

# Package ‘PenIC’

January 9, 2020

**Type** Package

**Version** 1.0.0

**Title** Semiparametric Regression Analysis of Interval-Censored Data  
using Penalized Splines

**Maintainer** Yan Liu <yanliuresearch@gmail.com>

**Description** Currently incorporate the generalized odds-rate model (a type of linear transformation model) for interval-censored data based on penalized monotonic B-Spline. More methods under other semiparametric models such as cure model or additive model will be included in future versions. For more details see Lu, M., Liu, Y., Li, C. and Sun, J. (2019) <arXiv:1912.11703>.

**Depends** R (>= 3.3.0)

**License** GPL (>= 2)

**Imports** stats, numDeriv, splines, Matrix, MASS

**Encoding** UTF-8

**LazyData** true

**NeedsCompilation** no

**Author** Yan Liu [aut, cre],  
Minggen Lu [aut]

**Repository** CRAN

**Date/Publication** 2020-01-09 16:40:07 UTC

## R topics documented:

PenIC-package . . . . .	2
dataPA . . . . .	2
EM_fit . . . . .	3
<b>Index</b>	<b>6</b>

---

 PenIC-package

*A statistical package for regression analysis of interval-censored data under the generalized odds-rates model using penalized B-splines*


---

### Description

This package is designed to conduct the semiparametric regression analysis of interval-censored data under the generalized odds-rates model. To estimate the unknown nondecreasing cumulative baseline hazard function, monotone B-splines are used. An expectation maximization (EM) algorithm is developed to facilitate model fitting.

### Details

Package: PenIC  
 Type: Package  
 Version: 1.0.0  
 Date: 2019-12-11

### Author(s)

Yan Liu and Minggen Lu

---

 dataPA

*Date generation function*


---

### Description

Generate interval-censored data under generalized odds-rate model, with different combinations of right-censoring rate and cumulative baseline hazard function.

### Usage

```
dataPA(N, case, alpha)
```

### Arguments

N	size of dataset
case	data generation configuration; takes value in 1, 2 and 3.
alpha	parameter of link function; alpha=0 for the PH model and alpha=1 for the PO model.

**Details**

The above function generate interval-censored data from generalized odds-rate model, under different simulation configurations. For further details please see Lu et al. (2019+).

**Value**

d1	vector indicating whether an observation is left-censored (1) or not (0).
d2	vector indicating whether an observation is interval-censored (1) or not (0).
d3	vector indicating whether an observation is right-censored (1) or not (0).
Li	the left endpoint of the observed interval; if an observation is left-censored, its corresponding entry should be 0.
Ri	the right endpoint of the observed interval; if an observation is right-censored, its corresponding entry should be Inf.
Z	design matrix of predictor variables (in columns); should be specified without an intercept term.

**References**

Lu, M., Liu, Y., Li, C. and Sun, J. (2019+). An efficient penalized estimation approach for a semi-parametric linear transformation model with interval-censored data. arXiv:1912.11703.

**Examples**

```
case <- 3
nsub <- 100

# Generate interval-censored data under PH model

dat <- dataPA(nsub,case,alpha=0)
rp <- c(mean(dat$d1),mean(dat$d2),mean(dat$d3))
rp

# [1] 0.63 0.22 0.15
```

---

EM_fit	<i>EM algorithm for fitting generalized odds-rate model with specified link function (i.e., alpha value) under interval-censored data</i>
--------	---

---

**Description**

Fits the generalized odds-rate model based on penalized B-splines to interval censored data via an EM algorithm.

**Usage**

```
EM_fit(g0,b0,d1,d2,d3,Li,Ri,Z,nsub,alpha,qn,order,t.seq,tol=1e-5,itmax=500,lamu=1e5)
```

**Arguments**

$g_0$	initial estimate of the spline coefficients; should be of length $qn+order+1$ .
$b_0$	initial estimate of regression coefficients; should be of length $\dim(Z)[2]$ .
$d_1$	vector indicating whether an observation is left-censored (1) or not (0).
$d_2$	vector indicating whether an observation is interval-censored (1) or not (0).
$d_3$	vector indicating whether an observation is right-censored (1) or not (0).
$L_i$	the left endpoint of the observed interval; if an observation is left-censored, its corresponding entry should be 0.
$R_i$	the right endpoint of the observed interval; if an observation is right-censored, its corresponding entry should be Inf.
$Z$	design matrix of predictor variables (in columns); should be specified without an intercept term.
$n_{sub}$	size of observed dataset.
$\alpha$	parameter of link function; $\alpha=0$ for the PH model and $\alpha=1$ for the PO model.
$qn$	the number of interior knots to be used; should not exceed square root of sample size.
$order$	the order of the basis functions; $order=3$ for cubic spline.
$tol$	the convergence criterion of the EM algorithm.
$t.seq$	an increasing sequence of points at which the cumulative baseline hazard function is evaluated.
$itmax$	maximum iterations of EM procedure.
$\lambda_{mu}$	upper limit of smoothing parameter.

**Details**

The above function fits the generalized odds-rate model (with specified value of  $\alpha$ ) to interval censored data via an EM algorithm using penalized monotone B-splines.

**Value**

$b$	estimates of the regression coefficients.
$g$	estimates of the spline coefficients.
$se$	the standard deviation of $b$ .
$base$	estimated cumulative baseline hazard function evaluated at the points $t.seq$ .
$\lambda$	final value of smooth parameter.
$flag$	the indicator whether the procedure converged; 0 if converged.

**References**

Lu, M., Liu, Y., Li, C. and Sun, J. (2019+). An efficient penalized estimation approach for a semi-parametric linear transformation model with interval-censored data. arXiv:1912.11703.

**Examples**

```
set.seed(1)
case <- 2
nsub <- 35

# Generate interval-censored data under PH model

dat <- dataPA(nsub,case,alpha=0)
rp <- c(mean(dat$d1),mean(dat$d2),mean(dat$d3))
rp

# [1] 0.2571429 0.3428571 0.4000000

t.seq <- seq(0.01,4,0.01)

# number of interior knots to be used
qn <- ceiling(nsub^(1/3))-2
order <- 3
d1 <- dat$d1
d2 <- dat$d2
d3 <- dat$d3
Ri <- dat$Ri
Li <- dat$Li
Z <- dat$Z
p <- ncol(Z)
b0 <- rep(0,p)
g0 <- sort(runif(qn+order+1,-1,1))

# Fit data under PH model

fit <- EM_fit(g0,b0,d1,d2,d3,Li,Ri,Z,nsub,alpha=0,qn,order,t.seq,tol=1e-2,itmax=100,lamu=1e5)
cbind(fit$b,fit$se)

#           [,1]      [,2]
#[1,] -1.0655212 0.5021835
#[2,]  0.7649178 0.3185045
```

# Index

[dataPA](#), [2](#)

[EM\\_fit](#), [3](#)

[PenIC-package](#), [2](#)