Package ‘ICBayes’

December 2, 2012

Type Package

Title Bayesian semiparametric models for interval-censored data

Version 1.0

Date 2012-10-27

Author Chun Pan, Bo Cai, Lianming Wang, and Xiaoyan Lin

Maintainer Chun Pan <chunpan2003@hotmail.com>

Description Functions to estimate regression coefficients and baseline survival function for interval-censored data under Bayesian semiparametric models

License GPL(>=2)

LazyLoad yes

Depends HI, survival

R topics documented:

ICBayes-package .................................................. 2
bcdata ............................................................... 2
case1ph ............................................................ 3
case1po ............................................................. 4
case2ph ............................................................ 5
case2po ............................................................. 7
case2probit ........................................................ 8
ICBayes ............................................................ 9
lungdata ........................................................... 12
plot.ICBayes ....................................................... 13
print.ICBayes ..................................................... 13
SurvtoLR .......................................................... 14

Index 15
ICBayes--package  Regression modeling for interval-censored data

Description

This package contains functions to carry out several semiparametric 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 baseline survival curves.

Details

Package: ICBayes
Type: Package
Version: 1.0
Date: 2012-10-27
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

An interval-censored data set presented 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 adjuvent chemotherapy were examined periodically for breast retraction. Endpoints of observed intervals were recorded 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  left points of observed intervals
R  right points of observed intervals
status  censorship indicator: 0=left-censor, 1=interval-censor, and 2=right-censor
x1  treatment indicator: 0=radiotherapy alone, 1=radiotherapy with adjuvent chemotherapy
**case1ph**

**Source**

**Examples**

```r
data(bcdata)
```

---

**Description**
Fit proportional hazards model for case 1 interval-censored data. Use MCMC method to estimate regression coefficients and baseline survival at designated points.

**Usage**

```r
case1ph(L, R, status, xcov, p, order, k, knots_step, knots, sig/zero.noslash, coef_range, a_eta, b_eta, grids_start, grids_end, kgrids, niter)
```

**Arguments**

- **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 censoring indicators: 1=left-censored, 0=right-censored.
- **xcov**: a matrix of covariates, each column corresponds to one covariate.
- **p**: number of covariates
- **order**: degree of I-splines (b_l) (see details). Recommended values are 2-4.
- **k**: number of I-splines. Recommend values are 20-50. Only one of k, knots_step, or knots is needed.
- **knots_step**: increment of knot sequence.
- **knots**: a sequence of points to define I-splines.
- **sig/zero.noslash**: 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).
- **grids_start**: start-point of a sequence of points where baseline survival function is to be estimated. Default is minimum observed time points.
- **grids_end**: end-point of a sequence of points where baseline survival function is to be estimated. Default is maximum observed time points.
- **kgrids**: length of grid sequence.
- **niter**: total number of iterations of MCMC chains.
Details

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

$$\sum_{l=1}^{k}(\gamma_l*b_l).$$

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 matrix of MCMC draws of $\beta_r$, $r=1, \ldots, p$.
- `parsurv0`: A matrix, each row contains the baseline survival at grids from one iteration.
- `grids`: A sequence of points where baseline survival is estimated.

Author(s)

Bo Cai

References


Usage

```r
case1po(L, R, status, xcov, p, order, k, knots_step, knots, sig0, coef_range, a_eta, b_eta, grids_start, grids_end, kgrids, niter)
```

Arguments

- `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 censoring indicators: 1=left-censored, 0=right-censored.
- `xcov`: A matrix of covariates, each column corresponds to one covariate.
- `p`: Number of covariates.
- `order`: Degree of I-splines ($b_l$) (see details). Recommended values are 2-4.
- `k`: Number of I-splines. Recommend values are 20-50. Only one of `k`, `knots_step`, or `knots` is needed.
- `knots_step`: Increment of knot sequence.
- `knots`: A sequence of points to define I-splines.
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 matrix of MCMC draws of $\beta_r$, $r=1, ..., p$.
- `parsurv`: A matrix, each row contains the baseline survival at `grids` from one iteration.
- `grids`: A sequence of points where baseline survival is estimated.

Author(s)

Xiaoyan Lin

References


Description

Fit proportional hazards model for case 2 interval-censored data. Use MCMC method to estimate regression coefficients and baseline survival at designated points.

Usage

```r
case2ph(L, R, status, xcov, p, order, k, knots_step, knots, sig/zero.noslash, coef_range, a_eta, b_eta, grids_start, grids_end, kgrids, niter)
```
Arguments

- **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 censoring indicators: 0=left-censored, 1=interval-censored, 2=right-censored.
- **xcov**: a matrix of covariates, each column corresponds to one covariate.
- **p**: number of covariates
- **order**: degree of I-splines ($b_1$) (see details). Recommended values are 2-4.
- **k**: number of I-splines. Recommend values are 20-50. Only one of $k$, **knots_step**, or **knots** is needed.
- **knots_step**: increment of knot sequence.
- **knots**: a sequence of points to define I-splines.
- **sig/zero.noslash**: 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).
- **grids_start**: start-point of a sequence of points where baseline survival function is to be estimated. Default is minimum observed time points.
- **grids_end**: end-point of a sequence of points where baseline survival function is to be estimated. Default is maximum observed time points.
- **kgrids**: length of grid sequence.
- **niter**: total number of iterations of MCMC chains.

Details

The baseline cumulative hazard is modeled 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 matrix of MCMC draws of $\beta_r, r=1, \ldots, p$.
- **parsurv/zero.noslash**: a matrix, each row contains the baseline survival at **grids** from one iteration.
- **grids**: a sequence of points where baseline survival is estimated.

Author(s)

Bo Cai

References

case2po

**PO model for case 2 interval-censored data**

**Description**

Fit proportional odds model for case 2 interval-censored data. Use MCMC method to estimate regression coefficients and baseline survival at designated points.

**Usage**

```r
case2po(method, L, R, status, xcov, p, order, k, knots_step, knots, v/zero.noslash, a_eta, b_eta, grids_start, grids_end, kgrids, niter)
```

**Arguments**

- `method`: a character string specifying the data augmentation method. Possible values: "PO1", "PO2" (see details).
- `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 censoring indicators: 0=left-censored, 1=interval-censored, 2=right-censored.
- `xcov`: a matrix of covariates, each column corresponds to one covariate.
- `p`: number of covariates.
- `order`: degree of I-splines \( b_l \) (see details). Recommended values are 2-4.
- `k`: number of I-splines. Recommend values are 20-50. Only one of \( k \), `knots_step`, or `knots` need to be specified.
- `knots_step`: increment of knot sequence.
- `knots`: a sequence of points to define I-splines.
- `v0`: precision of normal prior for \( \gamma_0 \).
- `a_eta`: shape parameter of Gamma prior for \( \gamma_0 \) (see details).
- `b_eta`: rate parameter of Gamma prior for \( \gamma_0 \) (see details).
- `grids_start`: start-point of a sequence of points where baseline survival function is to be estimated. Default is minimum observed time points.
- `grids_end`: end-point of a sequence of points where baseline survival function is to be estimated. Default is maximum observed time points.
- `kgrids`: length the grid sequence.
- `niter`: total number of iterations of MCMC chains.

**Details**

Two data augmentation methods are used (referred to as 'PO1' and 'PO2'). The first method leads to the exact PO model, while the second leads to an approximate PO model but all full conditional distributions are of standard forms. The baseline log odds function is modeled by a linear combination of I-splines:

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

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 matrix of MCMC draws of $\beta_r$, $r=1, ..., p$.
par surv0 a matrix, each row contains the baseline survival at grids from one iteration.
grids a sequence of points where baseline survival is estimated.

Author(s)

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

References


case2probit Probit model for case 2 interval-censored data

Description

Fit probit model to case 2 interval-censored data. Use MCMC method to estimate regression coefficients and baseline survival at designated points.

Usage

case2probit(L, R, status, xcov, p, order, k, knots_step, knots, v0, a_eta, b_eta, grids_start, grids_end, kgrids, niter)

Arguments

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 censoring indicators: 0=left-censored, 1=interval-censored, 2=right-censored.
xcov a matrix of covariates, each column corresponds to one covariate.
p number of covariates.
order degree of I-splines ($b_l$) (see details). Recommended values are 2-4.
k number of I-splines. Recommend values are 20-50. Only one of $k$, knots_step, or knots need to be specified.
knots_step increment of the knot sequence.
knots a sequence of points to define I-splines.
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).
grids_start start-point of a sequence of points where baseline survival function is to be estimated. Default is minimum observed time points.
grid_end  end-point of a sequence of points where baseline survival function is to be estimated. Default is maximum observed time points.

kgrids  length grid sequence.
niter  total number of iterations of MCMC chains.

Details

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

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 matrix of MCMC draws of \( \beta_r \), \( r=1, \ldots, p \).
- parsurv  a matrix, each row contains the baseline survival at \( \text{grid} \) from one iteration.
- grids  a sequence of points where baseline survival is estimated.

Author(s)

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

References


### Description

Calls the `case1ph`, `case2ph`, `case2po`, or `case2probit` function to fit the corresponding model. Give point estimates and credible intervals for regression coefficients and estimation and plot of baseline survival function.

### Usage

```r
ICBayes(L, ...)
```

```r
## Default S3 method:
ICBayes(model, method, L, R, status, xcov, p, order = 2, k = 20,
knots_step = NULL, knots = NULL, sig0 = 10, coef_range = 5, v0 = 0.1, a_eta = 1, b_eta = 1,
grids_start = NULL, grids_end = NULL, kgrids = NULL, niter = 11000, burnin = 1000, thin = 1,
conf.int = 0.95, plot_S = TRUE, chain.save = FALSE, dd1)
## S3 method for class 'formula'
ICBayes(formula, data, ...)
```
Arguments

model  a character string specifying the type of model. Possible values are "case1ph", "case2ph", "case2po", and "case2probit".
method  a character string specifying the data augmentation method. Used if model="case2po". Possible values: "PO1", "PO2" (see details).
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.
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, 2=right-censored.
xcov  a matrix of covariates, each column corresponds to one covariate.
p  number of covariates.
order  degree of I-splines (b_1) (see details). Recommended values are 2-4. Default is 2.
k  number of I-splines (default is 20). Recommend values are 20-50. Default is 20. Only one of k, knots_step, or knots need to be specified.
knots_step  increment of knot sequence. Default is NULL.
knots  a sequence of points to define I-splines. Default is NULL.
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".
v0  precision of normal prior for gamma_l. 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.
grids_start  start-point of a sequence of points where baseline survival function is to be estimated. Default is minimum observed time points.
grids_end  end-point of a sequence of points where baseline survival function is to be estimated. Default is maximum observed time points.
kgrids  length the grid sequence. Default is 100.
niter  total number of iterations of MCMC chains. Default is 11000.
burnin  number of iterations to discard at the beginning of an MCMC run. Default is 1000.
thin  specify thinning of MCMC draws. Default is 1.
conf.int  level for a two-sided credible interval on coefficient estimate(s). Default is 0.95.
plot_S  logical; if TRUE (default), baseline survival at grids are saved for plot.
chain.save  logical; if TRUE, MCMC chains for beta_r’s are saved in dd1.
formula  a symbolic description of the model to be fit.
data  a data frame containing the variables in the model.
...  values passed to other functions.
Details

For "case1ph", "case1po", and "case2ph" models, function arms is used to sample regression coefficient beta_r, 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:

\[ \text{sum}_{l=1}^{k}(\gamma_l \cdot b_l) \]

For "case2po" model, two data augmentation methods are used (referred to as 'PO1' and 'Po2'). The first method leads to the exact PO model, while the second leads to an approximation to the PO model. Baseline log odds function is modeled by a linear combination of I-splines:

\[ \gamma_0 + \text{sum}_{l=1}^{k}(\gamma_l \cdot b_l) \]

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

\[ \gamma_0 + \text{sum}_{l=1}^{k}(\gamma_l \cdot b_l) \]

For "case2po" and "case2probit" models, regression coefficient vector beta is sampled from a multivariate normal distribution.

For more information, please see reference.

Value

an object of class intervmod2 containing at least the following elements:

- coef: a vector of regression coefficient estimates
- coef_se: a vector of standard errors of regression coefficient estimates
- coef_ci: credible intervals for regression coefficients

If plot_S is TRUE, also store:

- grids: the sequence of points where baseline survival functions is estimated
- S_m: estimated baseline survival probabilities at grids

Author(s)

Chun Pan

References


See Also

case1ph, case1po, case2ph, case2po, case2probit
Examples

```r
# formula form
library(survival)
data(bcdata)
bcdata<-data.frame(bcdata) # must be a data frame
try<-ICBayes(Surv(L,R,type='interval2')~x1,data=bcdata,
model='case2ph',status=bcdata[,3],p=1)
# general form
try2<-ICBayes(model='case2ph',L=bcdata[,1],R=bcdata[,2],status=bcdata[,3],
xcov=bcdata[,4],p=1)
```

---

**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

```r
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

```r
data(lungdata)
```
plot.ICBayes

Plot baseline survival function

Description
Plot estimated baseline survival probabilities at specified grids, which are stored in ICBayes object. Given baseline survival at grids, then can calculate survival for a specific covariate vector at grids, hence plot survival curves for different treatment groups, for instance.

Usage
```r
## S3 method for class 'ICBayes'
plot(grids, S/zero.m)
```

Arguments
- `grids`: a sequence of points where baseline survival probabilities are estimated
- `S/zero.m`: estimated baseline survival at grids

Value
A plot of baseline survival function.

Examples
```r
data(bcdata)
try<-ICBayes(model='case2ph',L=bcdata[,1],R=bcdata[,2],status=bcdata[,3],
            xcov=bcdata[,4],p=1,coef_range=2,chain.save=FALSE)
plot.ICBayes(try$grids,try$S/zero.m)
```

print.ICBayes

print method for ICBayes object

Description
Print output from function ICBayes

Usage
```r
## S3 method for class 'ICBayes'
print(object)
```

Arguments
- `object`: an object of class "ICBayes", i.e., a fitted model
SurvtoLR

Value
an object of class ICBayes containing the following elements

- `coef`: a named vector of coefficient estimates
- `coef_se`: a named vector of standard errors of coefficient estimates
- `coef_ci`: a named matrix of credible intervals for coefficients

SurvtoLR

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

```r
library(survival)
L<-c(45, 6, 0, 46)
R<-c(NA, 10, 7, NA)
y<-Surv(L, R, type='interval2')
SurvtoLR(y)
```
Index

*Topic case 1 interval-censored data
  ICBayes, 9
*Topic case 2 interval-censored data
  ICBayes, 9
*Topic datasets
  bcdata, 2
  lungdata, 12

arms, 3–6, 10, 11

bcdata, 2

  case1ph, 3, 11
  case1po, 4, 11
  case2ph, 5, 11
  case2po, 7, 11
  case2probit, 8, 11

ICBayes, 9, 13
ICBayes-package, 2

lungdata, 12

plot.ICBayes, 13
print.ICBayes, 13

Surv, 14
SurvtoLR, 14