

# Package ‘bhm’

August 31, 2022

**Type** Package

**Title** Biomarker Threshold Models

**Version** 1.18

**Date** 2022-08-29

**Author** Bingshu E. Chen

**Maintainer** Bingshu E. Chen <bingshu.chen@queensu.ca>

**Depends** R (>= 3.5.0), coda, ggplot2, MASS, survival

**Imports** methods

**Description** Contains tools to fit both predictive and prognostic biomarker effects using biomarker threshold models and continuous threshold models. Evaluate the treatment effect, biomarker effect and treatment-biomarker interaction using probability index measurement. Test for treatment-biomarker interaction using residual bootstrap method.

**License** GPL (>= 2)

**LazyLoad** yes

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2022-08-31 17:00:05 UTC

## R topics documented:

bhm-package	2
bhm	3
bhmControl	6
brm	8
data	10
gImpLRT	11
llm	14
mpl	15
pIndex	17
pIndexControl	19
plot	21

print . . . . .	22
resboot . . . . .	23
rmscb . . . . .	25
rpicexp . . . . .	27
<b>Index</b>	<b>30</b>

---

bhm-package

*An R package for the biomarker threshold models*


---

## Description

This package fits biomarker threshold regression models for predictive and prognostic biomarker effects with binary data and survival data with an unknown biomarker cutoff point (Chen et al, 2014)<DOI:10.1016/j.csda.2013.05.015>. Multivariable models can also be fitted for adjusted biomarker effect (Fang et al, 2017)<DOI:10.1016/j.csda.2017.02.011>. Tools such as Probability index are included to measure treatment effect, biomarker effect or treatment-biomarker interaction(Jiang et al, 2016)<DOI:10.1002/sim.6907>.

## Details

"bhm" is a R package for Biomarker Threshold Models. Please use the following steps to install the most recent version of 'bhm' package:

1. First, you need to install the 'devtools' package. You can skip this step if you have 'devtools' installed in your R. Invoke R and then type

```
install.packages("devtools")
```

2. Load the devtools package.

```
library(devtools)
```

3. Install "bhm" package from github with R command

```
install_github("statapps/bhm")
```

"bhm" uses different statistical methods to identify cut-point (threshold parameter) for the biomarker in either generalized linear models or Cox proportional hazards model.

A stable version of View the "bhm" package is also available from the Comprehensive R Archive Network (<https://CRAN.R-project.org/package=bhm>) and can be installed using R command

```
install.packages("bhm")
```

## Author(s)

Bingshu E. Chen, Tian Fang, Jia Wang, Shuoshuo Liu

Maintainer: Bingshu E. Chen <[bingshu.chen@queensu.ca](mailto:bingshu.chen@queensu.ca)>

## References

- Chen, B. E., Jiang, W. and Tu, D. (2014). A hierarchical Bayes model for biomarker subset effects in clinical trials. *Computational Statistics and Data Analysis*. vol 71, page 324-334.
- Fang, T., Mackillop, W., Jiang, W., Hildesheim, A., Wacholder, S. and Chen, B. E. (2017). A Bayesian method for risk window estimatin with application to HPV vaccine trial. *Computational Statistics and Data Analysis*. 112, page 53-62.
- Jiang, S., Chen, B. E. and Tu, D.(2016). Inference on treatment-covariate interaction based on a nonparametric measure of treatment effects and censored survival data. *Statistics in Medicine*. 35, 2715-2725.
- Gavanji, P., Chen, B. E. and Jiang, W.(2018). Residual Bootstrap test for interactions in biomarker threshold models with survival data. *Statistics in Biosciences*.
- Chen, B. E. and Wang, J.(2020). Joint modelling of binary response and survival for clustered data in clinical trials. *Statistics in Medicine*. Vol 39. 326-339.
- Liu, S. S. and Chen, B. E. (2020). Continuous threshold models with two-way interactions in survival analysis. *Canadian Journal of Statistics*.

## See Also

[bhm](#), [brm](#), [coxph](#), [glm](#), [survival](#)

## Examples

```
# fit = bhm(y~biomarker+treatment)
# print(summary(fit))
```

---

bhm

*Fitting Biomarker Threshold Models*

---

## Description

{bhm} is a R package for Biomarker Threshold Models. It uses either Hierarchical Bayes method or profile likelihood method (Chen, et al, 2014 and Tian, et al, 2017) to identify a cut-point (thershold parameter) for the biomarker in either generalized linear models or Cox proportional hazards model. The model is specified by giving a symbolic description of the linear predictor and a description of the distribution family.

## Usage

```
bhm(x, ...)

## S3 method for class 'formula'
bhm(formula, family, data, control = list(...),...)

# use
#       bhm(y ~ biomarker)
```

```

#
# to fit a prognostic model with biomarker term only
#
# use
#
#       bhm(y ~ biomarker+treatment)
#
# to fit a predictive model with interaction between biomarker
# and treatment, use
#
#       bhmFit(x, y, family, control)
#
# to fit a model without the formula
#
# Biomarker shall be in the first dependent variable
#
# To summary a "bhm" object,
#
#       ## S3 method for class 'bhm'
summary(object, ...)

```

### Arguments

formula	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
family	a description of the response distribution and link function to be used in the model. The available family function are either "binomial" for fitting a logistic regression model or "surv" for fitting a Cox proportional hazards model
data	an optional data frame, list or environment (or object coercible by 'as.data.frame' to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which glm is called.
x, y	For "bhmFit", x is a design matrix of dimension n * p and y is a vector of observations of length n for "glm" models or a "Surv" survival object for "coxph" models.
control	a list of parameters for controlling the fitting process. See "bhmControl" for details
object	object returned from model fit
...	additional arguments to be passed to the low level regression fitting functions (see below)

### Details

'biomarker' is a Biomarker variable. This variable is required and shall be the first dependent variable in the formula.

"interaction" is an option of fitting model with interaction term. When interaction = TRUE, a predictive biomarker model will be fitted. When interaction = FALSE, a prognostic biomarker model will be fitted. Both Biomarker and Treatment variables are required if 'interaction' = TRUE and 'treatment' shall be the second variable in the formula.

"bhmFit" and "bhmGibbs" are the workhorse functions: they are not normally called directly but can be more efficient where the response vector, design matrix and family have already been calculated.

"x.cdf" is a function that maps biomarker values to interval (0, 1) using its empirical cumulative distribution function. After the threshold parameters are identified, the biomarker variable will be transformed back to its original scale.

## Value

bhm returns an object of class inheriting from "bhm" which inherits from the class glm or 'coxph'. See later in this section.

The function "summary" (i.e., "summary.bhm") can be used to obtain or print a summary of the results, for example, the 95

An object of class "bhm" is a list containing at least the following components:

c.max	a vector of the mean estimates for the threshold parameter(s)
coefficients	a named vector of coefficients from 'bhm'
c.fit	fitted conditional regression model given c = c.max
cg	Gibbs sample for threshold parameter c
bg	Gibbs sample for the coefficients beta

## Note

The logistic regression part are based on codes wrote by Tian Fang.

## Author(s)

Bingshu E. Chen (bingshu.chen@queensu.ca)

## References

Chen, B. E., Jiang, W. and Tu, D. (2014). A hierarchical Bayes model for biomarker subset effects in clinical trials. Computational Statistics and Data Analysis. vol 71, page 324-334.

## See Also

[glm](#), [coxph](#), [bhmControl](#)

## Examples

```
##
## Generate a random data set
n = 300
b = c(0.5, 1, 1.5)
data = gendat.surv(n, c0 = 0.40, beta = b)
```

```

age = runif(n, 0, 1)*100
tm = data[, 1]
status = data[, 2]
trt = data[, 3]
ki67 = data[, 4]
## fit a biomarker threshold survival model with one single cut point

# fit = bhm(Surv(tm, status)~ki67+trt+age, interaction = TRUE, B=5, R=10)

## here B=5 and R=10 is used for test run. In general, B > 500 and R > 2000 is
## recommend for the analysis of biomarker variable. To fit a model with
## two cut points, use:
##
##     fit = bhm(Surv(tm, status)~bmk+trt+age, B = 500, R = 2000, c.n = 2)
##
## To print the output, use
##
#     print(fit)
##

```

---

bhmControl

*Auxiliary function for bhm fitting*


---

## Description

Auxiliary function for `bhm` fitting. Typically only used internally by `'bhmFit'`, but may be used to construct a control argument to either function.

## Usage

```

bhmControl(method = 'Bayes', interaction, biomarker.main, alpha,
           B, R, thin, epsilon, c.n, beta0, sigma0)

```

## Arguments

<code>method</code>	choose either 'Bayes' for Bayes method with MCMC or 'profile' for profile likelihood method with Bootstrap. The default value is 'Bayes'
<code>interaction</code>	an option of fitting model with interaction term When <code>interaction = TRUE</code> , a predictive biomarker model will be fitted When <code>interaction = FALSE</code> , a prognostic biomarker model will be fitted The default value is <code>interaction = TRUE</code> .
<code>biomarker.main</code>	include biomarker main effect, default is <code>TRUE</code>
<code>B</code>	number of burn in
<code>R</code>	number of replications for Bayes meothd or number of Bootstrap for profile likelihood method
<code>thin</code>	thinning parameter for Gibbs samples, default is 2

epsilon	biomarker (transformed) step length for profile likelihood method, default is 0.01
alpha	significance level (e.g. alpha=0.05)
c.n	number of threshold (i.e. the cut point), default is 1
beta0	initial value for mean of the prior distribution of beta, default is 0
sigma0	initial value for variance of the prior distribution of beta, default is 10000

**Details**

Control is used in model fitting of "bhm".

**Value**

This function checks the internal consistency and returns a list of value as inputed to control model fit of bhm.

**Note**

Based on code from Tian Fang.

**Author(s)**

Bingshu E. Chen

**See Also**

[bhm](#)

**Examples**

```
## To fit a prognostic model for biomarker with two cut-points,  
## 500 burn-in samples and 10000 Gibbs samples,  
  
ctl = bhmControl(interaction = FALSE, B = 500, R = 10000, c.n = 2)  
  
##  
## then fit the following model  
##  
# fit = bhmFit(x, y, family = 'surv', control = ctl)  
##
```

**Description**

{brm} is an R function for Continuous Threshold Models. It uses the maximum likelihood method (Liu and Chen, 2020) to identify a cut-point (threshold parameter) for the biomarker in the Cox proportional hazards model. The model is specified by giving a symbolic description of the linear predictor and a description of the distribution family.

**Usage**

```
brm(x, ...)
```

```
## S3 method for class 'formula'
```

```
brm(formula, data = list(...), interaction = TRUE,
```

```
    method = c("gradient", "profile", "single"), q=1, epsilon = NULL, ...)
```

```
# use
```

```
#     brm(y ~ biomarker)
```

```
# or
```

```
#     brm(y ~ biomarker + x1 + x2, interactin = FALSE)
```

```
#
```

```
# to fit a prognostic model with biomarker term only (will be implement in the future)
```

```
#
```

```
# use
```

```
#
```

```
#     brm(y ~ biomarker+treatment+x1+x2+...)
```

```
#
```

```
# to fit a predictive model with interaciton between biomarker
```

```
# and treatment, adjusted for x1, x2, etc.
```

```
#
```

```
# use
```

```
#     brm(x, y, control, ...)
```

```
#
```

```
# to fit a model without formula
```

```
#
```

```
# Biomarker shall be in the first dependent variable
```

**Arguments**

formula	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
data	an optional data frame, list or environment (or object coercible by 'as.data.frame' to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which glm is called.



interaction	model with interaction term between biomarker and treatment variables. Default is TRUE.
method	method to fit a brm model, such as a gradient method or a single-index model. The default method is the "gradient" method.
q	number of biomarker variables. If $q > 1$ , a single-index model will be fitted. Default is $q = 1$ .
x	For "brm.default", x is a design matrix of dimension $n * p$ and y is a vector of observations of length n for a "Surv" survival object for "coxph" models.
...	additional arguments to be passed to the low level regression fitting functions (see below).
epsilon	Step width for the profile likelihood method, default is $(\max(w) - \min(w))/20$ .

### Details

'biomarker' is a Biomarker variable. This variable is required and shall be the first dependent variable in the formula.

"interaction" is an option of fitting model with interaction term. When interaction = TRUE, a predictive biomarker model will be fitted. When interaction = FALSE, a prognostic biomarker model will be fitted. Both Biomarker and Treatment variables are required if 'interaction' = TRUE and 'treatment' shall be the second variable in the formula.

"brm.default" is the workhorse functions: they are not normally called directly but can be more efficient where the response vector, design matrix and family have already been calculated.

### Value

brm returns an object of class inheriting from "brm" which inherits from the class glm or 'coxph'. See later in this section.

The function "summary" (i.e., "summary.brm") can be used to obtain or print a summary of the results, for example, the 95

An object of class "brm" is a list containing at least the following components:

coefficients	a named vector of coefficients from 'brm'
c.max	the maximum likelihood estimate for the threshold parameter(s).
theta	a vector contains both the coefficients and c.max.
var	the variance matrix of theta.
loglik	the log-likelihood with the final values of the coefficients.
linear.predictors	the vector of linear predictors, one per subject.
first	the first derivative vector at the solution.

### Author(s)

Shuoshuo Liu (shuoshuo.liu@psu.edu) and Bingshu E. Chen (bingshu.chen@queensu.ca)

## References

Liu, S. S. and Chen, B. E. (2020). Continuous threshold models with two-way interactions in survival analysis. *Canadian Journal of Statistics*. Vol. 48, page 751-772.

## See Also

[bhm](#), [coxph](#), [plot.brm](#), [print.brm](#), [residuals.brm](#), [summary.brm](#),

## Examples

```
##
## Generate a random data set
n = 100
b = c(0.5, 1, 1.5)
data = gendat.surv(n, c0 = 0.40, beta = b, type="brm")
age = runif(n, 0, 1)*100
tm = data[, 1]
status = data[, 2]
trt = data[, 3]
ki67 = data[, 4]
## fit a biomarker threshold survival model with one single cut point

fit = brm(Surv(tm, status)~ki67+trt+age)
##
## fit a prognostic continuous threshold model with biomarker only
##
# fit = brm(Surv(tm, status)~ki67)
##
## To print the output, use
##
# print(fit)
##
```

---

data

*dataset*

---

## Description

dataset for biomarker threshold model (bhm)

## Usage

# to generate survival data, use:

```
gendat.surv(n, c0, beta, type=c("brm", "bhm"))
```

# to generate glm data, use:

```
gendat.glm(n, c0, beta)
```

**Arguments**

n	sample size
c0	cut off point, for example c0 = 0.4
beta	regression coefficient, for example, beta = c(0.3, log(0.5), log(0.25))
type	type of biomarker threshold model, either bhm or brm, default is type = "brm"

**Format**

The format of the data set for analysis shall be a data frame with a response variable (either a Surv object for Cox model or a glm response variable object) and at least one dependent variable as the biomarker variable.

**Details**

data set of prostate cancer in the 'survival' package is used as an example in paper by Chen, et al. (2014).

**Source**

prostate dataset can be loaded with 'library(survival)'.

**References**

Chen, B. E., Jiang, W. and Tu, D. (2014). A hierarchical Bayes model for biomarker subset effects in clinical trials. Computational Statistics and Data Analysis. vol 71, page 324-334.

**Examples**

```
#data(data)
## maybe str(data) ; plot(data) ...
c0 = 0.4
b = c(-0.5, 1.5, 1.3)
data = gendat.surv(n=30, c0 = c0, beta = b)
```

---

glmpLRT

*Penalized likelihood ratio test for the generalized linear models.*


---

**Description**

{glmpLRT} is an R function for the penalized likelihood ratio test in generalized linear models. It uses the penalized likelihood method (Gavanji, Jiang and Chen, 2021) to identify a cut-point (threshold parameter) for the biomarker in the generalized linear models. The model is specified by giving a symbolic description of the linear predictor and a description of the distribution family.

**Usage**

```
## S3 method for class 'formula'
gImpLRT(formula, family = binomial,
  data=list(...), lambda = 15, c0 = 0.5, p1 = 1,
  method = c("pLRT", "Davies", "Bootstrap"),
  B=10, K = 50, epsilon = 0.025,...)
# use
#       gImpLRT(y ~ biomarker)
# or
#       gImpLRT(y ~ biomarker + x1 + x2, p1=0)
#
# to fit a prognostic model with biomarker term adjust for
# covariates x1, x2, etc.
#
# use
#
#       gImpLRT(y ~ biomarker+x1+x2+x3+x4+x5, p1=2...)
#
# to fit a predictive model with interacion between biomarker
# and x1, x2, adjusted for x3, x4, x5, etc.
#
# use
#       gImpLRT(x, y, control, ...)
#
# to fit a model without formula
#
# Biomarker shall be in the first dependent variable
```

**Arguments**

formula	an object of class "formula": a symbolic description of the model to be fitted. The details of model specificatoin are given under 'Details'.
family	a description of the error distribution and the link function to be used in the glm model. (See family for details of family functions.)
data	an optional data frame, list or enviroment containing the variables in the model.
method	the method to be used to fit a gImpLRT model. The default method is the "pLRT" method for the penalized likelihood ratio test. Other options include "Davis" for the Davis test and "Bootstrap" for the bootstrap test.
lambda	the tuning parameter for the penalized likelihood function, default is lambda = 15.
c0	the cut point for the threshold function, default is c0 = 0.5. Other options include c0 = 0.25 and c0 = 0.75.
p1	the number of covariates interact with the biomarker variable, the default is p1 = 1. If p1 = 0, then fit a model with biomarker main effect, adjusted for other potential covariates.

B	the number of bootstrap sample if the Bootstrap method is used.
K	smoothing parameter for the indicator function, the large value of K the better approximation of the indicator function, default is K = 50.
epsilon	Step width for the profile likelihood method, default is epsilon = 0.025.
...	additional arguments to be passed to the low level regression fitting functions (see below).

### Details

'biomarker' is a Biomarker variable. This variable is required and shall be the first dependent variable in the formula. It is not necessary to use the variable name 'biomarker', other variable name is allowed too.

'p1' controls the number of other covariates that interact with the biomarker variable.

"glmpLRT.default" is the workhorse functions: they are not normally called directly but can be more efficient where the response vector, design matrix and family have already been calculated.

"print", "plot" and "summary" methods can be applied to a fitted "glmpLRT" class to display the results.

### Value

glmpLRT returns an object of class inheriting from "glmpLRT" which inherits from the class glm. See later in this section.

An object of class "glmpLRT" is a list containing at least the following components:

coefficients	a named vector of coefficients from 'glmpLRT'
c.max	the maximum likelihood estimate for the threshold parameter(s).
loglik	the log-likelihood with the final values of the coefficients.
linear.predictors	the vector of linear predictors, one per subject.
mpv	p-value for the penalized likelihood ratio test.
rpv	p-value for the Davis test.
bpv	p-value for the Bootstrap test.

### Author(s)

Bingshu E. Chen (bingshu.chen@queensu.ca)

### References

Gavanji, P., Jiang, W. and Chen, B. E. (2021). Penalized likelihood ratio test for a biomarker threshold effect in clinical trials based on generalized linear models. Canadian Journal of Statistics.

### See Also

[glmpLRT](#), [glm](#), [plot.glmpLRT](#), [print.glmpLRT](#)

**Examples**

```
## A simulated example
bmk = rnorm(100, 3, 0.25)
age = rnorm(100, 50, 20)
trt = rbinom(100, 1, 0.5)
lp = exp(log(0.25) + 0.1*ifelse(bmk>2.5, 1, 0) + 0.69*trt)
p = lp/(1+lp)
y = rbinom(100, 1, p)
fit = glmPRT(y~bmk+trt+age, p1 = 0)
print(fit)
```

llm

*Fit an L-shape linear model***Description**

Fit an L-shape linear model with the cut point estimated by the profile likelihood method.

**Usage**

```
## S3 method for class 'formula'
llm(formula, data=list(...), epsilon = 0.025, ...)
```

**Arguments**

formula	an object of class "formula": a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
data	an optional data frame, list or environment containing the variables in the model.
epsilon	Step width for the profile likelihood method, default is 0.025
...	additional arguments to be passed to the low level regression fitting functions (see below).

**Details**

Define a L shape linear function that change slope at  $c_0$ :

when  $x < c_0$ ,  $y = b_1 + b_2 * x$

when  $x \geq c_0$ ,  $y = b_1 + b_2 * x + b_3 * (x - x_0) = (b_1 - b_3 * c_0) + (b_2 + b_3) * x$

**Value**

llm returns an object of class inheriting from "llm" which inherits from the class glm. See later in this section.

An object of class "llm" is a list containing at least the following components:

coefficients	a named vector of coefficients from 'llm'
residuals	the residuals, that is response minus fitted values.

fitted.values	the fitted mean values.
rank	the numeric rank of the fitted linear model.
df.residual	the residual degrees of freedom.
call	the matched call.
terms	the 'terms' object used.
c.max	the maximum likelihood estimate for the threshold parameter(s).
loglik	the log-likelihood with the final values of the coefficients.

**Author(s)**

Bingshu E. Chen (bingshu.chen@queensu.ca)

**References**

Liu, S. S. and Chen, B. E. (2020). Continuous threshold models with two-way interactions in survival analysis. *Canadian Journal of Statistics*. Vol. 48, page 751-772.

**See Also**

[brm](#), [lm](#), [glm](#)

**Examples**

```
#### simulate the data and fit a L-shape model.
n = 50 ; p <- 2
X = matrix(rnorm(n * p), n, p) # no intercept!
w = X[, 1]; age = X[, 2]

wc = w - 0.2; sigma = 0.25
y = rnorm(n, -0.1+0.7*w-1.2*ifelse(wc>0, wc, 0), sigma)

fit=llm(y~w+age)
print(fit)
print(summary(fit))
#### to plot the L-shape function
# plot(fit)
```

---

mpl

*Joint models for clustered data with binary and survival outcomes.*

---

**Description**

{mpl} is a function to fit a joint model for clustered binary and survival data using maximum penalized likelihood (MPL) method with Jackknife variance.

**Usage**

```

mpl(formula, ...)

## S3 method for class 'formula'
mpl(formula, formula.glm, formula.cluster, data, weights=NULL,
     subset = NULL, max.iter=100, tol = 0.005, jackknife=TRUE, ...)
#
# Use:
#
# fit = mpl(Surv(time, status)~w+z, y~x1+x2, ~cluster, data=data)
#

```

**Arguments**

formula	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the Cox proportiobal hazards model to be fitted for survival data.
formula.glm	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the generalized linear model to be fitted for binary data.
formula.cluster	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the cluster variable.
data	an optional data frame, list or environment (or object coercible by 'as.data.frame' to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the enviro nment from which mpl is called.
weights	an optional vector of weights to be used in the fitting process. Should be 'NULL' or a numeric vector. If non-NULL, weights options in the glm model and the coxph model will be assinged with the supplied weights.
subset	only a subset of data will be used for model fitting.
max.iter	Maximum number of iterations, default is max.iter = 100
tol	Tolrance for convergence, default is tol = 0.005
jackknife	Jackknife method for variance, default is jackknife = TRUE
...	additional arguments to be passed to the low level regression fitting functions (see below).

**Details**

mpl(Surv(time, event)~w+z, y~x1+x2, ~cluster) will fit penalized likelihood for binary and survival data with cluster effect. Function print(x) can be used to print a summary of mpl results.

**Value**

mpl returns an object of class inheriting from "mpl". When jackknife = TRUE, an object of class "mpl" is a list containing the following components:

theta	the maximum estimate of the regression coefficients and varaince component
-------	--



OR_HR	Odds ratios (OR) and hazard ratios (HR) for binary and survival outcomes, respectively
ase	Asymptotic standard error for theta, which is usually underestimated
jse	Jackknife standard error of theta based on resampling, this is considered to be more robust

**Note**

Based on code from J. Wang.

**Author(s)**

Bingshu E. Chen (bingshu.chen@queensu.ca)

**References**

Chen, B. E. and Wang, J. (2020). Joint modelling of binary response and survival for clustered data in clinical trials. *Statistics in Medicine*. Vol 39. 326-339.

**See Also**

[coxph](#), [glm](#), [print](#).

**Examples**

```
##
### No run
#
# fit = mpl(Surv(time, event)~trt+ki67, resp~trt+age, ~center.id)
#
```

---

pIndex

*Probability Index for Survival Time Difference*

---

**Description**

{pIndex} is a function to estimate and test difference of survival time among groups. It is defined as  $p = \Pr\{T_1 < T_2\}$ , where  $T_1$  is survival time for subjects in group 1 and  $T_2$  is survival time in group 2.

**Usage**

```
pIndex(x, ...)

## S3 method for class 'formula'
pIndex(formula, data, control = list(...),...)
###To estimate probability index for treatment and control groups (define by trt):
#
```

```
# fit = pIndex(Surv(time, status) ~ trt)
#
###To estimate probability index difference for treatment and control
###groups (define by trt) between biomarker positive and biomarker negative
###subjects(i.e. Treatment-biomarker interaction):
#
# fit = pIndex(Surv(time, status) ~ trt+biomarker)
#
```

### Arguments

formula	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
data	an optional data frame, list or environment (or object coercible by 'as.data.frame' to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which pIndex is called.
x	Here covariate x is a design matrix of dimension n * 1 (for two sample test) or dimension n * 2 (for treatment * biomarker interaction).
control	a list of parameters for controlling the fitting process. See 'pIndexControl' for details
...	additional arguments to be passed to the low level regression fitting functions (see below).

### Details

pIndex(y~x) will estimate probability index of two groups (eg. treatment vs control) define by x.  
 pIndex(y~x1 + x2) will estimate the difference of probability index of x1 (eg. treatment vs control) between biomarker positive and biomarker negative groups (x2). Function print(x) can be used to print a summary of pIndex results.

### Value

pIndex returns an object of class inheriting from "pIndex". When B > 0, an object of class "pIndex" is a list containing at least the following components:

theta	the estimated probability index
theta.b	Bootstrap or Jackknife sample of the probability index
sd	standard deviation of theta based on resampling
ci	(1-alpha) percent confidence interval based on resampling

### Note

This function is part of the bhm package.

### Author(s)

Bingshu E. Chen (bingshu.chen@queensu.ca)

## References

Jiang, S., Chen, B. E. and Tu, D.(2016). Inference on treatment-covariate interaction based on a nonparametric measure of treatment effects and censored survival data. *Statistics in Medicine*. 35, 2715-2725.

## See Also

[bhm](#), [pIndexControl](#),

## Examples

```
##
## Generate a random data set
n = 50
b = c(0.5, 1, 1.5)
data = gendat.surv(n, c0 = 0.40, beta = b, type='brm')
age = runif(n, 0, 1)*100
tm = data[, 1]
status = data[, 2]
trt = data[, 3]
ki67 = data[, 4]
#
#### No run
#
# fit = pIndex(Surv(tm, status) ~ trt + ki67)
#
```

---

pIndexControl

*Auxiliary function for pIndex fitting*

---

## Description

Auxiliary function for [pIndex](#) fitting. Typically only used internally by 'pIndexFit', but may be used to construct a control argument to either function.

## Usage

```
pIndexControl(method = c("Efron", "Elc", "Elw", "Pic"),
              model = c("default", "local", "threshold"),
              ci = c("Bootstrap", "Jackknife"), weights = NULL,
              kernel = NULL, h = 0.1, w = seq(0.05, 0.95, 0.05),
              alpha = 0.05, B = 0, pct = 0.5, tau=NULL)
```

## Arguments

method            choose either 'Efron' for Efron method, 'Elc' for conditional empirical likelihood, 'Elw' for weighted empirical likelihood method, and 'Pic' for piecewise exponential distribution. The default value is 'Efron'

model	'default' for default pIndex model, 'local' for kernel method, 'threshold' for threshold method
ci	Method to construct confidence interval, 'Bootstrap' for Bootstrap method and 'Jackknife' for Jackknife method
weights	case weight
kernel	kernel function types, including "gaussian", "epanechnikov", "rectangular", "triangular", "biweight", "cosine", "optcosine". The default value is 'gaussian'
h	bandwidth, default is 0.1
w	percentile of biomarker value for local fit
B	number of Bootstrap sample
alpha	significance level (e.g. alpha=0.05)
pct	Percentile of threshold (i.e. the cut point), default is 0.5
tau	maximum time tau to be used for pIndex

**Details**

Control is used in model fitting of 'pIndex'.

**Value**

This function checks the internal consistency and returns a list of value as inputed to control model fit of pIndex.

**Note**

Based on code from Bingshu E. Chen.

**Author(s)**

Bingshu E. Chen

**See Also**

[bhm](#), [pIndex](#)

**Examples**

```
## To calculate the probability index for a biomarker with conditional empirical likelihood method,
## and the corresponding 90 percent CI using Bootstrap method with 10000 bootstrap sample

ctl = pIndexControl(method = 'Elc', ci = 'Bootstrap', B = 10000, alpha = 0.1)

##
## then fit the following model
##
# fit = pIndex(y~x1 + x2, family = 'surv', control = ctl)
##
```

---

plot	<i>Plot a fitted biomarker threshold model</i>
------	--

---

### Description

Several different type of plots can be produced for biomarker threshold mdels. Plot method is used to provide a summary of outputs from "bhm", "pIndex", "resboot".

Use "methods(plot)" and the documentation for these for other plot methods.

### Usage

```
## S3 method for class 'bhm'
plot(x, type = c("profile", "density"), ...)
## S3 method for class 'brm'
plot(x, type = c("HR"), ...)
## S3 method for class 'pIndex'
plot(x, ...)
## S3 method for class 'resboot'
plot(x, ...)
## S3 method for class 'residuals.brm'
plot(x, type="Martingale", ...)
```

### Arguments

x	a class returned from "bhm", "pIndex" or "resboot" fit.
type	type of plot in bhm object, "profile" to plot profile likelihood, "density" to plot trace and density of the threshold distribution. "HR" to plot hazard ratio of the "brm" boject.
...	other options used in plot().

### Details

plot.bhm is called to plot either the profilelikelihood function or the threshold density function.

plot.pIndex is called to plot local probability index ([pIndex](#)) of a continuous biomarker.

plot.resboot is called to plot the bootstrap distribution of the likelihood ratio test statistics for biomarker threshold models ([resboot](#)).

The default method, plot.default has its own help page. Use methods("plot") to get all the methods for the plot generic.

### Author(s)

Bingshu E. Chen

### See Also

The default method for plot [plot.default](#). [glm](#) [bhm](#) [pIndex](#) [resboot](#)

**Examples**

```
#
# plot(fit)
#
##### plot for bhm object
#
# plot(fit, type = 'density')
#
```

---

print	<i>print a fitted object or a summary of fitted object</i>
-------	--

---

**Description**

print and summary are used to provide a short summary of outputs from "bhm", "brm", "mpl", "pIndex", "resboot".

**Usage**

```
## S3 method for class 'bhm'
print(x, ...)
## S3 method for class 'brm'
print(x, digits = 4, ...)
## S3 method for class 'mpl'
print(x, digits = 3, ...)
## S3 method for class 'pIndex'
print(x, ...)
## S3 method for class 'picreg'
print(x, digits=3, ...)
## S3 method for class 'resboot'
print(x, ...)
## S3 method for class 'summary.bhm'
print(x, ...)
```

**Arguments**

x	a class returned from bhm, pIndex or resboot fit
digits	number of digits to be printed
...	other options used in print()

**Details**

print.bhm is called to print object or summary of object from the biomarker threshold models [bhm](#). print.pIndex is called to print object or summary of object from the probability index model [pIndex](#). print.resboot is called to print object or summary of object from the residual bootstrap method for biomarker threshold models [resboot](#). summary(fit) provides detail summary of 'bhm' model fit, including parameter estimates, standard errors, and 95 percent CIs.

The default method, `print.default` has its own help page. Use `methods("print")` to get all the methods for the `print` generic.

### Author(s)

Bingshu E. Chen

### See Also

The default method for `print` is `print.default`. Other methods include `glm`, `bhm`, `brm`, `mpl`, `pIndex`, `resboot`.

### Examples

```
#
# print(fit)
#
```

---

resboot

*Residual Bootstrap Test (RBT) for treatment-biomarker interaction*

---

### Description

`{resboot}` is a function to test the existence of treatment-biomarker interaction in biomarker threshold model

$$g(Y) = b_0 + b_1 * I(w > c) + b_2 * z + b_3 * I(w > c) * z.$$

### Usage

```
resboot(x, ...)

## S3 method for class 'formula'
resboot(formula, family, data=list(...), B = 100, epsilon = 0.01, ...)
#
###To test the null hypothesis of interaction between treatment variable
###(define by z) and biomarker variables (define by w) for survival dataa,
###use:
#
# fit = resboot(Surv(time, status) ~ w + z + w:z)
#
```

### Arguments

<code>formula</code>	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
<code>family</code>	default is <code>family = 'Surv'</code> for survival data.

data	an optional data frame, list or environment (or object coercible by 'as.data.frame' to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which resboot is called.
x	Here covariate x is a design matrix of dimension $n * 1$ (for two sample test) or dimension $n * 2$ (for treatment * biomarker interaction).
B	Number of bootstraps, default is $B = 100$
epsilon	Biomarker (transformed) step length for profile likelihood method, default is $\epsilon = 0.01$
...	additional arguments to be passed to the low level regression fitting functions (see below).

### Details

resboot( $y \sim w + z + w:z$ ) will give residual bootstrap p-value for interaction between biomarker variable ( $w$ ) and treatment variable ( $z$ ). The null hypothesis is given by  $H_0: b_3 = 0$ , where  $b_3$  is the regression coefficient for the interaction term  $I(w > c) * z$ . Function print(x) can be used to print a summary of resboot results.

### Value

resboot returns an object of class inheriting from "resboot". When  $B > 0$ , an object of class "resboot" is a list containing at least the following components:

theta	the estimated maximum of likelihood ratio statistics
theta.b	Bootstrap sample of theta
sd	standard deviation of theta based on resampling
ci	(1-alpha) percent confidence interval for theta based on resampling

### Note

Based on code from Parisa Gavanji.

### Author(s)

Bingshu E. Chen (bingshu.chen@queensu.ca)

### References

Gavanji, P., Chen, B. E. and Jiang, W.(2018). Residual Bootstrap test for interactions in biomarker threshold models with survival data. Statistics in Biosciences.

### See Also

[bhm coxph](#)



**Examples**

```
##
## Generate a random data set
n = 30
b = c(0.5, 1, 1.5)
data = gendat.surv(n, c0 = 0.40, beta = b)
tm = data[, 1]
status = data[, 2]
trt = data[, 3]
ki67 = data[, 4]
#
### No run
#
# fit = resboot(Surv(tm, status) ~ ki67+trt+ki67:trt)
#
```

rmscb

*Fitting Restricted Mean Survival Time Models with a Continuous Biomarker*

**Description**

{rmscb} is an R function for restricted mean survival time (RMST) as a continuous function for a biomarker variables. The model is specified by giving a symbolic description of the linear predictor and a description of the distribution family.

**Usage**

```
rmscb(x, ...)
```

## S3 method for class 'formula'

```
rmscb(formula, data, subset, na.action, tau=5, h=0.2, w0=NULL,
      sig.level = 0.95, rho = 2,...)
```

# use

```
#       rmscb(y ~ biomarker)
```

#

# to fit a prognostic model with biomarker term only

#

# use

```
#       rmscb(y ~ biomarker + trt)
```

#

# to fit the difference of RMSTs between two treatment groups.

#

# use

## Default S3 method:

```
rmscb(x, y, control, ...)
```

#

# to fit a model without formula, where the biomarker shall be in the

# first dependent variable

**Arguments**

formula	an object of class "formula"(or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
data	an optional data frame, list or environment (or object coercible by 'as.data.frame' to a data frame) containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which "rmscb" is called.
tau	a prespecified time point at which the restricted mean survival time will be calculated.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the 'na.action' setting of 'options', and is 'na.fail' if that is unset. The 'factory-fresh' default is 'na.omit'. Another possible value is 'NULL', no action. Value 'na.exclude' can be useful.
h	the bandwidth, default is $h = 0.2$ , if $h = \text{NULL}$ , then the bandwidth will be selected using the cross validation method.
w0	the values of biomarker at which the RMST $E(T w=w_0)$ will be estimated.
sig.level	the significant level of the simultaneous confidence band will be constructed.
rho	the mode for the prediction error used in the cross validation bandwidth selection.
x	for 'rmscb.default', x is a design matrix of dimension $n * p$
y	y is a vector of observations of length n for a "Surv" survival object.
control	a list of parameters for controlling the fitting process. See "rmsControl" for details
...	additional arguments to be passed to the low level regression fitting functions (see below).

**Details**

'biomarker' is a Biomarker variable. This variable is required and shall be the first dependent variable in the formula.

'rmscb.default' is the workhorse functions: they are not normally called directly but can be more efficient where the response vector, design matrix and family have already been calculated.

**Value**

rmscb returns an object of class inheriting from "rmscb". See later in this section.

The function "summary" (i.e., "summary.rmscb") can be used to obtain or print a summary of the results, for example, the 95 percent CI of the parameters.

An object of class "rmscb" is a list containing at least the following components:

w0                      w0 from the input.

rms            a named vector of restricted mean survival time from the "rmscb".  
 LB            lower bound of the simultaneous confidence band.  
 UB            upper bound of the simultaneous confidence band.

### Author(s)

Wen Teng, Wenyu Jiang and Bingshu E. Chen (bingshu.chen@queensu.ca)

### References

Teng, W., Jiang, W. and Chen, B. E. (2022). Continuous threshold models with two-way interactions in survival analysis. *Statistics in Medicine*, submitted.

### Examples

```
##
## Generate a random data set
n = 100
age = runif(n, 0, 1)*100
tm = rexp(n, 1/10)
status = rbinom(n, 1, 0.5)
trt= rbinom(n, 1, 0.5)

## fit a restricted mean survival time with one biomarker

fit = rmscb(Surv(tm, status)~age)
print(fit)
## plot(fit)
## summary(fit)
```

---

 rpicexp

*The Piecewise Exponential Distribution*


---

### Description

Density, distribution function, quantile function, hazard function  $h(t)$ , cumulative hazard function  $H(t)$ , and random generation for the piecewise exponential distribution with rate equal to 'rate' and cut points equal to 'cuts'.

### Usage

```
dpicexp(x, rate=1, cuts=c(0, 10), log = FALSE)
ppicexp(q, rate=1, cuts=c(0, 10), lower.tail = TRUE, index = NULL)
qpicexp(p, rate=1, cuts=c(0, 10), lower.tail = TRUE)
rpicexp(n, rate=1, cuts=c(0, 10))

hpicexp(x, rate, cuts, index=NULL)
Hpicexp(x, rate, cuts, index=NULL)
```

```
#
## to fit a piece exponential survival model use:
#
# picfit(y, cuts=c(0, 10))
#
```

### Arguments

x, q	vector of quantiles.
p	vector of probabilities.
n	number of observations. If 'length(n) > 1', the length is taken to be the number required.
rate	vector rate parameter, defaulting to 1.
cuts	cut points, defaulting 0 to 10.
log	logical; if TRUE, probability p are given as log(p).
lower.tail	logical; if TRUE(default), probabilities are $P[X \leq x]$ , otherwise, $P[X > x]$ .
index	index of x, q in the interval defined by cuts, it saves time if index is known.

### Details

If the rate is not specified, it assumes the default value of 1.

### Value

dpicexp gives the density, ppicexp gives the distribution function, qplicexp gives the quantile function, and rplicexp generates random deviates.

The length of the result is determined by n for rplicexp.

Only the first elements of the logical arguments are used.

### Note

The cumulative hazard  $H(t) = -\log(1-F(t))$  is  $\log(1-\text{ppicexp}(t, \text{rate}, \text{cuts}))$ , or more efficiently call function  $\text{Hplicexp}(t, \text{rate}, \text{cuts})$ .

### Author(s)

Bingshu E. Chen (bingshu.chen@queensu.ca)

### References

Chen, B. E., Cook, R. J., Lawless, J. F. and Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine*. Vol 24, 671-691.

### See Also

[exp](#) for the exponential function.

[Distributions](#) for other standard distributions, including [dgamma](#) for the gamma distribution and [dweibull](#) for the Weibull distribution.

**Examples**

```
##  
### No run  
# n = 100  
# rate = c(1, 1, 0.5, 0.125)  
# cuts = c(0, 1, 2.5, 5, 10)  
# x = rpicexp(n, rate, cuts)  
#  
### compare rexp and rpicexp  
#  
#print(ppicexp(2.5, rate = .5))  
#print(pexp(2.5, rate = 0.5))  
#  
#
```

# Index

- \* **Biomarker interaction**
    - bhm, 3
    - brm, 8
    - gImpLRT, 11
    - pIndex, 17
    - resboot, 23
  - \* **Biomarker threshold models**
    - bhm-package, 2
  - \* **Biomarker**
    - rmscb, 25
  - \* **Continuous threshold models**
    - bhm-package, 2
    - brm, 8
  - \* **Cox regression**
    - mpl, 15
  - \* **Indicator threshold models**
    - bhm, 3
    - bhm-package, 2
  - \* **Jackknife**
    - mpl, 15
  - \* **Joint model**
    - mpl, 15
  - \* **Likelihood ratio test**
    - gImpLRT, 11
  - \* **Logistic regression**
    - bhm, 3
    - mpl, 15
  - \* **Penalized likelihood ratio test**
    - gImpLRT, 11
  - \* **Penalized likelihood**
    - mpl, 15
  - \* **Piecewise exponential distribution**
    - rpicexp, 27
  - \* **Piecewise exponential regression**
    - rpicexp, 27
  - \* **Piecewise exponential survival fit**
    - rpicexp, 27
  - \* **Predictive effect**
    - bhm, 3
    - brm, 8
    - gImpLRT, 11
    - resboot, 23
    - rmscb, 25
  - \* **Probability index**
    - bhm-package, 2
    - pIndex, 17
  - \* **Prognostic effect**
    - bhm, 3
    - brm, 8
    - gImpLRT, 11
    - rmscb, 25
  - \* **Residual Bootstrap**
    - bhm-package, 2
  - \* **Residual bootstrap**
    - resboot, 23
  - \* **Restricted mean survival time**
    - rmscb, 25
  - \* **control**
    - bhmControl, 6
    - pIndexControl, 19
  - \* **datasets**
    - data, 10
  - \* **linear model**
    - llm, 14
  - \* **plot**
    - plot, 21
  - \* **print**
    - print, 22
  - \* **summary**
    - plot, 21
    - print, 22
  - \* **threshold models**
    - llm, 14
- bhm, 3, 3, 6, 7, 10, 19–24  
bhm-doc (bhm-package), 2  
bhm-package, 2  
bhmControl, 5, 6  
bhmFit (bhm), 3

bhmGibbs (bhm), 3  
brm, 3, 8, 15, 23

coxph, 3, 5, 10, 17, 24

data, 10  
dgamma, 28  
Distributions, 28  
dpicexp (rpicexp), 27  
dweibull, 28

exp, 28

gendat.glm (data), 10  
gendat.surv (data), 10  
glm, 3, 5, 13, 15, 17, 21, 23  
glmpLRT, 11, 13

Hpicexp (rpicexp), 27  
hpicexp (rpicexp), 27

llm, 14  
lm, 15

mpl, 15, 23  
mplFit (mpl), 15

picfit (rpicexp), 27  
picreg (rpicexp), 27  
pIndex, 17, 19–23  
pIndexControl, 19, 19  
pIndexFit (pIndex), 17  
pIndexLocal (pIndex), 17  
pIndexThreshold (pIndex), 17  
plot, 21  
plot.brm, 10  
plot.default, 21  
plot.glmpLRT, 13  
plot.glmpLRT (glmpLRT), 11  
plot.llm (llm), 14  
plot.rmscb (rmscb), 25  
ppicexp (rpicexp), 27  
print, 17, 22  
print.brm, 10  
print.default, 23  
print.glmpLRT, 13  
print.glmpLRT (glmpLRT), 11  
print.llm (llm), 14  
print.rmscb (rmscb), 25  
prolikControl (bhmControl), 6  
prolikFit (bhm), 3  
qpicexp (rpicexp), 27  
resboot, 21–23, 23  
residuals.brm, 10  
residuals.brm (brm), 8  
rmscb, 25  
rmsControl (rmscb), 25  
rpicexp, 27  
summary.bhm (bhm), 3  
summary.brm, 10  
summary.brm (brm), 8  
summary.glmpLRT (glmpLRT), 11  
summary.picreg (rpicexp), 27  
summary.rmscb (rmscb), 25  
survival, 3  
thm.fit (bhm), 3  
thm.lik (bhm), 3  
x.cdf (bhm), 3