Package ‘ICsurv’

February 19, 2015

Type Package

Title A package for semiparametric regression analysis of interval-censored data

Version 1.0

Date 2014-6-9

Author Christopher S. McMahan and Lianming Wang

Maintainer Lianming Wang <wangl@stat.sc.edu>

Description Currently using the proportional hazards (PH) model. More methods under other semi-parametric regression models will be included in later versions.

License GPL (>= 2)

Imports MASS (>= 7.3-33), matrixcalc (>= 1.0-3)

NeedsCompilation no

Repository CRAN

Date/Publication 2014-06-10 18:22:44

R topics documented:

- ICsurv-package ......................................................... 2
- fast.PH.ICsurv.EM ...................................................... 2
- fast.PH.Louis.ICsurv ................................................... 4
- Hemophilia ............................................................... 5
- Ispline ................................................................. 6
- PH.ICsurv.EM ............................................................. 6
- PH.Louis.ICsurv ......................................................... 8

Index 10
ICsurv--package

A statistical package for regression analysis of interval-censored data under the semiparametric proportional hazards (PH) model

Description

This package allows for semiparametric regression analysis of interval-censored data under the proportional hazards model (PH). Monotone splines are used to estimate the unknown nondecreasing cumulative baseline hazard function. Model fitting is completed through the implementation of an expectation maximization (EM) algorithm.

Details

Package: ICsurv
Type: Package
Version: 1.0
Date: 2013-10-18

Author(s)

Christopher S. McMahan and Lianming Wang

Maintainers: Christopher S. McMahan <mcmaha2@clemson.edu> and Lianming Wang <wangl@stat.sc.edu>

fast.PH.ICsurv.EM

EM algorithm for general interval-censored data under the proportional hazards model

Description

This function is equivalent to the function "PH.ICsurv.EM" but runs much faster.

Usage

fast.PH.ICsurv.EM(d1, d2, d3, Li, Ri, Xp, n.int, order, g0, b0, tol, t.seq, equal = TRUE)

Arguments

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.

Xp  design matrix of predictor variables (in columns), should be specified without an intercept term.

n.int the number of interior knots to be used.

order the order of the basis functions.

g0  initial estimate of the spline coefficients; should be of length n.int+order.

b0  initial estimate of regression coefficients; should be of length dim(Xp)[2].

tol the convergence criterion of the EM algorithm, see details for further description.

t.seq an increasing sequence of points at which the cumulative baseline hazard function is evaluated.

equal logical, if TRUE knots are spaced evenly across the range of the endpoints of the observed intervals and if FALSE knots are placed at quantiles.

Details

The above function fits the semiparametric proportional hazards model (PH), proposed in Wang et al. (2014+), to interval censored data via an EM algorithm. For a discussion of starting values, number of interior knots, order, and further details please see Wang et al. (2014+). The EM algorithm converges when the maximum of the absolute difference in the parameter estimates (to include the regression and spline coefficients) is less than tol. The Hessian matrix of the observed likelihood is given in the output. The variance-covariance matrix of the estimated regression and spline coefficients can be obtained by taking the inverse of the Hessian matrix. When the Hessian matrix is singular, the variance matrix of the regression parameters is obtained by using the inverse of blocked matrix, which only involves taking inverse of lower dimensional matrices. To further provide robustness, the generalized inverse function "ginv" in the supporting package "MASS" is used in this case. A function in the supporting R package "matrixcalc" is used to check whether the Hessian matrix is singular.

Value

b  estimates of the regression coefficients.

g  estimates of the spline coefficients.

hz  estimated cumulative baseline hazard function evaluated at the points t.seq.

Hessian the Hessian matrix. Its inverse is the variance covariance matrix of b and g.

var.b the variance covariance matrix of b

flag  the indicator whether the Hessian matrix is non-singular. When flag=0, the variance estimate may not be accurate.

AIC  the Akaike information criterion.

BIC  the Bayesian information/Schwarz criterion.

ll  the value of the maximized log-likelihood.
References


---

**fast.PH.Louis.ICsurv**  
*Calculating the Hessian matrix using Louis’s Method (1982)*

### Description

This function is a fast version of **PH.Louis.ICsurv**. It calculates the negative of the Hessian of the log of the observed data likelihood, obtained via Louis’s method, evaluated at the last step of the EM algorithm described in Wang et al. (2014+). This is a support function for **fast.PH.ICsurv.EM**.

### Usage

```r
fast.PH.Louis.ICsurv(b, g, bli, bri, d1, d2, d3, xp)
```

### Arguments

- `b`  
  estimates of the regression coefficients obtained at the convergence of the EM algorithm.

- `g`  
  estimates of the spline coefficients obtained from at the convergence of the EM algorithm.

- `bli`  
  an I-spline basis matrix of dimension c(length(knots)+order-2, length(x)), corresponding to the left end points of the observed intervals.

- `bri`  
  an I-spline basis matrix of dimension c(length(knots)+order-2, length(x),), corresponding to the left end points of the observed intervals.

- `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).

- `xp`  
  design matrix of predictor variables (in columns), should be specified without an intercept term.

### Details

To obtain the Hessian matrix of the observed likelihood evaluated at the last step of the EM algorithm.

### Value

Hessian matrix.
Description

This data were collected as a part of a multi-center prospective study aimed at assessing the HIV-1 infection rate among people with hemophilia. In particular, the individuals in the study were at risk from contracting HIV-1 from blood products made from donor’s plasma. In this study, 544 patients were categorized into one of four groups based on their average annual dose of blood products. Specifically, patients were classified into high, medium, low, or no dose group. The goal of the study was to compare the HIV-1 infection rate between these dose groups. The time at which patients contracted HIV-1 is not known exactly, but were known to occur before, between, or after certain observation times. For more details pertaining to this data set see Goedert et al. (1989) and Kroner et al. (1994).

Usage

data(Hemophilia)

Format

A data frame with 544 observations on the following 8 variables.

d1 Censoring indicator, 1 if patient’s infection time was left censored, 0 otherwise.
d2 Censoring indicator, 1 if patient’s infection time was interval censored, 0 otherwise.
d3 Censoring indicator, 1 if patient’s infection time was right censored and, otherwise.
l Left endpoint of the observation interval
r Right endpoint of the observation interval
low Binary variable indicating the patient was a member of the low dose group.
medium Binary variable indicating the patient was a member of the medium dose group.
high Binary variable indicating the patient was a member of the high dose group.

References


**Ispline**

**Description**
Generates the I-spline basis matrix associated with integrated spline basis functions. Created by Cai and Wang in October, 2009.

**Usage**
Ispline(x, order, knots)

**Arguments**
- x: the predictor variables.
- order: the order of the basis functions.
- knots: a sequence of increasing points specifying the placement of the knots.

**Value**
An I-spline basis matrix of dimension c(length(knots)+order-2, length(x)).

**References**

**PH.ICsurv.EM**

**Description**
Fits the semiparametric proportional hazards model (PH), proposed in Wang et al. (2014+), to interval censored data via an EM algorithm.

**Usage**
PH.ICsurv.EM(d1, d2, d3, Li, Ri, Xp, n.int, order, g0, b0, tol, t.seq, equal = TRUE)
Arguments

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.
Xp design matrix of predictor variables (in columns), should be specified without an intercept term.
n.int the number of interior knots to be used.
order the order of the basis functions.
g0 initial estimate of the spline coefficients; should be of length n.int+order.
b0 initial estimate of regression coefficients; should be of length dim(Xp)[2].
tol the convergence criterion of the EM algorithm, see details for further description.
t.seq an increasing sequence of points at which the cumulative baseline hazard function is evaluated.
equal logical, if TRUE knots are spaced evenly across the range of the endpoints of the observed intervals and if FALSE knots are placed at quantiles.

Details

The above function fits the semiparametric proportional hazards model (PH), proposed in Wang et al. (2014+), to interval censored data via an EM algorithm. For a discussion of starting values, number of interior knots, order, and further details please see Wang et al. (2014+). The EM algorithm converges when the maximum of the absolute difference in the parameter estimates (to include the regression and spline coefficients) is less than tol. The Hessian matrix of the observed likelihood is given in the output. The variance-covariance matrix of the estimated regression and spline coefficients can be obtained by taking the inverse of the Hessian matrix. When the Hessian matrix is singular, the variance matrix of the regression parameters is obtained by using the inverse of blocked matrix, which only involves taking inverse of lower dimensional matrices. To further provide robustness, the generalized inverse function "ginv" in the supporting package "MASS" is used in this case. A function in the supporting R package "matrixcalc" is used to check whether the Hessian matrix is singular.

Value

b estimates of the regression coefficients.
g estimates of the spline coefficients.
hz estimated cumulative baseline hazard function evaluated at the points t.seq.
Hessian the Hessian matrix. Its inverse is the variance covariance matrix of b and g.
var.b the variance covariance matrix of b
flag
the indicator whether the Hessian matrix is non-singular. When flag=0, the variance estimate may not be accurate.

AIC
the Akaike information criterion.

BIC
the Bayesian information/Schwarz criterion.

ll
the value of the maximized log-likelihood.

References

Examples
data(Hemophilia)
d1<-Hemophilia[,1]
d2<-Hemophilia[,2]
d3<-Hemophilia[,3]
Li<-Hemophilia[,4]
Ri<-Hemophilia[,5]
Xp<-as.matrix(Hemophilia[,c(6,7,8)])

fit <- PH.ICsurv.EM(d1, d2, d3, Li, Ri, Xp, n.int=8, order=3, g0=rep(1,11), b0=rep(0,3), tol=0.001, t.seq=seq(0,0.5,1), equal = TRUE)

fit$b
# [1] 1.837590 3.018500 3.418981

fit$var.b

# [1,] 0.008526765 -0.010905780 0.01199610
# [2,] -0.010905780 0.012659520 -0.01462116
# [3,] 0.009960950 -0.014621160 0.06824411

**Description**
Calculates the negative of the Hessian of the log of the observed data likelihood, obtained via Louis’s method, evaluated at the last step of the EM algorithm described in Wang et al. (2014+). This is a support function for PH.ICsurv.EM.

**Usage**
PH.Louis.ICsurv(b, g, bli, bRi, d1, d2, d3, Xp)
Arguments

b  estimates of the regression coefficients obtained at the convergence of the EM algorithm.

g  estimates of the spline coefficients obtained from at the convergence of the EM algorithm.

bLi  an I-spline basis matrix of dimension c(length(knots)+order-2, length(x)), corresponding to the left end points of the observed intervals.

bRi an I-spline basis matrix of dimension c(length(knots)+order-2, length(x)), corresponding to the left end points of the observed intervals.

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).

xp design matrix of predictor variables (in columns), should be specified without an intercept term.

Details

To obtain the Hessian matrix of the observed likelihood evaluated at the last step of the EM algorithm.

Value

Hessian matrix.

References


Index

*Topic datasets
  Hemophilia, 5
  fast.PH.ICsurv.EM, 2, 4
  fast.PH.Louis.ICsurv, 4

Hemophilia, 5

ICsurv-package, 2
Ispline, 6

PH.ICsurv.EM, 6, 8
PH.Louis.ICsurv, 4, 8