphics region in inches. Defaults to 8.5.  |
| ...         | Generic arguments passed to other plotting functions.  |

### Details

The plot limits the display of trajectories to those only covariates that are used for peeling.

The plot includes box descriptive statistics (such as support), survival endpoint statistics (such as Maximum Event-Free Time (MEFT), Minimum Event-Free Probability (MEVP), LHR, LRT) and prediction performance (such as CER).

### Value

Invisible. None. Displays the plot(s) on the specified device.

### Note

End-user plotting function.

### Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "Local Sparse Bump Hunting." J. Comp Graph. Statistics, 19(4):900-92.

## Examples

```
#####
# Loading the library and its dependencies
#####
library("PRIMsrc")

#####
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#####
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 1, K = 5,
                      vs = TRUE, cpv = FALSE,
                      decimals = 2, probval = 0.5,
                      arg = "beta=0.05,
                          alpha=0.1,
                          minn=10,
                          L=NULL,
                          peelcriterion=\"lr\"",
                      parallel = FALSE, conf = NULL, seed = 123)

plot_boxtraj(object = CVCOMBREP.synt1,
             main = paste("RCCV peeling trajectories for model #1", sep=""),
             xlab = "Box Mass", ylab = "Covariate Range",
             topplot = CVCOMBREP.synt1$cvfit$cv.used,
             device = NULL, file = "Covariate Trajectory Plots", path=getwd(),
             horizontal = FALSE, width = 8.5, height = 8.5)
```

## Description

Function for plotting the cross-validated tuning profiles of a PRSP object. It uses the user's choice of statistics among the Log Hazard Ratio (LHR), Log-Rank Test (LRT) or Concordance Error Rate (CER) as a function of the model tuning parameter, that is, the optimal number of peeling steps of the peeling sequence (inner loop of our PRSP algorithm).



**Usage**

```
plot_profile(object,
             main = NULL,
             xlab = "Peeling Steps",
             ylab = "Mean Profiles",
             add.sd = TRUE,
             add.legend = TRUE,
             add.profiles = TRUE,
             pch = 20,
             col = 1,
             lty = 1,
             lwd = 2,
             cex = 2,
             device = NULL,
             file = "Profile Plot",
             path=getwd(),
             horizontal = FALSE,
             width = 8.5,
             height = 5.0, ...)
```

**Arguments**

|              |  |
|--------------|--|
| object       | Object of class PRSP as generated by the main function <a href="#">sbh</a> .   |
| main         | Character vector. Main Title. Defaults to NULL.  |
| xlab         | Character vector. X axis label. Defaults to "Peeling Steps".   |
| ylab         | Character vector. Y axis label. Defaults to "Mean Profiles".   |
| add.sd       | Logical scalar. Shall the standard error bars be plotted? Defaults to TRUE.  |
| add.legend   | Logical scalar. Shall the legend be plotted? Defaults to TRUE.   |
| add.profiles | Logical scalar. Shall the individual profiles (for all replicates) be plotted? Defaults to TRUE.   |
| pch          | Integer scalar of symbol number for all the profiles. Defaults to 20.  |
| col          | Integer scalar of line color of the mean profile. Defaults to 1.   |
| lty          | Integer scalar of line type of the mean profile. Defaults to 1.  |
| lwd          | Integer scalar of line width of the mean profile. Defaults to 2.   |
| cex          | Integer scalar of symbol expansion for all the profiles. Defaults to 2.  |
| device       | Graphic display device in {NULL, "PS", "PDF"}. Defaults to NULL (standard output screen). Currently implemented graphic display devices are "PS" (Postscript) or "PDF" (Portable Document Format). |
| file         | File name for output graphic. Defaults to "Profile Plot".  |
| path         | Absolute path (without final (back)slash separator). Defaults to working directory path.   |
| horizontal   | Logical scalar. Orientation of the printed image. Defaults to FALSE, that is potrait orientation.  |
| width        | Numeric scalar. Width of the graphics region in inches. Defaults to 8.5.   |
| height       | Numeric scalar. Height of the graphics region in inches. Defaults to 5.  |
| ...          | Generic arguments passed to other plotting functions.  |

## Details

Model validation is done by applying the optimization criterion on the user's choice of specific statistic. The goal is to find the optimal value of the K-fold cross-validated number of steps by maximization of LHR or LRT, or minimization of CER.

Currently, this done internally for visualization purposes, but it will ultimately offer the option to do be interactive with the end-user as well for parameter choosing/model selection.

## Value

Invisible. None. Displays the plot(s) on the specified device.

## Note

End-user plotting function.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

## Examples

```
#=====
# Loading the library and its dependencies
#=====
library("PRIMsrc")

#=====
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#=====
CVCMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 1, K = 5,
```

```

vs = TRUE, cpv = FALSE,
decimals = 2, probval = 0.5,
arg = "beta=0.05,
      alpha=0.1,
      minn=10,
      L=NULL,
      peelcriterion=\"lr\"",
parallel = FALSE, conf = NULL, seed = 123)

plot_profile(object = CVCOMBREP.synt1,
             main = "RCCV tuning profiles for model #1",
             xlab = "Peeling Steps", ylab = "Mean Profiles",
             pch=20, col="black", lty=1, lwd=2, cex=2,
             add.sd = TRUE, add.legend = TRUE, add.profiles = TRUE,
             device = NULL, file = "Profile Plot", path=getwd(),
             horizontal = FALSE, width = 8.5, height = 5)

```

predict.PRSP

*Predict Function***Description**

S3-generic predict function to predict the box membership and box vertices on an independent set.

**Usage**

```

## S3 method for class 'PRSP'
predict(object, newdata, steps, na.action = na.omit, ...)

```

**Arguments**

|           |  |
|-----------|--|
| object    | Object of class PRSP as generated by the main function <a href="#">sbh</a> .   |
| newdata   | An object containing the new input data: either a numeric matrix or numeric vector. A vector will be transformed to a ( $\#$ sample x 1) matrix. |
| steps     | Integer vector. Vector of peeling steps at which to predict the box memberships and box vertices. Defaults to the last peeling step only.        |
| na.action | A function to specify the action to be taken if NAs are found. The default action is na.omit, which leads to rejection of incomplete cases.      |
| ...       | Further generic arguments passed to the predict function.  |

**Value**

List containing the following 2 fields:

|          |   |
|----------|---|
| boxind   | Logical matrix of predicted box membership indicator (columns) by peeling steps (rows). TRUE = in-box, FALSE = out-of-box.  |
| vertices | List of size the number of chosen peeling steps where each entry is a numeric matrix of predicted box vertices: lower and upper bounds (rows) by covariate (columns). |

**Note**

End-user predict function.

**Author(s)**

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

**References**

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

**Examples**

```
#=====
# Loading the library and its dependencies
#=====
library("PRIMsrc")

#=====
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#=====
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 1, K = 5,
                      vs = TRUE, cpv = FALSE,
                      decimals = 2, probval = 0.5,
                      arg = "beta=0.05,
                          alpha=0.1,
                          minn=10,
                          L=NULL,
                          peelcriterion=\"lr\"",
                      parallel = FALSE, conf = NULL, seed = 123)

n <- 100
p <- 3
x <- matrix(data=runif(n=n*p, min=0, max=1),
            nrow=n, ncol=p, byrow=FALSE,
```

```
dimnames=list(1:n, paste("X", 1:p, sep=""))
pred <- predict(object=CVCOMBREP.synt1, newdata=x, steps=CVCOMBREP.synt1$cvfit$cv.nsteps)
pred
```

PRIMsrc.news

*Display the **PRIMsrc** Package News***Description**

Function to display the log file NEWS of updates of the **PRIMsrc** package.

**Usage**

```
PRIMsrc.news(...)
```

**Arguments**

... Further arguments passed to or from other methods.

**Value**

None.

**Note**

End-user function.

**Author(s)**

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

**References**

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "Local Sparse Bump Hunting." J. Comp Graph. Statistics, 19(4):900-92.

---

print.PRSP

---

*Print Function*


---

## Description

S3-generic print function to display the cross-validated estimated values of the PRSP object.

## Usage

```
## S3 method for class 'PRSP'
print(x, ...)
```

## Arguments

**x**                      Object of class PRSP as generated by the main function [sbh](#).  
**...**                    Further generic arguments passed to the print function.

## Value

Display of the cross-validated fitted values of its argument.

## Note

End-user print function.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods*." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods*." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting*." J. Comp Graph. Statistics, 19(4):900-92.

## Examples

```
#####
# Loading the library and its dependencies
#####
library("PRIMsrc")

#####
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#####
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 1, K = 5,
                      vs = TRUE, cpv = FALSE,
                      decimals = 2, probval = 0.5,
                      arg = "beta=0.05,
                           alpha=0.1,
                           minn=10,
                           L=NULL,
                           peelcriterion=\"lr\"",
                      parallel = FALSE, conf = NULL, seed = 123)

print(CVCOMBREP.synt1)
```

Real.1

*Real Dataset #1: Clinical Dataset ( $p < n$  case)*

## Description

Publicly available HIV clinical data from the Women's Interagency HIV cohort Study (WIHS). Inclusion criteria of the study were that women at enrolment were (i) alive, (ii) HIV-1 infected, and (iii) free of clinical AIDS symptoms. Women were followed until the first of the following occurred: (i) treatment initiation (HAART), (ii) AIDS diagnosis, (iii) death, or administrative censoring. The studied outcomes were the competing risks "AIDS/Death (before HAART)" and "Treatment Initiation (HAART)". However, here, for simplification purposes, only the first of the two competing events (i.e. the time to AIDS/Death), was used in this dataset example. Likewise, the entire study enrolled 1164 women, but only the complete cases were used in this clinical dataset example for simplification. Variables included history of Injection Drug Use ("IDU") at enrollment, African American ethnicity ("Race"), age ("Age"), and baseline CD4 count ("CD4"). The question in this dataset example was whether it is possible to achieve a prognostication of patients for AIDS and HAART. See below Bacon et al. (2005) and the WIHS website for more details.

## Usage

Real.1

## Format

Dataset consists of a numeric data.frame containing  $n = 485$  complete observations (samples) by rows and  $p = 4$  clinical covariates by columns, not including the censoring indicator and (censored) time-to-event variables. It comes as a compressed Rda data file.

**Author(s)**

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

**Source**

See real data application in Dazard et al., 2015.

**References**

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.
- Bacon M.C, von Wyl V., Alden C. et al. 2005. "*Semi-Supervised Methods to Predict Patient Survival from Gene Expression Data.*" Clin Diagn Lab Immunol 12(9): 1013-1019.

**See Also**

[statepiaps.jhsph.edu/wihs/](http://statepiaps.jhsph.edu/wihs/)

---

Real.2

Real Dataset #2: Genomic Dataset ( $p \gg n$  case)

---

**Description**

Publicly available lung cancer genomic data from the Chemores Cohort Study. This data is part of an integrated study of mRNA, miRNA and clinical variables to characterize the molecular distinctions between squamous cell carcinoma (SCC) and adenocarcinoma (AC) in Non Small Cell Lung Cancer (NSCLC) aside large cell lung carcinoma (LCC). Tissue samples were analysed from a cohort of 123 patients who underwent complete surgical resection at the Institut Mutualiste Montsouris (Paris, France) between 30 January 2002 and 26 June 2006. In this genomic dataset, the expression levels of Agilent miRNA probes ( $p = 939$ ) were included from the  $n = 123$  samples of the Chemores cohort. The data contains normalized expression levels. See below the paper by Lazar et al. (2013) and Array Express data repository for complete description of the samples, tissue preparation, Agilent array technology, data normalization, etc. This dataset represents a situation where the number of covariates dominates the number of complete observations, or  $p \gg n$  case.



**Usage**

Real.2

**Format**

Dataset consists of a numeric `data.frame` containing  $n = 123$  complete observations (samples) by rows and  $p = 939$  genomic covariates by columns, not including the censoring indicator and (censored) time-to-event variables. It comes as a compressed Rda data file.

**Author(s)**

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

**Source**

See real data application in Dazard et al., 2015.

**References**

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods*." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods*." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting*." J. Comp Graph. Statistics, 19(4):900-92.
- Lazar V. et al. (2013). "*Integrated molecular portrait of non-small cell lung cancers*." BMC Medical Genomics 6:53-65. PMID: 24299561

**See Also**

Array Express data repository at the European Bioinformatics Institute. Accession number: #E-MTAB-1134 (MIR). [www.ebi.ac.uk/arrayexpress/](http://www.ebi.ac.uk/arrayexpress/)  
CHEMORES Consortium and website. [www.chemores.ki.se/index.html](http://www.chemores.ki.se/index.html)

sbh

*Cross-Validated Survival Bump Hunting***Description**

Main end-user function for fitting a cross-validated Survival Bump Hunting (SBH) model. Returns a cross-validated PRSP object, as generated by our Patient Recursive Survival Peeling or PRSP algorithm, containing cross-validated estimates of end-points statistics of interest.

**Usage**

```
sbh(dataset,
      B = 10, K = 5, A = 1000,
      vs = TRUE, cpv = FALSE, decimals = 2,
      cvtype=c("combined", "averaged", "none", NULL),
      cvcriterion=c("lrt", "cer", "lhr", NULL),
      arg = "beta=0.05,alpha=0.05,minn=5,L=NULL,peelcriterion=\"lr\"",
      probval = NULL, timeval = NULL,
      parallel = FALSE, conf = NULL, seed = NULL)
```

**Arguments**

|             |   |
|-------------|---|
| dataset     | data.frame or numeric matrix of input dataset containing the observed survival and status indicator variables in the first two columns, respectively, and all the covariates thereafter. If a data.frame is provided, it will be coerced to a numeric matrix.   |
| B           | Positive integer scalar of the number of replications of the cross-validation procedure. Defaults to 10.  |
| K           | Integer giving the number of folds (partitions) into which the observations should be randomly split for the cross-validation procedure. Setting K also specifies the type of cross-validation to be done: <ul style="list-style-type: none"> <li>• K = 1 carries no cross-validation out.</li> <li>• K in <math>\{2, \dots, n - 1\}</math> carries out eqnK-fold cross-validation.</li> <li>• K = n carries out leave-one-out cross-validation.</li> </ul> |
| A           | Positive integer scalar of the number of permutations for the computation of cross-validated p-values. Defaults to 1000.  |
| vs          | logical scalar. Flag for optional variable (covariate) pre-selection. Defaults to TRUE.   |
| cpv         | logical scalar. Flag for computation of permutation p-values. Defaults to FALSE.  |
| decimals    | integer scalar. Number of user-specified significant decimals to output results. Defaults to 2.   |
| cvtype      | Character vector describing the cross-validation technique in {"combined", "averaged", "none", NULL}. If NULL, automatically reset to "none".   |
| cvcriterion | character vector describing the optimization criterion in {"lrt", "lhr", "cer", NULL}. If NULL, automatically reset to "none".  |
| arg         | Character vector describing the PRSP parameters:  |

- `alpha` = fraction to peel off at each step. Defaults to 0.05.
- `beta` = minimum support size resulting from the peeling sequence. Defaults to 0.05.
- `minn` = minimum number of observation that we want to be able to detect in a box. Defaults to 5.
- `L` = fixed peeling length. Defaults to NULL.
- `peelcriterion` in {"hr" for Log-Hazard Ratio (LHR), "lr" for Log-Rank Test (LRT), "ch" for Cumulative Hazard Summary (CHS)}. Defaults to "lr".

Note that the parameters in `arg` come as a string of characters between double quotes, where all parameter evaluations are separated by commas (see example).

|                       |   |
|-----------------------|---|
| <code>probval</code>  | Numeric scalar of the survival probability at which we want to get the endpoint box survival time. Defaults to NULL.  |
| <code>timeval</code>  | Numeric scalar of the survival time at which we want to get the endpoint box survival probability. Defaults to NULL.  |
| <code>parallel</code> | Logical. Is parallel computing to be performed? Optional. Defaults to FALSE.  |
| <code>conf</code>     | List of parameters for cluster configuration. Inputs for R package <b>parallel</b> function <code>makeCluster</code> (R package <b>parallel</b> ) for cluster setup. Optional, defaults to NULL. See details for usage. |
| <code>seed</code>     | Positive integer scalar of the user seed to reproduce the results.  |

## Details

At this point, the main function `sbh` performs the search of the *first* box of the recursive coverage (outer) loop of our Patient Recursive Survival Peeling (PRSP) algorithm.

Also, the main function relies on an optional variable pre-selection procedure that is run before the PRSP algorithm. At this point, this is done by Elastic-Net (EN) penalization of the partial likelihood, where both mixing (`alpha`) and overall shrinkage (`lambda`) parameters are simultaneously estimated by cross-validation using the `glmnet::cv.glmnet` function of the R package **glmnet**.

The PRSP object contains cross-validated estimates of all the decision-rules of pre-selected covariates and all other statistical quantities of interest at each iteration of the peeling sequence (inner loop of the PRSP algorithm).

It enables the display of results graphically of/for model tuning/selection, all peeling trajectories, covariate traces and survival distributions (see plotting functions for more details).

The function offers a number of options for the type of cross-validation desired:  $K$ -fold (replicated)-averaged or-combined, as well as peeling and optimization criteria for model fitting, tuning and selection and a few more parameters for the PRSP algorithm.

In case a replicated cross-validation is performed, a "summary" of the outputs is done over the  $B$  replicates, which requires some explanation:

- Even though the PRSP algorithm uses only one covariate at a time at each peeling step, the reported matrix of "Replicated CV" box decision rules may show several covariates being used in a given step, simply because these decision rules are averaged over the  $B$  replicates (see equation #21 in Dazard et al. 2015). This is also reflected in the reported "Replicated CV" importance and usage plots of covariate traces.
- Likewise, the output matrix of "Replicated CV" box membership indicator does not necessarily match exactly the output vector of "Replicated CV" box support (and corresponding box sample size) for all peeling steps. The reason is that the reported "Replicated CV" box membership indicators are computed (at each peeling step) as the point-wise majority vote over

the  $B$  replicates (see equation #22 in Dazard et al. 2015), whereas the "Replicated CV" box support vector (and corresponding box sample size) is averaged (at each peeling step) over the  $B$  replicates.

The function takes advantage of the R package **parallel**, which allows users to create a cluster of workstations on a local and/or remote machine(s), enabling scaling-up with the number of CPU cores specified and efficient parallel execution.

Discrete (or nominal) covariates should be made (or re-arranged into) ordinal variables.

If the computation of cross-validated p-value is desired, then running with the parallelization option is generally advised as it may take a while. In the case of large ( $p > n$ ) or very large ( $p \gg n$ ) datasets, it is required to use the parallelization option preferably on a hyperperformance cluster of workstations.

To run a parallel session (and parallel RNG) of the PRIMsrc procedures (`parallel=TRUE`), argument `conf` is to be specified (i.e. non NULL). It must list the specifications of the following parameters for cluster configuration: "names", "cpus", "type", "homo", "verbose", "outfile". These match the arguments described in function `makeCluster` of the R package **parallel**. All fields are required to properly configure the cluster, except for "names" and "cpus", which are the values used alternatively in the case of a cluster of type "SOCK" (socket), or in the case of a cluster of type other than "SOCK" (socket), respectively. See examples below.

- "names": `names` : character vector specifying the host names on which to run the job. Could default to a unique local machine, in which case, one may use the unique host name "localhost". Each host name can potentially be repeated to the number of CPU cores available on the corresponding machine.
- "cpus": `spec` : integer scalar specifying the total number of CPU cores to be used across the network of available nodes, counting the workernodes and masternode.
- "type": `type` : character vector specifying the cluster type ("SOCK", "PVM", "MPI").
- "homo": `homogeneous` : logical scalar to be set to FALSE for inhomogeneous clusters.
- "verbose": `verbose` : logical scalar to be set to FALSE for quiet mode.
- "outfile": `outfile` : character vector of the output log file name for the workernodes.

Note that argument `B` is internally reset to `conf$cpus*ceiling(B/conf$cpus)` in case the parallelization is used (i.e. `conf` is non NULL), where `conf$cpus` denotes the total number of CPUs to be used (see above).

The actual creation of the cluster, its initialization, and closing are all done internally. In addition, when random number generation is needed, the creation of separate streams of parallel RNG per node is done internally by distributing the stream states to the nodes (For more details see function `makeCluster` (R package **parallel**) and/or <http://www.stat.uiowa.edu/~luke/R/cluster/cluster.html>).

The use of a seed allows to reproduce the results within the same type of session: the same seed will reproduce the same results within a non-parallel session or within a parallel session, but it will not necessarily give the exact same results (up to sampling variability) between a non-parallelized and parallelized session due to the difference of management of the seed between the two (see parallel RNG and value of retuned seed below).

## Value

Object of class PRSP (Patient Recursive Survival Peeling) List containing the following 19 fields:

|                    |  |
|--------------------|--|
| <code>x</code>     | numeric matrix of original dataset.                  |
| <code>times</code> | numeric vector of observed failure / survival times. |

|                |   |
|----------------|---|
| status         | numeric vector of observed event indicator in {1,0}.  |
| B              | positive integer of the number of replications used in the cross-validation procedure.  |
| K              | positive integer of the number of folds used in the cross-validation procedure.   |
| A              | positive integer of the number of permutations used for the computation of permutation p-values.  |
| vs             | logical scalar of returned flag of optional variable pre-selection.   |
| cpv            | logical scalar of returned flag of optional computation of permutation p-values.  |
| decimals       | integer of the number of user-specified significant decimals.   |
| cvtype         | character vector of the cross-validation technique used.  |
| cvcriterion    | character vector of optimization criterion used.  |
| arg            | character vector of the parameters used:  |
| probval        | Numeric scalar of survival probability used.  |
| timeval        | Numeric scalar of survival time used.   |
| cvfit          | List with 10 fields of cross-validated estimates: <ul style="list-style-type: none"> <li>• cv.maxsteps: numeric scalar of maximal ceiled-mean of number of peeling steps over the replicates.</li> <li>• cv.steps: numeric scalar of optimal number of peeling steps according to the optimization criterion.</li> <li>• cv.trace: numeric vector of the modal trace values of covariate usage for all peeling steps.</li> <li>• cv.boxind: logical matrix in TRUE, FALSE of individual observation box membership indicator (columns) for all peeling steps (rows).</li> <li>• cv.rules: data.frame of decision rules on the covariates (columns) for all peeling steps (rows).</li> <li>• cv.signnumeric vector in {-1,+1} of directions of peeling for all pre-selected covariates.</li> <li>• cv.selectednumeric vector of pre-selected covariates in reference to original index.</li> <li>• cv.usednumeric vector of covariates used for peeling in reference to original index.</li> <li>• cv.stats: numeric matrix of box endpoint quantities of interest (columns) for all peeling steps (rows).</li> <li>• cv.pval: numeric vector of log-rank permutation p-values of separation of survival distributions.</li> </ul> |
| cvprofiles     | List of ( $B$ ) of numeric vectors, one for each replicate, of the cross-validated statistics used in the optimization criterion (one set by user) as a function of the number of peeling steps.  |
| cvmeanprofiles | List of numeric vectors of the cross-validated mean statistics over the replicates. used in the optimization criterion (one set by user) as a function of the number of peeling steps.  |
| plot           | logical scalar of the returned flag for plotting results (TRUE if CV successful).   |
| config         | List with 7 fields of parameters used for configuring the parallelization including parallel and conf.  |
| seed           | User seed(s) used: integer of a single value, if parallelization is used integer vector of values, one for each replication, if parallelization is not used.  |

## Note

Unique end-user function for fitting the Survival Bump Hunting model.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods*." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods*." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting*." J. Comp Graph. Statistics, 19(4):900-92.

## See Also

- makeCluster (R package **parallel**)
- cv.glmnet (R package **glmnet**)
- glmnet (R package **glmnet**)

## Examples

```
#####
# Loading the library and its dependencies
#####
library("PRIMsrc")

#####
# Package news
# Package citation
#####
PRIMsrc.news()
citation("PRIMsrc")

#####
# Demo with a synthetic dataset
# Use help for descriptions
#####
data("Synthetic.1", package="PRIMsrc")
?Synthetic.1

#####
```

```

# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
# Without parallelization
# Without computation of permutation p-values
#=====
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 1, K = 5,
                      vs = TRUE, cpv = FALSE,
                      decimals = 2, probval = 0.5,
                      arg = "beta=0.05,
                           alpha=0.1,
                           minn=10,
                           L=NULL,
                           peelcriterion=\"lr\"",
                      parallel = FALSE, conf = NULL, seed = 123)

summary(CVCOMBREP.synt1)
print(CVCOMBREP.synt1)

## Not run:
#=====
# Examples of parallel backend parametrization
#=====
# Example #1 - 1-Quad (4-core double threaded) PC
# Running WINDOWS
# With SOCKET communication
#=====
if (.Platform$OS.type == "windows") {
  cpus <- detectCores()
  conf <- list("names" = rep("localhost", cpus),
              "cpus" = cpus,
              "type" = "SOCK",
              "homo" = TRUE,
              "verbose" = TRUE,
              "outfile" = "")
}
#=====
# Example #2 - 1 master node + 3 worker nodes cluster
# All nodes equipped with identical setups and multicores
# Running LINUX
# With SOCKET communication
#=====
if (.Platform$OS.type == "unix") {
  masterhost <- Sys.getenv("HOSTNAME")
  slavehosts <- c("compute-0-0", "compute-0-1", "compute-0-2")
  nodes <- length(slavehosts) + 1
  cpus <- 8
  conf <- list("names" = c(rep(masterhost, cpus),
                           rep(slavehosts, cpus)),
              "cpus" = nodes * cpus,
              "type" = "SOCK",
              "homo" = TRUE,
              "verbose" = TRUE,
              "outfile" = "")
}

```

```

}
#####
# Example #3 - Multinode multicore per node cluster
# Running LINUX
# with MPI communication
# Here, a file named ".nodes" (e.g. in the home directory)
# contains the list of nodes of the cluster
#####
if (.Platform$OS.type == "unix") {
  hosts <- scan(file=paste(Sys.getenv("HOME"), "/.nodes", sep=""),
               what="",
               sep="\n")
  hostnames <- unique(hosts)
  nodes <- length(hostnames)
  cpus <- length(hosts)/length(hostnames)
  conf <- list("cpus" = nodes * cpus,
              "type" = "MPI",
              "homo" = TRUE,
              "verbose" = TRUE,
              "outfile" = "")
}

#####
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
# With parallelization
# With computation of permutation p-values
#####
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
                      cvtype = "combined", cvcriterion = "lrt",
                      B = 10, K = 5, A = 1024,
                      vs = TRUE, cpv = TRUE,
                      decimals = 2, probval = 0.5,
                      arg = "beta=0.05,
                           alpha=0.1,
                           minn=10,
                           L=NULL,
                           peelcriterion=\"lr\"",
                      parallel = TRUE, conf = conf, seed = 123)

summary(CVCOMBREP.synt1)
print(CVCOMBREP.synt1)

## End(Not run)

```

summary.PRSP

*Summary Function***Description**

S3-generic summary function to summarize the main parameters used to generate the PRSP object.



**Usage**

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

**Arguments**

object            Object of class PRSP as generated by the main function [sbh](#).  
 ...              Further generic arguments passed to the summary function.

**Value**

Summarizes the main parameters used to generate its argument.

**Note**

End-user summary function.

**Author(s)**

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

**References**

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods*." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods*." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting*." J. Comp Graph. Statistics, 19(4):900-92.

**Examples**

```
#####
# Loading the library and its dependencies
#####
library("PRIMsrc")

#####
# Simulated dataset #1 (n=250, p=3)
# Replicated Combined Cross-Validation (RCCV)
# Peeling criterion = LRT
# Optimization criterion = LRT
#####
CVCOMBREP.synt1 <- sbh(dataset = Synthetic.1,
```

```

      cvtype = "combined", cvcriterion = "lrt",
      B = 1, K = 5,
      vs = TRUE, cpv = FALSE,
      decimals = 2, probval = 0.5,
      arg = "beta=0.05,
            alpha=0.1,
            minn=10,
            L=NULL,
            peelcriterion=\"lr\"",
      parallel = FALSE, conf = NULL, seed = 123)

summary(CVCOMBREP.synt1)

```

Synthetic.1

*Synthetic Dataset #1:  $p < n$  case***Description**

Dataset from simulated regression survival model #1 as described in Dazard et al. (2015). Here, the regression function uses all of the predictors, which are also part of the design matrix. Survival time was generated from an exponential model with rate parameter  $\lambda$  (and mean  $\frac{1}{\lambda}$ ) according to a Cox-PH model with hazard  $\exp(\eta)$ , where  $\eta(\cdot)$  is the regression function. Censoring indicator were generated from a uniform distribution on  $[0, 3]$ . In this synthetic example, all covariates are continuous, i.i.d. from a multivariate uniform distribution on  $[0, 1]$ .

**Usage**

Synthetic.1

**Format**

Each dataset consists of a numeric matrix containing  $n = 250$  observations (samples) by rows and  $p = 3$  variables by columns, not including the censoring indicator and (censored) time-to-event variables. It comes as a compressed Rda data file.

**Author(s)**

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

**Source**

See simulated survival model #1 in Dazard et al., 2015.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods." (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods." In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "Local Sparse Bump Hunting." J. Comp Graph. Statistics, 19(4):900-92.

---

 Synthetic.1b

---

 Synthetic Dataset #1b:  $p < n$  case
 

---

## Description

Dataset from simulated regression survival model #1b as described in Dazard et al. (2015). Here, the regression function uses all of the predictors, which are also part of the design matrix. In this example, the signal is limited to a box-shaped region  $R$  of the predictor space. Survival time was generated from an exponential model with rate parameter  $\lambda$  (and mean  $\frac{1}{\lambda}$ ) according to a Cox-PH model with hazard  $\exp(\eta)$ , where  $\eta(\cdot)$  is the regression function. Censoring indicator were generated from a uniform distribution on  $[0, 3]$ . In this synthetic example, all covariates are continuous, i.i.d. from a multivariate uniform distribution on  $[0, 1]$ .

## Usage

Synthetic.1b

## Format

Each dataset consists of a numeric matrix containing  $n = 250$  observations (samples) by rows and  $p = 3$  variables by columns, not including the censoring indicator and (censored) time-to-event variables. It comes as a compressed Rda data file.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## Source

See simulated survival model #1b in Dazard et al., 2015.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

---

Synthetic.2

*Synthetic Dataset #2:  $p < n$  case*

---

## Description

Dataset from simulated regression survival model #2 as described in Dazard et al. (2015). Here, the regression function uses some informative predictors. The rest represent un-informative noisy covariates, which are not part of the design matrix. Survival time was generated from an exponential model with rate parameter  $\lambda$  (and mean  $\frac{1}{\lambda}$ ) according to a Cox-PH model with hazard  $\exp(\eta)$ , where  $\eta(\cdot)$  is the regression function. Censoring indicator were generated from a uniform distribution on  $[0, 3]$ . In this synthetic example, all covariates are continuous, i.i.d. from a multivariate uniform distribution on  $[0, 1]$ .

## Usage

Synthetic.2

## Format

Each dataset consists of a numeric matrix containing  $n = 250$  observations (samples) by rows and  $p = 3$  variables by columns, not including the censoring indicator and (censored) time-to-event variables. It comes as a compressed Rda data file.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## Source

See simulated survival model #2 in Dazard et al., 2015.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

---

 Synthetic.3

---

 Synthetic Dataset #3:  $p < n$  case
 

---

## Description

Dataset from simulated regression survival model #3 as described in Dazard et al. (2015). Here, the regression function does not include any of the predictors. This means that none of the covariates is informative (noisy), and are not part of the design matrix. Survival time was generated from an exponential model with rate parameter  $\lambda$  (and mean  $\frac{1}{\lambda}$ ) according to a Cox-PH model with hazard  $\exp(\eta)$ , where  $\eta(\cdot)$  is the regression function. Censoring indicator were generated from a uniform distribution on  $[0, 3]$ . In this synthetic example, all covariates are continuous, i.i.d. from a multivariate uniform distribution on  $[0, 1]$ .

## Usage

Synthetic.3

## Format

Each dataset consists of a numeric matrix containing  $n = 250$  observations (samples) by rows and  $p = 3$  variables by columns, not including the censoring indicator and (censored) time-to-event variables. It comes as a compressed Rda data file.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## Source

See simulated survival model #3 in Dazard et al., 2015.

## References

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

---

Synthetic.4

*Synthetic Dataset #4:  $p > n$  case*

---

## Description

Dataset from simulated regression survival model #4 as described in Dazard et al. (2015). Here, the regression function uses 1/10 of informative predictors in a  $p > n$  situation with  $p = 1000$  and  $n = 100$ . The rest represents non-informative noisy covariates, which are not part of the design matrix. Survival time was generated from an exponential model with rate parameter  $\lambda$  (and mean  $\frac{1}{\lambda}$ ) according to a Cox-PH model with hazard  $\exp(\eta)$ , where  $\eta(\cdot)$  is the regression function. Censoring indicator were generated from a uniform distribution on  $[0, 2]$ . In this synthetic example, all covariates are continuous, i.i.d. from a multivariate standard normal distribution.

## Usage

Synthetic.4

## Format

Each dataset consists of a numeric matrix containing  $n = 100$  observations (samples) by rows and  $p = 1000$  variables by columns, not including the censoring indicator and (censored) time-to-event variables. It comes as a compressed Rda data file.

## Author(s)

- "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>
- "Michael Choe, M.D." <mjc206@case.edu>
- "Michael LeBlanc, Ph.D." <mleblanc@fhcrc.org>
- "Alberto Santana, MBA." <ahs4@case.edu>

Maintainer: "Jean-Eudes Dazard, Ph.D." <jxd101@case.edu>

Acknowledgments: This project was partially funded by the National Institutes of Health NIH - National Cancer Institute (R01-CA160593) to J-E. Dazard and J.S. Rao.

## Source

See simulated survival model #4 in Dazard et al., 2015.

**References**

- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2015). "*Cross-validation and Peeling Strategies for Survival Bump Hunting using Recursive Peeling Methods.*" (Submitted).
- Dazard J-E., Choe M., LeBlanc M. and Rao J.S. (2014). "*Cross-Validation of Survival Bump Hunting by Recursive Peeling Methods.*" In JSM Proceedings, Survival Methods for Risk Estimation/Prediction Section. Boston, MA, USA. American Statistical Association IMS - JSM, p. 3366-3380.
- Dazard J-E. and J. S. Rao (2010). "*Local Sparse Bump Hunting.*" J. Comp Graph. Statistics, 19(4):900-92.

# Index

## \*Topic **AIDS Prognostication**

Real.1, [23](#)

## \*Topic **Bump Hunting**

plot.PRSP, [5](#)  
plot\_boxkm, [8](#)  
plot\_boxtrace, [10](#)  
plot\_boxtraj, [13](#)  
plot\_profile, [16](#)  
predict.PRSP, [19](#)  
PRIMsrc-package, [2](#)  
print.PRSP, [22](#)  
sbh, [26](#)  
summary.PRSP, [32](#)

## \*Topic **Cross-Validation**

plot.PRSP, [5](#)  
plot\_boxkm, [8](#)  
plot\_boxtrace, [10](#)  
plot\_boxtraj, [13](#)  
plot\_profile, [16](#)  
predict.PRSP, [19](#)  
PRIMsrc-package, [2](#)  
print.PRSP, [22](#)  
sbh, [26](#)  
summary.PRSP, [32](#)

## \*Topic **Exploratory Survival/Risk Analysis**

plot.PRSP, [5](#)  
plot\_boxkm, [8](#)  
plot\_boxtrace, [10](#)  
plot\_boxtraj, [13](#)  
plot\_profile, [16](#)  
predict.PRSP, [19](#)  
PRIMsrc-package, [2](#)  
print.PRSP, [22](#)  
sbh, [26](#)  
summary.PRSP, [32](#)

## \*Topic **Non-Parametric Method**

plot.PRSP, [5](#)  
plot\_boxkm, [8](#)  
plot\_boxtrace, [10](#)  
plot\_boxtraj, [13](#)  
plot\_profile, [16](#)  
predict.PRSP, [19](#)

PRIMsrc-package, [2](#)

print.PRSP, [22](#)

sbh, [26](#)

summary.PRSP, [32](#)

## \*Topic **Real Dataset**

Real.1, [23](#)

Real.2, [24](#)

## \*Topic **Rule-Induction Method**

plot.PRSP, [5](#)  
plot\_boxkm, [8](#)  
plot\_boxtrace, [10](#)  
plot\_boxtraj, [13](#)  
plot\_profile, [16](#)  
predict.PRSP, [19](#)  
PRIMsrc-package, [2](#)  
print.PRSP, [22](#)  
sbh, [26](#)  
summary.PRSP, [32](#)

## \*Topic **Survival/Risk Estimation & Prediction**

plot.PRSP, [5](#)  
plot\_boxkm, [8](#)  
plot\_boxtrace, [10](#)  
plot\_boxtraj, [13](#)  
plot\_profile, [16](#)  
predict.PRSP, [19](#)  
PRIMsrc-package, [2](#)  
print.PRSP, [22](#)  
sbh, [26](#)  
summary.PRSP, [32](#)

## \*Topic **Tumor sample comparisons**

Real.2, [24](#)

## \*Topic **datasets**

Synthetic.1, [34](#)  
Synthetic.1b, [35](#)  
Synthetic.2, [36](#)  
Synthetic.3, [37](#)  
Synthetic.4, [38](#)

## \*Topic **documentation**

PRIMsrc.news, [21](#)

plot, [3](#)

plot (plot.PRSP), [5](#)

plot.PRSP, [5](#)



plot\_boxkm, [3](#), [8](#)  
plot\_boxtrace, [3](#), [10](#)  
plot\_boxtraj, [3](#), [13](#)  
plot\_profile, [3](#), [16](#)  
predict, [3](#)  
predict(predict.PRSP), [19](#)  
predict.PRSP, [19](#)  
PRIMsrc (PRIMsrc-package), [2](#)  
PRIMsrc-package, [2](#)  
PRIMsrc.news, [2](#), [21](#)  
print, [2](#)  
print(print.PRSP), [22](#)  
print.PRSP, [22](#)  
  
Real.1, [4](#), [23](#)  
Real.2, [4](#), [24](#)  
  
sbh, [3](#), [6](#), [8](#), [11](#), [14](#), [17](#), [19](#), [22](#), [26](#), [33](#)  
summary, [2](#)  
summary(summary.PRSP), [32](#)  
summary.PRSP, [32](#)  
Synthetic.1, [4](#), [34](#)  
Synthetic.1b, [4](#), [35](#)  
Synthetic.2, [4](#), [36](#)  
Synthetic.3, [4](#), [37](#)  
Synthetic.4, [4](#), [38](#)