Package ‘ICBayes’

September 24, 2017

Title Bayesian Semiparametric Models for Interval-Censored Data
Version 1.1
Date 2017-9-24
Author Chun Pan, Bo Cai, Lianming Wang, and Xiaoyan Lin
Maintainer Chun Pan <chunpan2003@hotmail.com>

Description
Contains functions to fit Bayesian semiparametric regression survival models (proportional hazards model, proportional odds model, and probit model) to interval-censored time-to-event data.

License GPL (>= 2)
LazyLoad yes
Depends HI, survival, coda
NeedsCompilation no
Repository CRAN
Date/Publication 2017-09-24 21:00:21 UTC

R topics documented:

ICBayes-package .................................................... 2
bdata ................................................................. 2
case1ph ............................................................. 3
case1po .............................................................. 5
case2ph ............................................................. 6
case2probit .......................................................... 8
coeff.ICBayes ....................................................... 9
ICBayes ............................................................. 10
logLik.ICBayes ....................................................... 12
lungdata ............................................................. 13
plot.ICBayes ......................................................... 14
summary.ICBayes ..................................................... 14
SurvtoLR ............................................................. 15

Index 16
ICBayes-package  

Bayesian Semiparametric Models for Interval-Censored Data

Description
This package contains functions to fit several survival regression models (including the proportional hazard model, the proportional odds model, and the probit model) for interval-censored data under Bayesian framework. Estimations are available for both regression coefficients and survival functions. The Bayesian model selection criterion log pseudo marginal likelihood (LPML) is computed.

Details

Package: ICBayes  
Type: Package  
Version: 1.1  
Date: 2016-12-24  
License: GPL>=2  
LazyLoad: yes

Author(s)
Chun Pan, Bo Cai, Lianming Wang, and Xiaoyan Lin  
Maintainer: Chun Pan <chunpan2003@hotmail.com>

bcdata  

Breast Cosmesis Data

Description
A general interval-censored data set analyzed in Finkelstein and Wolfe (1985) and can be found in Sun (2006, page 7). Early breast cancer patients treated with radiotherapy alone or radiotherapy with adjuvant chemotherapy were examined periodically for breast retraction. Time was measured in months.

Usage
data(bcdata)
**Format**

A matrix with 96 rows and 4 columns. Each row (L, R, status, x1) corresponds to a patient in the study.

- **L** a numeric vector of left-points of observed time intervals
- **R** a numeric vector of right-points of observed time intervals
- **status** a vector of censorship indicators: 0=left-censored, 1=interval-censored, and 2=right-censored
- **x1** a vector of treatment indicators: 0=radiotherapy alone, 1=radiotherapy with adjuvent chemotherapy

**Source**


**Examples**

```r
data(bcdata)
```

<table>
<thead>
<tr>
<th>case1ph</th>
<th>PH model for case 1 interval-censored data</th>
</tr>
</thead>
</table>

**Description**

Fit proportional hazards model for case 1 interval-censored data. Use MCMC method to estimate regression coefficients, baseline survival, and survival at user-specified covariate values.

**Usage**

```r
case1ph(L, R, status, xcov, x_user, order, sig0, coef_range, a_eta, b_eta, knots, grids, niter, seed)
```

**Arguments**

- **L** a numeric vector of left timepoints of observed time intervals.
- **R** a numeric vector of right timepoints of observed time intervals.
- **status** a vector of censoring indicators: 1=left-censored, 0=right-censored.
- **xcov** a matrix of covariates, each column corresponds to one covariate.
- **x_user** a user specified vector of covariate values
- **order** degree of I-splines (b_l) (see details). Recommended values are 2-4.
- **sig0** standard deviation of normal prior for each regression coefficient beta_r.
coef_range specify support domain of target density for beta_r sampled by arms (see details).

a_eta shape parameter of Gamma prior for gamma_l (see details).

b_eta rate parameter of Gamma prior for gamma_l (see details).

knots a sequence of points to define I-splines.

grids a sequence of points where baseline survival function is to be estimated.

niter total number of iterations of MCMC chains.

seed a user specified random seed, default is NULL.

**Details**

The baseline cumulative hazard is approximated by a linear combination of I-splines:

\[
\sum_{l=1}^{k}(gamma_l*b_lI).
\]

Function arms is used to sample each regression coefficient beta_r, and coef_range specifies the support of the indFunc in arms.

**Value**

a list containing the following elements:

parbeta a niter by p matrix of MCMC draws of beta_r, r=1, ..., p.

parsurv0 a niter by length(grids) matrix, each row contains the baseline survival at grids from one iteration.

parsurv a niter by length(grids)*G matrix, each row contains the survival at grids from one iteration. G is the number of sets of user-specified covariate values.

parfinv a niter by n matrix, each row contains the inverse PDF of observed interval-censored data from one iteration. This is used for computing LPML later.

grids a sequence of points where baseline survival is estimated.

**Author(s)**

Bo Cai

**References**

case1po

PO model for case 1 interval-censored data

Description

Fit proportional odds model for case 1 interval-censored data. Use MCMC method to estimate regression coefficients, baseline survival, and survival function at user-specified covariate values.

Usage

case1po(L, R, status, xcov, x_user, order, sig0, coef_range, a_eta, b_eta, knots, grids, niter, seed)

Arguments

L a numeric vector of left timepoints of observed time intervals.
R a numeric vector of right timepoints of observed time intervals.
status a vector of censoring indicators: 1=left-censored, 0=right-censored.
xcov a matrix of covariates, each column corresponds to one covariate.
x_user a vector of user specified covariate values.
order degree of I-splines (b_l) (see details). Recommended values are 2-4.
sig0 standard deviation of normal prior for each regression coefficient beta_r.
coef_range specify support domain of target density for beta_r sampled by arms (see details).
a_eta shape parameter of Gamma prior for gamma_1 (see details).
b_eta rate parameter of Gamma prior for gamma_1 (see details).
knots a sequence of points to define I-splines.
grids a sequence of points where baseline survival function is to be estimated.
niter total number of iterations of MCMC chains.
seed a user specified random seed, default is NULL.

Details

The baseline odds function is approximated by a linear combination of I-splines:

\[ \sum_{l=1}^{k} \gamma_l \cdot b_{l}. \]

Function arms is used to sample each regression coefficient beta_r, and coef_range specifies the support of the indFunc in arms.
Value

a list containing the following elements:

parbeta a niter by p matrix of MCMC draws of beta_r, r=1,...,p.
parsurv0 a niter by length(grids) matrix, each row contains the baseline survival at grids from one iteration.
parsurv a niter by length(grids)*G matrix, each row contains the survival at grids from one iteration. G is the number of sets of user-specified covariate values.
parinv a niter by n matrix, each row contains the inverse PDF of observed interval-censored data from one iteration. This is used for computing LPML later.

Author(s)

Xiaoyan Lin

References


description

Fit proportional hazards model for general interval-censored data. Use MCMC method to estimate regression coefficients, baseline survival, and survival function at user-specified covariate values.

Usage

case2ph(L, R, status, xcov, x_user, order, sig0, coef_range, a_eta, b_eta, knots, grids, niter, seed)

Arguments

L a numeric vector of left timepoints of observed time intervals.
R a numeric vector of right timepoints of observed time intervals.
status a vector of censoring indicators: 1=left-censored, 0=right-censored.
xcov a matrix of covariates, each column corresponds to one covariate.
x_user a user specified vector of covariate values.
order degree of I-splines (b_l) (see details). Recommended values are 2-4.
sig0 standard deviation of normal prior for each regression coefficient beta_r.
coef_range specify support domain of target density for beta_r sampled by arms (see details).
a_eta          shape parameter of Gamma prior for gamma_1 (see details).
b_eta          rate parameter of Gamma prior for gamma_1 (see details).
knots          a sequence of points to define I-splines.
grids          a sequence of points where baseline survival function is to be estimated.
niter          total number of iterations of MCMC chains.
seed           a user specified random seed, default is NULL.

Details

The baseline cumulative hazard is modeled by a linear combination of I-splines:

\[ \sum_{l=1}^k \gamma_l b_l I \]

Function arms is used to sample each regression coefficient beta_r, and coef_range specifies the support of the indFunc in arms.

Value

a list containing the following elements:

parbeta       a niter by p matrix of MCMC draws of beta_r, r=1, ..., p.
parsurv0      a niter by length(grids) matrix, each row contains the baseline survival at grids from one iteration.
parsurv       a niter by length(grids)*G matrix, each row contains the survival at grids from one iteration. G is the number of sets of user-specified covariate values.
parfinv       a niter by n matrix, each row contains the inverse PDF of observed interval-censored data from one iteration. This is used for computing LPML later.
grids         a sequence of points where baseline survival is estimated.

Author(s)

Bo Cai

References

case2probit  Probit model for general interval-censored data

Description

Fit probit model to general interval-censored data. Use MCMC method to estimate regression coefficients, baseline survival, and survival function at user-specified covariate values.

Usage

case2probit(l, R, status, xcov, x_user, order,
    v0, a_eta, b_eta, knots, grids, niter, seed)

Arguments

  L  a numeric vector of left timepoints of observed time intervals.
  R  a numeric vector of right timepoints of observed time intervals.
status  a vector of censoring indicators: 0=left-censored, 1=interval-censored, 2=right-censored.
xcov  a matrix of covariates, each column corresponds to one covariate.
x_user  a vector of user specified covariate values.
order  degree of I-splines (b_1) (see details). Recommended values are 2-4.
v0  precision of normal prior for \gamma_0.
a_eta  shape parameter of Gamma prior for \gamma_1 (see details).
b_eta  rate parameter of Gamma prior for \gamma_1 (see details).
knots  a sequence of points to define I-splines.
grids  a sequence of points where baseline survival function is to be estimated. Default is minimum observed time points.
niter  total number of iterations of MCMC chains.
seed  a user specified random seed, default is NULL.

Details

The baseline function is modeled by a linear combination of I-splines:
\[ \gamma_0 + \sum_{l=1}^k (\gamma_1 \cdot b_1). \]

Regression coefficient vector \( \beta \) is sampled from a multivariate normal distribution. For more information, please see reference.
Value

A list containing the following elements:

- `parbeta`: A `niter` by `p` matrix of MCMC draws of \( \beta_r \), \( r = 1, ..., p \).
- `parsurv0`: A `niter` by `length(grids)` matrix, each row contains the baseline survival at `grids` from one iteration.
- `parsurv`: A `niter` by `length(grids)*G` matrix, each row contains the survival at `grids` from one iteration. \( G \) is the number of sets of user-specified covariate values.
- `parfinv`: A `niter` by `n` matrix, each row contains the inverse PDF of observed interval-censored data from one iteration. This is used for computing LPML later.
- `grids`: A sequence of points where baseline survival is estimated.

Author(s)

Lianming Wang and Xiaoyan Lin. R version by Bo Cai.

References


description

evaluates estimated regression coefficient(s)

Usage

```r
# S3 method for class 'ICBayes'
coef(object, ...)
```

Arguments

- `object`: Class ICBayes object
- `...`: Other arguments if any

Value

A scalar or vector
Description

Calls the case1ph, case2ph, or case2probit function to fit the corresponding model. Give point estimates and credible intervals for regression coefficients and estimation and plot of survival functions.

Usage

ICBayes(L, ...)

## Default S3 method:
ICBayes(L, R, model, status, xcov, x_user=NULL, order=2,
sig0=10, coef_range=5, v0=0.1, a_eta=1, b_eta=1,
knots=NULL, grids=NULL, conf.int=0.95,
niter=5000, burnin=1000, thin=1, seed=NULL, ...)
## S3 method for class 'formula'
ICBayes(formula, data, ...)

Arguments

- **L** a column vector of left-points of observed time intervals.
- **R** a column vector of right-points of observed time intervals. Use NA to denote infinity.
- **model** a character string specifying the type of model. Possible values are "case1ph", "case2ph", "case2po", and "case2probit".
- **status** a vector of censoring indicators. If model="case1ph", then 1=left-censored, 0=right-censored. If model="case2ph", "case2po", or "case2probit", then 0=left-censored, 1=interval-censored, 2=right-censored.
- **xcov** a matrix of covariates, each column corresponds to one covariate.
- **x_user** a vector of covariate values, default is NULL. Need to specify for survival estimation.
- **order** degree of I-splines (b_l) (see details). Recommended values are 2-4. Default is 2.
- **sig0** standard deviation of normal prior for each regression coefficient beta_r. Used if model="case1ph", "case1po", or "case2ph". Default is 10.
- **coef_range** specify support domain of target density for beta_r using arms (see details). Used if model="case1ph", "case1po", or "case2ph". Default is 5.
- **v0** precision of normal prior for gamma_0. Used if model="case2po" or "case2probit". Default is 0.1.
- **a_eta** shape parameter of Gamma prior for gamma_l (see details). Default is 1.
- **b_eta** rate parameter of Gamma prior for gamma_l (see details). Default is 1.
ICBayes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>knots</td>
<td>A sequence of points to define I-splines. Default is a sequence of time points from min to max with length=10.</td>
</tr>
<tr>
<td>grids</td>
<td>A sequence of points where survival function is to be estimated. Default is a sequence of time points from min to max with length=100.</td>
</tr>
<tr>
<td>conf.int</td>
<td>Level for a two-sided credible interval on coefficient estimate(s). Default is 0.95.</td>
</tr>
<tr>
<td>niter</td>
<td>Total number of iterations of MCMC chains. Default is 11000.</td>
</tr>
<tr>
<td>burnin</td>
<td>Number of iterations to discard at the beginning of an MCMC run. Default is 1000.</td>
</tr>
<tr>
<td>thin</td>
<td>Specify thinning of MCMC draws. Default is 1.</td>
</tr>
<tr>
<td>seed</td>
<td>A use-specified random seed. Default is NULL.</td>
</tr>
<tr>
<td>formula</td>
<td>A symbolic description of the model to be fit.</td>
</tr>
<tr>
<td>data</td>
<td>A data frame containing the variables in the model.</td>
</tr>
<tr>
<td>...</td>
<td>Values passed to other functions.</td>
</tr>
</tbody>
</table>

### Details

For "case1ph", "case1po", and "case2ph" models, function **arms** is used to sample regression coefficients **beta** and **coef.range** specifies the support of the indfunc in **arms**. The baseline cumulative hazard in "case1ph" and "case2ph" models and the baseline odds function in "case1po" are modeled by a linear combination of I-splines:

\[
\sum_{l=1}^{k}(\gamma_l \cdot b_{l})
\]

For "case2probit" model, baseline function is modeled by a linear combination of I-splines:

\[
\gamma_0 + \sum_{l=1}^{k}(\gamma_l \cdot b_{l})
\]

For "case2probit" model, regression coefficient vector **beta** is sampled from a multivariate normal distribution.

For more information, please see reference.

### Value

An object of class ICBayes containing the following elements:

- **coef**: A vector of regression coefficient estimates
- **coef_ssd**: A vector of sample standard deviations of regression coefficient estimates
- **coef_ci**: Credible intervals for regression coefficients
- **LPML**: Log pseudo marginal likelihood for model selection, the larger the better
- **grids**: The sequence of points where baseline survival functions is estimated
- **S0_m**: Estimated baseline survival probabilities at grids
- **S_m**: A length(grids)\*G by 2 matrix that contains estimated survival probabilities at grids for \(x\_user\), where \(G\) is the number of sets of covariate values
- **S_ci**: Credible intervals for survival probabilities at grids for \(x\_user\)
- **mcmc_beta**: A niter by p matrix of mcmc chains for regression coefficients, where niter is the number of iterations and p is the number of covariates
- **mcmc_surv**: A niter by length(grids)\*G matrix of mcmc chains for survival probabilities at grids, where niter is the number of iterations and G is the number of sets of covariate values
Author(s)

Chun Pan

References


See Also

case1ph, case1po, case2ph, case2probit

Examples

# To save time in checking package, niter is set to only 500 iterations.
# formula form
data(bcdata)
bcdata<-data.frame(bcdata)  # must be a data frame
try<-ICBayes(Surv(L,R,type='interval')~x1,data=bcdata,
model='case2ph',status=bcdata[,3],x_user=c(0,1),knots=seq(0.1,60.1,length=10),
grids=seq(0.1,60.1,by=1),niter=500,burnin=100)
# general form
try2<-ICBayes(model='case2ph',L=bcdata[,1],R=bcdata[,2],status=bcdata[,3],
xcov=bcdata[,4],x_user=c(0,1),knots=seq(0.1,60.1,length=10),
grids=seq(0.1,60.1,by=1),niter=500,burnin=100)

logLik.ICBayes

### logLik method for class ICBayes

Description

log-likelihood of the observed interval-censored data estimated by log pseudo marginal likelihood

Usage

```r
## S3 method for class 'ICBayes'
logLik(object, ...)
```
**lungdata**

**Arguments**

- object
- class ICBayes object
- ... other arguments if any

**Value**

- a scalar

---

**lungdata**

* Lung cancer data

**Description**

A case 1 interval-censored data set first presented in Hoel and Walberg (1972) and can be found in Sun (2006, page 6). In the study, 144 male RFM mice were raised under two conditions: conventional environment (96 mice) and germfree environment (48 mice). Each mouse was "sacrificed" at a random time to see if it had lung tumors. Time was measured in days.

**Usage**

data(lungdata)

**Format**

A matrix with 144 rows and 4 columns. Each row (L, R, status, treatment) corresponds to a mouse in the study.

- L left-points of observed intervals
- R right-points of observed intervals
- status censorship indicator: 1=left-censor and 0=right-censor
- treatment treatment indicator: 1=conventional environment, 2=germfree environment

**Source**


**Examples**

data(lungdata)
**plot.ICBayes**  
*Plot baseline survival function*

**Description**
Plot estimated baseline survival function at grids, which are stored in the ICBayes object.

**Usage**
```r
## S3 method for class 'ICBayes'
plot(x, y, ...)
```

**Arguments**
- `x`: a sequence of points where baseline survival probabilities are estimated
- `y`: estimated baseline survival at grids
- `...`: other arguments

**Value**
A plot of baseline survival function.

**Examples**
```r
data(bcdata)
try<-'ICBayes(Surv(L,R,type='interval2')~x1, data= data.frame(bcdata),
model='case2ph', status=bcdatal3], p=1, x_user=c(1),
knots=seq(0.1, 60.1, length=10), grids=seq(0.1, 60.1, by=1), niter=500, burnin=100)
plot.ICBayes(try$grids, try$50_m)
```

---

**summary.ICBayes**  
*summary method for class ICBayes*

**Description**
Present output from function ICBayes

**Usage**
```r
## S3 method for class 'ICBayes'
summary(object, ...)
```

**Arguments**
- `object`: class ICBayes object
- `...`: other arguments if any
SurvtoLR

Value

an object of class ICBayes containing the following elements

- **coef**: a named vector of coefficient estimates
- **coef_ssd**: a named vector of sample standard deviations of coefficient estimates
- **coef_ci**: a named matrix of credible intervals for coefficients
- **LPML**: log pseudo marginal likelihood

---

**Transform Surv object to data matrix with L and R columns**

Description

Take a `Surv` object and transforms it into a data matrix with two columns, L and R, representing the left and right points of observed time intervals. For right-censored data, R = NA.

Usage

```
SurvtoLR(x)
```

Arguments

- **x**: a `Surv` object

Details

The input Surv object should be in the form of `Surv(L, R, type='interval2')`, where R = NA for right-censored data.

Value

A data matrix with two variables:

- **L**: left-points of observed time intervals
- **R**: right-points of observed time intervals

References


Examples

```
library(survival)
L <- c(45, 6, 8, 46)
R <- c(NA, 10, 7, NA)
y <- Surv(L, R, type='interval2')
SurvtoLR(y)
```
Index

*Topic case 1 interval-censored data
ICBayes, 10

*Topic case 2 interval-censored data
ICBayes, 10

*Topic datasets
bcdata, 2
lungdata, 13

arms, 4–7, 10, 11

bcdata, 2

case1ph, 3, 12
case1po, 5, 12
case2ph, 6, 12
case2probit, 8, 12
coeff.ICBayes, 9

ICBayes, 10, 14
ICBayes-package, 2

logLik.ICBayes, 12
lungdata, 13

plot.ICBayes, 14

summary.ICBayes, 14
Surv, 15
SurvtoLR, 15