Package ‘mixPHM’

July 23, 2015

Type Package
Title Mixtures of Proportional Hazard Models
Version 0.7-2
Date 2015-07-23
Description Fits multiple variable mixtures of various parametric proportional hazard models using the EM-Algorithm. Proportionality restrictions can be imposed on the latent groups and/or on the variables. Several survival distributions can be specified. Missing values and censored values are allowed. Independence is assumed over the single variables.

License GPL-2
Imports graphics, stats, survival, lattice
Depends R (>= 3.0.0)
Encoding UTF-8
LazyData yes
LazyLoad yes
ByteCompile yes
NeedsCompilation no
Author Patrick Mair [cre, aut], Marcus Hudec [aut]
Maintainer Patrick Mair <mair@fas.harvard.edu>
Repository CRAN
Date/Publication 2015-07-23 15:14:01

R topics documented:

mixPHM-package ................................................................. 2
msBIC .................................................................................. 2
phmclust .............................................................................. 4
plot_hazard ......................................................................... 6
screeBIC ............................................................................. 7
stableEM .............................................................................. 8
webshop .............................................................................. 10
WilcoxH ............................................................................. 10
Description

This package fits multiple variable mixtures of various parametric proportional hazard models using the EM-Algorithm. Proportionality restrictions can be imposed on the latent groups and/or on the variables. Several survival distributions can be specified. Missing and censored values are allowed. Independence is assumed over the single variables.

Details

Package: mixPHM
Type: Package
Version: 0.7-2
Date: 2015-07-23
License: GPL-2

Author(s)

Patrick Mair, Marcus Hudec
Maintainer: Patrick Mair <mair@fas.harvard.edu>

References


Description

This function fits models for different proportionality restrictions.
Usage

msBIC(x, K, method = "all", Sdist = "weibull", cutpoint = NULL, 
EMoption = "classification", EMstop = 0.01, maxiter = 100)

Arguments

x, K, method, Sdist, cutpoint, EMoption, EMstop, maxiter

Arguments

x
Data frame or matrix of dimension n*p with survival times (NA's allowed).

K
A vector with number of mixture components.

method
A vector with the methods provided in phmclust: With "separate" no restrictions are imposed. "main.g" relates to a group main effect, "main.p" to the variables main effects. "main.gp" reflects the proportionality assumption over groups and variables. "int.gp" allows for interactions between groups and variables. If method is "all", each model is fitted.

Sdist
Various survival distributions such as "weibull", "exponential", and "rayleigh".

cutpoint
Cutpoint for censoring

EMoption
"classification" is based on deterministic cluster assignment, "maximization" on deterministic assignment, and "randomization" provides a posterior-based randomized cluster assignment.

EMstop
Stopping criterion for EM-iteration.

maxiter
Maximum number of iterations.

Details

Based on the output BIC matrix, model selection can be performed in terms of the number of mixture components and imposed proportionality restrictions.

Value

Returns an object of class BICmat with the following values:

BICmat
Matrix with BIC values

K
Vector with different components

method
Vector with proportional hazard methods

Sdist
Survival distribution

See Also

cscreBIC

Examples

```r
## Fitting 3 Weibull proportional hazard models (over groups, pages) for K=2,3 components
data(webshop)
res <- msBIC(webshop, K = c(2,3), method = c("main.p","main.g"), maxiter = 10)
res
```
**phmclust**

*Fits mixtures of proportional hazard models*

**Description**

This function allows for the computation of proportional hazards models with different distribution assumptions on the underlying baseline hazard. Several options for imposing proportionality restrictions on the hazards are provided. This function offers several variations of the EM-algorithm regarding the posterior computation in the M-step.

**Usage**

```r
phmclust(x, K, method = "separate", Sdist = "weibull", cutpoint = NULL, EMstart = NA, EMoption = "classification", EMstop = 0.01, maxiter = 100)
```

**Arguments**

- **x**: Data frame or matrix of dimension n*p with survival times (NA’s allowed).
- **K**: Number of mixture components.
- **method**: Imposing proportionality restrictions on the hazards: With "separate" no restrictions are imposed, "main.g" relates to a group main effect, "main.p" to variable main effects. "main.gp" reflects the proportionality assumption over groups and variables. "int.gp" allows for interactions between groups and variables.
- **Sdist**: Various survival distributions such as "weibull", "exponential", and "rayleigh".
- **cutpoint**: Integer value with upper bound for observed dwell times. Above this cutpoint, values are regarded as censored. If NULL, no censoring is performed.
- **EMstart**: Vector of length n with starting values for group membership, NA indicates random starting values.
- **EMoption**: "classification" is based on deterministic cluster assignment, "maximization" on deterministic assignment, and "randomization" provides a posterior-based randomized cluster assignment.
- **EMstop**: Stopping criterion for EM-iteration.
- **maxiter**: Maximum number of iterations.

**Details**

The method "separate" corresponds to an ordinary mixture model. "main.g" imposes proportionality restrictions over variables (i.e., the group main effect allows for free-varying variable hazards). "main.p" imposes proportionality restrictions over groups (i.e., the variable main effect allows for free-varying group hazards). If clusters with only one observation are generated, the algorithm stops.
**phmclust**

### Value

Returns an object of class `mws` with the following values:

- **K**: Number of components
- **iter**: Number of EM iterations
- **method**: Proportionality restrictions used for estimation
- **Sdist**: Assumed survival distribution
- **likelihood**: Log-likelihood value for each iteration
- **pvisit**: Matrix of prior probabilities due to NA structure
- **se.pvisit**: Standard errors for priors
- **shape**: Matrix with shape parameters
- **scale**: Matrix with scale parameters
- **group**: Final deterministic cluster assignment
- **posteriors**: Final probabilistic cluster assignment
- **npar**: Number of estimated parameters
- **aic**: Akaike information criterion
- **bic**: Bayes information criterion
- **clmean**: Matrix with cluster means
- **se.clmean**: Standard errors for cluster means
- **clmed**: Matrix with cluster medians

### References


### See Also

`stableEM`, `msBIC`

### Examples

```r
data(webshop)

## Fitting a Weibull mixture model (3 components) is fitted with classification EM
## Observations above 600sec are regarded as censored

res1 <- phmclust(webshop, K = 3, cutpoint = 600)
res1
summary(res1)
```
# Fitting a Rayleigh Weibull proportional hazard model (2 components, proportional over groups)

```r
res2 <- phmclust(webshop, K = 2, method = "main.p", Sdist = "rayleigh")
res2
summary(res2)
```

---

## plot_hazard

### Plot functions

Plotting functions for hazard rates, survival times and cluster profiles.

### Usage

```r
plot_hazard(x, gr.subset, var.subset, group = TRUE, xlim = NA, ylim = NA,
            xlab = "Survival Time", ylab = "Hazard Function", main = "Hazard Functions", type = "l",
            lty = 1, lwd = 1, col = NA, legpos = "right", ...)
```

```r
plot_survival(x, gr.subset, var.subset, group = TRUE, xlim = NA, ylim = NA,
              type = "l", lty = 1, lwd = 1, col = NA, legpos = "right", ...)
```

```r
plot_profile(x, method = "mean", type = "b", pch = 19, lty = 1, lwd = 1, col = NA,
             xlab = "Variables", leglab = NA, ylab = NA, main = NA, legpos = "topright", ...)
```

### Arguments

- `x`: object of class `mws` from `phmclust`
- `gr.subset`: Optional vector for plotting subset of clusters
- `var.subset`: Optional vector for plotting subset of variables
- `group`: if TRUE hazard/survival plots are produced for each group, if FALSE for each variable
- `method`: "mean" for cluster mean profile plot and "median" for cluster median profile plot
- `xlim`: limits for x-axis
- `ylim`: limits for y-axis
- `xlab`: label for x-axis
- `ylab`: label for y-axis
- `main`: title of the plot
- `leglab`: label for the legend
- `type`: type of plot
- `lty`: line type
Description

This function produces a scree plot on the basis of the BIC values in msBIC.

Usage

```
screeBIC(x, lty = 1, col = NA, pch = 19, type = "b", main = "BIC Screeplot",
        xlab = "Number of Components", ylab = "BIC", legpos = "topright", ...)
```

Arguments

- `x`: Object of class `mws` from `msBIC`
- `lty`: Line type
- `col`: Line colors; if NA, colors are determined automatically
- `pch`: Value for plotting points

Examples

```r
# Plots for mixture Weibull model with 3 components
data(webshop)
res <- phmclust(webshop, 3)

# Hazard plot for first and third group, all pages
plot_hazard(res, gr.subset = c(1,3), group = TRUE, xlab = "Dwell Time")

# Survival plot for each group, first 6 pages
plot_survival(res, var.subset = 1:6, group = FALSE, xlab = "Dwell Time")

# Cluster profile plot
plot_profile(res, xlab = "Pages", ylab = "Mean Dwell Time", main = "Cluster Profile")
```
stableEM

<table>
<thead>
<tr>
<th>type</th>
<th>Type of plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>main</td>
<td>Plot title</td>
</tr>
<tr>
<td>xlab</td>
<td>Label for x-axis</td>
</tr>
<tr>
<td>ylab</td>
<td>Label for y-axis</td>
</tr>
<tr>
<td>legpos</td>
<td>position of the legend</td>
</tr>
<tr>
<td>...</td>
<td>Additional plot parameters</td>
</tr>
</tbody>
</table>

**See Also**

msBIC

**Examples**

```r
# Fitting all Weibull proportional hazard models for K=2,3,4 components
data(webshop)
res <- msBIC(webshop, K = c(2,3,4), method = "all", maxiter = 5)
screeBIC(res)
```

### Description

This function performs the clustering for different EM starting values in order to find a stable solution.

### Usage

```r
stableEM(x, K, numEMstart = 5, method = "separate", Sdist = "weibull", cutpoint = NULL, EMoption = "classification", EMstop = 0.0001, maxiter = 1000, print.likvec = TRUE)
```

### Arguments

- **x**: Data frame or matrix of dimension n*p with survival times (NA's allowed).
- **K**: Number of mixture components.
- **numEMstart**: Number of different starting solutions
- **method**: Imposing proportionality restrictions on the hazards: With separate no restrictions are imposed, main.g relates to a group main effect, main.p to the variables main effects. main.gp reflects the proportionality assumption over groups and variables. int.gp allows for interactions between groups and variables.
- **Sdist**: Various survival distributions such as weibull, exponential, and rayleigh.
- **cutpoint**: Integer value with upper bound for observed dwell times. Above this cutpoint, values are regarded as censored. If NULL, no censoring is performed.
**EMoption** classification is based on deterministic cluster assignment, maximization on deterministic assignment, and randomization provides a posterior-based randomized cluster assignment.

**EMstop** Stopping criterion for EM-iteration.

**maxiter** Maximum number of iterations.

**print.likvec** If TRUE the likelihood values for different starting solutions are printed.

**Details**

After the computation of the models for different starting solutions using the function `phmclust` the best model is chosen, i.e., the model with the largest likelihood value. The output values refer to this final model.

**Value**

Returns an object of class `mws` with the following values:

- **K** Number of components
- **iter** Number of EM iterations
- **method** Method with proportionality restrictions used for estimation
- **sdist** Assumed survival distribution
- **likelihood** Log-likelihood value for each iteration
- **pvisit** Matrix of prior probabilities due to NA structure
- **se.pvisit** Standard errors for priors
- **shape** Matrix with shape parameters
- **scale** Matrix with scale parameters
- **group** Final deterministic cluster assignment
- **posteriors** Final probabilistic cluster assignment
- **npar** Number of estimated parameters
- **aic** Akaike information criterion
- **bic** Bayes information criterion
- **clmean** Matrix with cluster means
- **se.clmean** Standard errors for cluster means
- **clmed** Matrix with cluster medians

**See Also**

- `phmclust`<br>`msbic`

**Examples**

```r
## Exponential mixture model with 2 components for 4 different starting solutions
data(webshop)
res <- stableEM(webshop, K = 2, numEMstart = 4, Sdist = "exponential")
res
summary(res)
```
### webshop

**Webshop dataset for mixPHM package**

**Description**

This artificial data set represents dwell times in seconds of 333 sessions on 7 webpage categories of a webshop. Missing values indicate that the corresponding session did not visit a particular page.

**Usage**

```r
data(webshop)
```

**Format**

Numeric matrices of data frames with subjects as rows and variables as columns. Missing values are coded as `NA` (which corresponds to 0 survival time).

**Examples**

```r
data(webshop)
str(webshop)
```

---

### WilcoxH

**Tests of Zero Correlations Among P Variables**

**Description**

This function computes Wilcoxon H-test and the Steiger-Hakstian-Test for testing $H_0: R = I$.

**Usage**

```r
WilcoxH(x, use = "pairwise.complete.obs")
```

**Arguments**

- `x` : Data frame or matrix of dimension $n*p$ with survival times (NA’s allowed).
- `use` : Treatment of NA’s for the computation of the correlation matrix (see `cor()`). Either "all.obs", "complete.obs", or "pairwise.complete.obs"

**Details**

This test is robust against violations of normality. Since `phmclust()` assumes independence across pages, this test can be used to explore the appropriateness of the data.
Value

Returns an object of class "wilcoxh" with the following values:

- `Rmat`: Correlation matrix
- `SH.res`: Results for Steiger-Hakstian-Test
- `WH.res`: Results for Wilcox H-test

References


See Also

- `phmclust`

Examples

```r
data(webshop)
res <- WilcoxH(webshop)
res
```
Index

*Topic **datasets**
  webshop, 10

*Topic **hplot**
  plot_hazard, 6
  screeBIC, 7

*Topic **models**
  msBIC, 2
  phmclust, 4
  stableEM, 8
  WilcoxH, 10

*Topic **package**
  mixPHM-package, 2

  logLik.mws (phmclust), 4

  mixPHM (mixPHM-package), 2
  mixPHM-package, 2
  msBIC, 2, 5, 8, 9
  phmclust, 4, 7, 9, 11
  plot_hazard, 6
  plot_profile (plot_hazard), 6
  plot_survival (plot_hazard), 6
  print.msBIC (msBIC), 2
  print.mws (phmclust), 4
  print.wilcoxh (WilcoxH), 10

  screeBIC, 3, 7
  stableEM, 5, 8
  summary.mws (phmclust), 4

  webshop, 10
  WilcoxH, 10