Package ‘SurvLong’

January 10, 2020

Type Package

Title Analysis of Proportional Hazards Model with Sparse Longitudinal Covariates

Version 1.1

Date 2020-01-08

Author Hongyuan Cao, Mathew M. Churpek, Donglin Zeng, Jason P. Fine, and Shannon T. Holloway

Maintainer Shannon T. Holloway <sthollow@ncsu.edu>

Description Provides kernel weighting methods for estimation of proportional hazards models with intermittently observed longitudinal covariates. Cao H., Churpek M. M., Zeng D., and Fine J. P. (2015) <doi:10.1080/01621459.2014.957289>.

License GPL-2

Imports methods, stats

NeedsCompilation no

Repository CRAN

Date/Publication 2020-01-10 13:00:05 UTC

R topics documented:

`SurvLong-package` *Analysis of Proportional Hazards Model with Sparse Longitudinal Covariates*

Description

Kernel weighting methods for estimation of proportional hazards models with intermittently observed longitudinal covariates. Cao H., Churpek M. M., Zeng D., and Fine J. P. (2015) <doi:10.1080/01621459.2014.957289>

Details

Package:	SurvLong
Type:	Package
Version:	1.1
Date:	2020-01-08
License:	GPL-2

Author(s)

Hongyuan Cao, Mathew M. Churpek, Donglin Zeng, Jason P. Fine, and Shannon T. Holloway

Maintainer: Shannon T. Holloway <sthollow@ncsu.edu>

References

Cao H., Churpek M. M., Zeng D., Fine J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. Journal of the American Statistical Association, 110, 1187-1196.

See Also

`fullKernel`, `halfKernel`, `lastValue`, `nearValue`

`fullKernel` *Full Kernel Estimation with Forward and Backward Lagged Covariates*

Description

A kernel weighting scheme to evaluate the effects of longitudinal covariates on the occurrence of events when the time-dependent covariates are measured intermittently. Regression parameter estimation uses full kernel imputation of missing values with both forward and backward lagged covariates.

Usage

```
fullKernel(X, Z, tau, kType = "epan", bw = NULL, tol = 0.001,
           maxiter = 100, verbose = TRUE)
```

Arguments

`X`	an object of class data.frame. The structure of the data.frame must be {patient ID, event time, event indicator}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA. The event indicator is 1 if the event occurred; 0 if censored.
`Z`	an object of class data.frame. The structure of the data.frame must be {patient ID, time of measurement, measurement(s)}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA.
`tau`	an object of class numeric. The desired time point.
`kType`	An object of class character indicating the type of smoothing kernel to use in the estimating equation. Must be one of {"epan", "uniform", "gauss"}, where "epan" is the Epanechnikov kernel and "gauss" is the Gaussian kernel.
`bw`	If provided, bw is an object of class numeric or a numeric vector containing the bandwidths for which parameter estimates are to be obtained. If NULL, an optimal bandwidth will be determined using an adaptive selection procedure. The range of the bandwidth search space is taken to be 2*(Q3 - Q1)*n^-0.7 to 2*(Q3 - Q1)*n^-0.3, where Q3 is the 0.75 quantile and Q1 is the 0.25 quantile of the measurement times for the covariate and n is the effective number of patients, taken as the total number of patients that experienced an event.
`tol`	An object of class numeric. The minimum change in the regression parameters deemed to indicate convergence of the Newton-Raphson method.
`maxiter`	An object of class numeric. The maximum number of iterations used to estimate regression parameters.
`verbose`	An object of class logical. TRUE results in screen prints.

Value

A list is returned. If bandwidths are provided, each element of the list is a matrix, where the ith row corresponds to the ith bandwidth of input argument “bw," and the columns correspond to the model parameters. If the bandwidth is determined automatically, each element of the list is a named vector calculated at the optimal bandwidth.

`betaHat`	The estimated model coefficients.
`stdErr`	The standard error for each coefficient.
`zValue`	The estimated z-value for each coefficient.
`pValue`	The p-value for each coefficient.

If the bandwidth is determined automatically, three additional list elements are returned:

`optBW`	The estimated optimal bandwidth.
`minMSE`	The mean squared error at the optimal bandwidth.
`MSE`	The vector of MSE for each bandwidth.

Author(s)

Hongyuan Cao, Mathew M. Churpek, Donglin Zeng, Jason P. Fine, and Shannon T. Holloway

References

Cao H., Churpek M. M., Zeng D., Fine J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. Journal of the American Statistical Association, 110, 1187-1196.

See Also

halfKernel, lastValue, nearValue

Examples

```
data(SurvLongData)

exp <- fullKernel(X = X, Z = Z, tau = 1.0, bw = 0.015)
```

---

halfKernel — *Half Kernel Estimation with Backward Lagged Covariates*

Description

A kernel weighting scheme to evaluate the effects of longitudinal covariates on the occurrence of events when the time-dependent covariates are measured intermittently. Regression parameter estimation using half kernel imputation of missing values with backward lagged covariates.

Usage

```
halfKernel(X, Z, tau, kType = "epan", bw = NULL, tol = 0.001,
           maxiter = 100, verbose = TRUE)
```

Arguments

`X`	an object of class data.frame. The structure of the data.frame must be {patient ID, event time, event indicator}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA. The event indicator is 1 if the event occurred; 0 if censored.
`Z`	an object of class data.frame. The structure of the data.frame must be {patient ID, time of measurement, measurement(s)}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA.
`tau`	an object of class numeric. The desired time point.

`kType`	An object of class character indicating the type of smoothing kernel to use in the estimating equation. Must be one of {"epan", "uniform", "gauss"}, where "epan" is the Epanechnikov kernel and "gauss" is the Gaussian kernel.
`bw`	If provided, bw is an object of class numeric or a numeric vector containing the bandwidths for which parameter estimates are to be obtained. If NULL, an optimal bandwidth will be determined using an adaptive selection procedure. The range of the bandwidth search space is taken to be 2*(Q3 - Q1)*n^-0.7 to 2*(Q3 - Q1)*n^-0.3, where Q3 is the 0.75 quantile and Q1 is the 0.25 quantile of the measurement times for the covariate and n is the effective number of patients, taken as the total number of patients that experienced an event.
`tol`	An object of class numeric. The minimum change in the regression parameters deemed to indicate convergence of the Newton-Raphson method.
`maxiter`	An object of class numeric. The maximum number of iterations used to estimate regression parameters.
`verbose`	An object of class logical. TRUE results in screen prints.

Value

A list is returned. If bandwidths are provided, each element of the list is a matrix, where the ith row corresponds to the ith bandwidth of argument “bw," and the columns correspond to the model parameters. If the bandwidth is determined automatically, each element is a named vector calculated at the optimal bandwidth.

`betaHat`	The estimated model coefficients.
`stdErr`	The standard error for each coefficient.
`zValue`	The estimated z-value for each coefficient.
`pValue`	The p-value for each coefficient.

If the bandwidth is determined automatically, three additional list elements are returned:

`optBW`	The estimated optimal bandwidth.
`minMSE`	The mean squared error at the optimal bandwidth.
`MSE`	The vector of MSE for each bandwidth.

Author(s)

Hongyuan Cao, Mathew M. Churpek, Donglin Zeng, Jason P. Fine, and Shannon T. Holloway

References

Cao H., Churpek M. M., Zeng D., Fine J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. Journal of the American Statistical Association, 110, 1187-1196.

See Also

`fullKernel`, `lastValue`, `nearValue`

Examples

```
data(SurvLongData)

exp <- halfKernel(X = X, Z = Z, tau = 1.0, bw = 0.015)
```

---

lastValue *Last Value Carried Forward Method*

Description

A simple approach to evaluate the effects of longitudinal covariates on the occurrence of events when the time-dependent covariates are measured intermittently. Regression parameter are estimated using last value carried forward imputation of missing values.

Usage

```
lastValue(X, Z, tau, tol = 0.001, maxiter = 100, verbose = TRUE)
```

Arguments

X	an object of class data.frame. The structure of the data.frame must be {patient ID, event time, event indicator}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA. The event indicator is 1 if the event occurred; 0 if censored.
Z	an object of class data.frame. The structure of the data.frame must be {patient ID, time of measurement, measurement(s)}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA.
tau	an object of class numeric. The desired time point.
tol	An object of class numeric. The minimum change in the regression parameters deemed to indicate convergence of the Newton-Raphson method.
maxiter	An object of class numeric. The maximum number of iterations used to estimate regression parameters.
verbose	An object of class logical. TRUE results in screen prints.

Value

A list is returned.

betaHat	The estimated model coefficients.
stdErr	The standard error for each coefficient.
zValue	The estimated z-value for each coefficient.
pValue	The p-value for each coefficient.

Author(s)

Hongyuan Cao, Mathew M. Churpek, Donglin Zeng, Jason P. Fine, and Shannon T. Holloway

References

Cao H., Churpek M. M., Zeng D., Fine J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. Journal of the American Statistical Association, 110, 1187-1196.

See Also

fullKernel, halfKernel, nearValue

Examples

```
data(SurvLongData)
# A truncated dataset to keep example run time brief
exp <- lastValue(X = X[1:200,], Z = Z, tau = 1.0)
```

---

nearValue *Nearest Value Method*

Description

A simple approach to evaluate the effects of longitudinal covariates on the occurrence of events when the time-dependent covariates are measured intermittently. Regression parameter are estimated using the nearest value to impute missing values.

Usage

```
nearValue(X, Z, tau, tol = 0.001, maxiter = 100, verbose = TRUE)
```

Arguments

X	an object of class data.frame. The structure of the data.frame must be {patient ID, event time, event indicator}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA. The event indicator is 1 if the event occurred; 0 if censored.
Z	an object of class data.frame. The structure of the data.frame must be {patient ID, time of measurement, measurement(s)}. Patient IDs must be of class integer or be able to be coerced to class integer without loss of information. Missing values must be indicated as NA.
tau	an object of class numeric. The desired time point.
tol	An object of class numeric. The minimum change in the regression parameters deemed to indicate convergence of the Newton-Raphson method.
maxiter	An object of class numeric. The maximum number of iterations used to estimate regression parameters.
verbose	An object of class logical. TRUE results in screen prints.

Value

A list is returned.

betaHat	The estimated model coefficients.
stdErr	The standard error for each coefficient.
zValue	The estimated z-value for each coefficient.
pValue	The p-value for each coefficient.

Author(s)

Hongyuan Cao, Mathew M. Churpek, Donglin Zeng, Jason P. Fine, and Shannon T. Holloway

References

Cao H., Churpek M. M., Zeng D., Fine J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. Journal of the American Statistical Association, 110, 1187-1196.

See Also

fullKernel, halfKernel, lastValue

Examples

```
data(SurvLongData)
# A truncated dataset to keep example run time brief
exp <- nearValue(X = X[1:100,], Z = Z, tau = 1.0)
```

SurvLongData *Generated Sparse Longitudinal Data*

Description

For the purposes of the package examples, the data set was adapted from the numerical simulations of the original manuscript. Specifically, data was generated for 400 subjects. The total number of covariate observation times was Poisson distributed with intensity rate 8. The covariate observation times are generated from a uniform distribution Unif(0,1) independently. The covariate process is piecewise constant, with values being multivariate normal with mean 0, variance 1 and correlation exp(-|i - j|/20). The survival time were generated from the Cox model lambdat|Z(r),r<=t=lambda0 exp(beta Z(t)), where beta = 1.5, and lambda0 = 1.0. Covariates are dataset Z. Event times and indicators are dataset X.

Format

X is a data frame with 400 observations on the following 3 variables.

`ID` patient identifier, there are 400 patients.

`Time` the time to event or censoring

`Delta` a numeric vector with 0 denoting censoring and 1 event

Z is a data frame with 3237 observations on the following 3 variables.

`ID` patient identifier, there are 400 patients.

`obsTime` the covariate observation times.

`x1` the covariate generated through a piecewise constant function.

Source

Generated by Shannon T. Holloway in R.

References

Cao H., Churpek M. M., Zeng D., Fine J. P. (2015). Analysis of the proportional hazards model with sparse longitudinal covariates. Journal of the American Statistical Association, 110, 1187-1196.

Index