Package 'PCDSpline'

July 2, 2014

Type Package

Title Semiparametric regression analysis of panel count data using monotone splines

Version 1.0

Date 2014-06-13

Author Bin Yao <yaob@email.sc.edu> and Lianming Wang <wangl@stat.sc.edu>

Maintainer Bin Yao <yaob@email.sc.edu>

Description

Semiparametric regression analysis of panel count data under the non-homogeneous Poisson process model with and without Gamma frailty using monotone splines.

License GPL (>= 2)

Imports nleqslv (>= 2.2), matrixcalc (>= 1.0-3)

NeedsCompilation no

Repository CRAN

Date/Publication 2014-06-13 18:18:27

R topics documented:

PCDSpline-package *A statistical package for regression analysis of panel count data under the non-homogeneous Poisson process models with and without frailty*

Description

This package allows for semiparametric regression analysis of panel count data under the non-homogeneous Poisson process models with and without frailty. The Gamma frailty model allows to account for the within-subject correlation. Monotone splines of Ramsay (1988) are used to estimate the unknown baseline mean function. Fitting the models to panel count data via EM algorithm.

Details

Package:	PCDSpline
Type:	Package
Version:	1.0
Date:	2014-06-13
License:	GPL (>= 2)

Author(s)

Bin Yao <yaob@email.sc.edu> and Lianming Wang <wangl@stat.sc.edu>

BladTumor *Bladder Tumor Cancer Data*

Description

Bladder tumor data were from the bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group, and used by many people to demonstrate methodology for recurrent event modelling. In this study 118 patients who had superficial bladder tumors were randomized into one of three treatment groups: placebo (48), thiotepa (38), and pyridoxine (32). During the study at each follow-up visit, new tumors since the last visit were counted, measured, and removed transurethrally. For each patient the initial number of tumors and the size of largest initial tumors were also recorded. For more details about this study see Byar et al. (1977).

Usage

```
data(BladTumor)
```

Format

A data frame with 116 observations on the following 8 variables.

subject:	patient ID
time:	observation time
count:	cumulative number of tumors
number:	initial number of tumors (8=8 or more)
size:	size(cm) of largest initial tumors
pyridoxine:	dummy variable for pyridoxine treatment
thiotepa:	dummy variable for thiotepa treatment
count1:	number of new tumors since last observation time

Details

This data include 116 subjects who have at least one follow-up observation after the study enrollment.

Note

To further use all the functions of this package one must convert the original data structure into the specified data structre which is in a list form. For more details please see the following example using bladder tumor data.

Source

Wang, X. and Yan, J. (2011). Fitting semiparametric regressions for panel count survival data with an R package spef. *Computer Methods and Programs in Biomedicine* 104,2 278-285

References

Byar, D.P., Blackard,C., and the VACURG. (1977). Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage I bladder cancer. *Urology* 10, 556-561.

See Also

BladTumor1

Examples

```
data(BladTumor)

n<-max(BladTumor$subject)
#record the number of observations for all patients
k<-as.numeric(table(BladTumor$subject))
K<-max(k)
t<-matrix(,n,K)
z<-matrix(,n,K)

x1<-c();x2<-c();x3<-c();x4<-c();

for (r in 1:n){
rownum<-which(BladTumor$subject==r)
#record all observation times
```

```
t[r,][1:k[r]]<-BladTumor[rownum,]$time
#record all panel counts from non-overlapping intervals
z[r,][1:k[r]]<-BladTumor[rownum,]$count1
x1[r]<-BladTumor[which(BladTumor$subject==r),]$number[1]
x2[r]<-BladTumor[which(BladTumor$subject==r),]$size[1]
x3[r]<-BladTumor[which(BladTumor$subject==r),]$pyridoxine[1];
x4[r]<-BladTumor[which(BladTumor$subject==r),]$thiotepa[1]
}

x<-cbind(x1,x2,x3,x4)
BladTumor1<-list(t=t,x=x,z=z,k=k,K=K)
```

`BladTumor1` *Converted Bladder Tumor Cancer Data*

Description

This data set is converted from `BladTumor` to have the format to be used in the function. One should convert the original data into this specified data structure and do the further analysis.

Usage

```
data(BladTumor1)
```

Format

t:	observation times matrix
x:	covariate matrix
z:	panel counts from non-overlapping intervals matrix
k:	the number of observations from all patients
K:	the largest number of observations

See Also

`BladTumor`

`Ispline` *Ispline*

Description

Generates the I-spline basis matrix associated with integrated spline basis functions. Created by Cai and Wang in October, 2009. Details can be found in Ramsay (1988).

Usage

```
Ispline(x, order, knots)
```

Arguments

x	a vector of observations.
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

Ramsay, J. (1988). Monotone regression splines in action. *Statistical Science* 3, 425-441.

PCD.Corr — *Pearson's coefficient of correlation between panel counts within two non-overlapping intervals*

Description

Calculating Pearson's coefficient of correlation between panel counts within two non-overlapping time intervals (t1, t2] and (t3, t4]. The arguments of this function are the output of function PCDReg.wf. For the corresponding formula see Yao, Wang and He (2014+).

Usage

```
PCD.Corr(x, beta, nu, gamma, t1, t2, t3, t4, order, knots)
```

Arguments

x	the covariate vector.
beta	estimates of regression coefficients.
nu	estimate of gamma frailty variance parameter.
gamma	estimates of spline coefficients.
t1, t2	interval endpoints. t1 must be less than t2.
t3, t4	interval endpoints. t3 must be less than t4.
order	the order of basis functions.
knots	knots used in the analysis.

Value

corr.rho	Pearson's coefficient of correlation.

Note

The two intervals (t1, t2] and (t3, t4] must not be overlapped.

References

Yao, B., Wang, L., and He, X. (2014+). Semiparametric regression analysis of panel count data allowing for within-subject correlation.

See Also

PCDReg.wf

Examples

```
##Simulated Data

n=13; #number of subjects

##generate the number of observations for each subject
k=rpois(n,6)+1; K=max(k);

##generate random time gaps for each subject
y=matrix(,n,K);
for (i in 1:n){y[i,1:k[i]]=rexp(k[i],1)}

##get observation time points for each subject
t=matrix(,n,K);
for (i in 1:n){
  for (j in 2:K){
    t[i,1] = y[i,1]
    t[i,j] = y[i,j]+t[i,j-1]
  }
}

##covariate x1 and x2 generated from Normal(0,0.5^2) and Bernoulli(0.5) respectively
x1=rnorm(n,0,0.5); x2=rbinom(n,1,0.5); x=cbind(x1,x2)

##true regression parameters and frailty variance parameter
beta1=1; beta2=-1; nu=0.5;
parms=c(beta1,beta2)
phi=rgamma(n,nu,nu)

##true baseline mean function
mu=function(t){2*t^(0.5)}

##get the number of events between time intervals
z=matrix(,n,K);
xparms=c();for (s in 1:nrow(x)){xparms[s]<-sum(x[s,]*parms)}
for (i in 1:n){
 z[i,1]<-rpois(1,mu(t[i,1])*exp(xparms[i])*phi[i])
 if (k[i]>1){
 z[i,2:k[i]]<-rpois(k[i]-1,(mu(t[i,2:k[i]])-mu(t[i,1:(k[i]-1)]))*exp(xparms[i])*phi[i])
```

```
 }
}

TestD<-list(t=t, x=x, z=z, k=k, K=K)

fit<-PCDReg.wf(DATA = TestD, order = 1, placement = TRUE, nknot = 3, myknots,
               binit = c(0.5,-0.5), ninit = 0.1, ginit = seq(0.1,2),
               t.seq = seq(0,15,0.2), tol = 10^(-3))

x1=c(1,1);
b1=fit$beta; n1=fit$nu; g1=fit$gamma;
t1=0; t2=6; t3=6; t4=12;
order=1; knots=fit$knots;

PCD.Corr(x1, b1, n1, g1, t1, t2, t3, t4, order, knots)
```

`PCDHess.wf` *Calculating the Hessian matrix using Louis's method (1982)*

Description

Calculating the Hessian matrix using Louis's Method under the Gamma frailty non-homogeneous Poisson process model. This is a support function for `PCDReg.wf`.

Usage

```
PCDHess.wf(DATA, beta, gamma, nu, order, knots)
```

Arguments

`DATA`	use specified data structure.
`beta`	estimates of regression coefficients.
`gamma`	estimates of spline coefficients.
`nu`	estimates of gamma frailty variance parameter.
`order`	the order of basis functions.
`knots`	the equally spaced knots.

Details

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

Value

`HESS`	Hessian matrix.

References

Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society*, Series B 44, 226-233.

Yao,B., Wang, L., and He, X. (2014+).Semiparametric regression analysis of panel count data allowing for within-subject correlation.

`PCDReg.nf` *Regression analysis of panel count data under the non-homogeneous Poisson process model*

Description

Fits the nonhomogeneous Poisson process model to panel count data using EM algorithm.

Usage

```
PCDReg.nf(DATA, order, placement, nknot, myknots, binit, ginit, t.seq, tol)
```

Arguments

`DATA`	use specified data structure.
`order`	the order of basis functions.
`placement`	logical, if TRUE knots are placed evenly across the observed intervals based on the input data set; if FALSE knots should be specified by the user. see myknots.
`nknot`	the number of knots to be used.
`myknots`	a sequence of increasing points whose length is nknot.
`binit`	initial estimate of regression coefficients.
`ginit`	initial estimate of spline coefficients whose length should be (order+nknot-2) or (order+length(myknots)-2).
`t.seq`	an increasing sequence of points at which the baseline mean function is evaluated.
`tol`	the convergence criterion of the EM algorithm.

Details

The above function fits the non-homogeneous Poisson process model to panel count data via EM algorithm.

Value

beta	estimates of regression coefficients.
gamma	estimates of spline coefficients.
var.b	the variance covariance matrix of regression coefficients.
Hess	Hessian matrix.
knots	the knots used in the analysis.
bmf	estimated baseline mean function evaluated at the points t.seq; use pmf to plot the baseline mean fuction.
AIC	the Akaike information criterion.
BIC	the Bayesian information/Schwarz criterion.
flag	the indicator whether the Hessian matrix is non-singular. When flag="TRUE",the variance estimate may not be accurate.

Note

The non-homogeneous Poisson process model involves no Gamma frailty.

References

Yao, B., Wang, L., and He, X. (2014+). Semiparametric regression analysis of panel count data allowing for within-subject correlation.

See Also

PCDReg.wf

Examples

```
##Simulated Data

n=13; #number of subjects

##generate the number of observations for each subject
k=rpois(n,6)+1; K=max(k);

##generate random time gaps for each subject
y=matrix(,n,K);
for (i in 1:n){y[i,1:k[i]]=rexp(k[i],1)}

##get observation time points for each subject
t=matrix(,n,K);
for (i in 1:n){
  for (j in 2:K){
    t[i,1] = y[i,1]
    t[i,j] = y[i,j]+t[i,j-1]
  }
}
```

```
##covariate x1 and x2 generated from Normal(0,0.5^2) and Bernoulli(0.5) respectively
x1=rnorm(n,0,0.5); x2=rbinom(n,1,0.5); x=cbind(x1,x2)

##true regression parameters and frailty variance parameter
beta1=1; beta2=-1; nu=0.5;
parms=c(beta1,beta2)
phi=rgamma(n,nu,nu)

##true baseline mean function
mu=function(t){2*t^(0.5)}

##get the number of events between time intervals
z=matrix(,n,K);
xparms<-c();for (s in 1:nrow(x)){xparms[s]<-sum(x[s,]*parms)}
for (i in 1:n){
 z[i,1]<-rpois(1,mu(t[i,1])*exp(xparms[i])*phi[i])
 if (k[i]>1){
 z[i,2:k[i]]<-rpois(k[i]-1,(mu(t[i,2:k[i]])-mu(t[i,1:(k[i]-1)]))*exp(xparms[i])*phi[i])
 }
}

TestD<-list(t=t, x=x, z=z, k=k, K=K)

fit<-PCDReg.nf(DATA = TestD, order = 1, placement = TRUE, nknot=3,
               myknots, binit = c(-0.5,0.5) , ginit = seq(0.1,2),
               t.seq = seq(0,15,0.2), tol=10^(-3))
```

PCDReg.wf *Regression analysis of panel count data under the Gamma frailty non-homogeneous Poisson process model*

Description

Fits the Gamma frailty non-homogeneous Poisson process model to panel count data using EM algorithm.

Usage

```
PCDReg.wf(DATA, order, placement, nknot, myknots, binit, ninit, ginit, t.seq, tol)
```

Arguments

DATA	use specified data structure.
order	the order of basis functions.
placement	logical, if TRUE knots are placed evenly across the observed intervals based on the input data set; if FALSE knots should be specified by the user. see myknots.
nknot	the number of knots to be used.
myknots	knots specified by the user whose length is nknot.

binit	initial estimate of regression coefficients.
ninit	initial estimate of gamma frailty variance parameter.
ginit	initial estimate of spline coefficients whose length should be (order+nknot-2) or (order+length(myknots)-2).
t.seq	an increasing sequence of points at which the baseline mean function is evaluated.
tol	the convergence criterion of the EM algorithm.

Details

The above function fits the Gamma frailty non-homogeneous Poisson process model to panel count data via EM algorithm. To use this function, the data must have the same structure as in BladTumor1. For a discussion of order, number of interior knots and further details please see Yao et al. (2014+). The EM algorithm converges when the maximum of the absolute difference in the parameter estimates is less than tol.

Value

beta	estimates of regression coefficients.
nu	estimates of gamma frailty variance parameter.
gamma	estimates of spline coefficients.
var.bn	the variance covariance matrix of regression coefficients estimates and gamma frailty variance parameter estimates.
knots	the knots used in the analysis; equally spaced knots or knots specified by the user.
bmf	estimated baseline mean function evaluated at the points t.seq; use pmf to plot the baseline mean function.
AIC	the Akaike information criterion.
BIC	the Bayesian information/Schwarz criterion.
flag	the indicator whether the Hessian matrix is non-singular. When flag="TRUE",the variance estimate may not be accurate.

Note

Use specified data structure.

References

Yao, B., Wang, L., and He, X. (2014+). Semiparametric regression analysis of panel count data allowing for within-subject correlation.

See Also

PCDHess.wf

Examples

```
##Simulated Data

n=13; #the number of subjects

##generate the number of observations for each subject
k=rpois(n,6)+1; K=max(k);

##generate random time gaps for each subject
y=matrix(,n,K);
for (i in 1:n){y[i,1:k[i]]=rexp(k[i],1)}

##get observation time points for each subject
t=matrix(,n,K);
for (i in 1:n){
  for (j in 2:K){
    t[i,1] = y[i,1]
    t[i,j] = y[i,j]+t[i,j-1]
  }
}

##covariate x1 and x2 generated from Normal(0,0.5^2) and Bernoulli(0.5) respectively
x1=rnorm(n,0,0.5); x2=rbinom(n,1,0.5); x=cbind(x1,x2)

##true regression parameters and frailty variance parameter
beta1=1; beta2=-1; nu=0.5;
parms=c(beta1,beta2)
phi=rgamma(n,nu,nu)

##true baseline mean function
mu=function(t){2*t^(0.5)}

##get the number of events between time intervals
z=matrix(,n,K);
xparms=c();for (s in 1:nrow(x)){xparms[s]=sum(x[s,]*parms)}
for (i in 1:n){
 z[i,1]=rpois(1,mu(t[i,1])*exp(xparms[i])*phi[i])
 if (k[i]>1){
 z[i,2:k[i]]=rpois(k[i]-1,(mu(t[i,2:k[i]])-mu(t[i,1:(k[i]-1)]))*exp(xparms[i])*phi[i])
 }
}

TestD<-list(t=t, x=x, z=z, k=k, K=K)

fit<-PCDReg.wf(DATA = TestD, order = 1, placement = TRUE, nknot = 3, myknots,
               binit = c(0.5,-0.5), ninit = 0.1, ginit = seq(0.1,2),
               t.seq = seq(0,15,0.2), tol = 10^(-3))
```

Index