Package 'mixPHM'

October 13, 2022

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   *Mixtures of proportional hazard models*

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

Mair, P., and Hudec, M. (2009). Multivariate Weibull mixtures with proportional hazard restric-
tions for dwell time based session clustering with incomplete data. Journal of the Royal Statistical
Society, Series C (Applied Statistics), 58(5), 619-639.

Kalbfleisch, J.D., and Prentice, R.L. (1980). The statistical analysis of failure time data. New York:
Wiley.

Celaux, G., and Govaert, G. (1992). A classification EM algorithm for clustering and two stochastic
versions. Computational Statistics and Data Analysis, 14, 315-332.

---

msBIC   *PHM model selection with BIC*

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	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 distrubtions 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 assignement.
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

[screeBIC](#)

Examples

```
##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

```
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 distrubtions 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 assignement.
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.

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

Mair, P., and Hudec, M. (2009). Multivariate Weibull mixtures with proportional hazard restrictions for dwell time based session clustering with incomplete data. Journal of the Royal Statistical Society, Series C (Applied Statistics), 58(5), 619-639.

Celaux, G., and Govaert, G. (1992). A classification EM algorithm for clustering and two stochastic versions. Computational Statistics and Data Analysis, 14, 315-332.

See Also

stableEM, msBIC

Examples

```
data(webshop)

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

---

plot_hazard   *Plot functions*

Description

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

Usage

```
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", ...)

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

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

lwd	line width
pch	type of plotting points
col	colors; if NA it is determined in the function
legpos	position of the legend; "topright","topleft","bottomright", "bottomleft","left","right","top", or "center"
...	Additional plot options

See Also

phmclust

Examples

```
##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")
```

---

screeBIC   *Scree plot of BIC's*

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

type	Type of plot
main	Plot title
xlab	Label for x-axis
ylab	Label for y-axis
legpos	position of the legend
...	Additional plot parameters

See Also

[msBIC](#)

Examples

```
##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)
```

---

stableEM   *Stable EM solution*

Description

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

Usage

```
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 distrubtions 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 assignement.
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 propotionality 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](#),[msBIC](#)

Examples

```
## Exponental 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

```
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

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

WilcoxH   *Tests of Zero Correlations Among P Variables*

Description

This function computes Wilcox H-test and the Steiger-Hakstian-Test for testing H0: R = I.

Usage

```
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

Wilcox, R. (1997). Tests of independence and zero correlations among P variables. Biometrical Journal, 2, 183-193.

See Also

phmclust

Examples

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

Index