Package 'DOvalidation'

May 7, 2026

Type Package

Title Kernel Hazard Estimation with Best One-Sided and Double One-Sided Cross-Validation

Version 1.1.0

Date 2017-10-20

Author M.L. Gamiz, E. Mammen, M.D. Martinez-Miranda and J.P. Nielsen

Maintainer Maria Dolores Martinez-Miranda <mmiranda@ugr.es>

Description Local linear hazard estimator and its multiplicatively bias correction, including three bandwidth selection methods: best one-sided cross-validation, double one-sided cross-validation, and standard cross-validation.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2017-10-20 10:37:11 UTC

Contents

DOvalidation-package . . . 2
b.BO . . . 3
b.BO.MBC . . . 5
b.CV . . . 6
b.CV.MBC . . . 8
b.OSCV . . . 10
b.OSCV.MBC . . . 12
discretise.data . . . 14
DK . . . 15
hazard.LL . . . 16
hazard.LL.RH . . . 19
hazard.MBC . . . 20
Iceland . . . 22
K.epa . . . 23
K.sextic . . . 24
UK . . . 25
US . . . 26

Index 27

`DOvalidation-package` *Kernel Hazard Estimation with Best One-Sided and Double One-Sided Cross-Validation*

Description

Local linear hazard estimator and its multiplicatively bias correction, including three bandwidth selection methods: best one-sided cross-validation, double one-sided cross-validation, and standard cross-validation.

Details

Package:	DOvalidation
Type:	Package
Version:	1.1.0
Date:	2017-10-20
License:	GPL-2

Author(s)

M.L. Gamiz, E. Mammen, M.D. Martinez-Miranda and J.P. Nielsen

Maintainer: Maria Dolores Martinez-Miranda <mmiranda@ugr.es>

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

Examples

```
data(Iceland)
Oi<-Iceland$D
Ei<-Iceland$E
ti<-40:110  # time is age and it goes from 40 to 110 years
## Local linear hazard estimator
## and its multiplicatively bias corrected version (MBC)
## with best one-sided cross-validated bandwidths
## Note: use functions b.BO and b.BO.MBC to get these bandwidths
```

```
##       (48.7 and 14.6, respectively)
res.LL<-hazard.LL(xi=ti,Oi=Oi,Ei=Ei,x=ti,b=14.6)
res.MBC<-hazard.MBC(xi=ti,Oi=Oi,Ei=Ei,x=ti,b=48.7)
plot(ti,res.LL$hLL,main='Hazard estimates',xlab='age',ylab='',
    type='l',col=4,lwd=2)
lines(ti,res.MBC$hMBC,col=2,lwd=2)
legend("topleft",bt="n",c("Local linear", "MBC"),col=c(4,2),lwd=2)
```

---

b.BO *Best One-Sided Cross-Validation for Local Linear Hazards*

Description

Bandwidth selection for local linear hazard estimation using best one-sided cross-validation

Usage

```
b.BO(grid.b, nb , K = "sextic", type.bo = "Oi", xi, Oi, Ei, wei = "same")
```

Arguments

grid.b	Optional. A vector of bandwidths to minimise the cross-validation score. If not specified it will be considered an equally-spaced grid of nb bandwidths between "amp/(M+1)" and "amp/2" for "amp" being the range of xi and "M" its length.
nb	Optional. The number of bandwidths used to minimise the cross-validation score. If grid.b is provided then the argument nb will be ignored (if specified).
K	Indicates the kernel function to be considered in the local linear hazard estimator. Choose between values "epa" (for the epanechnikov kernel) or "sextic" (see details in hazard.LL for the definition).
type.bo	Choose between "Oi" or "Ei" to find the best side using the occurrences or the exposures, respectively.
xi	Vector of time points where the count data are given.
Oi	Vector with the number (counts) of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
wei	Indicates the weights used in the cross-validation score. Choose between the value "exposure" or "same". See details below.

Details

It is assumed that the data are given as count data i.e. number of occurrences and exposures.

The BO-validated bandwidth is calculated as the minimizer of a cross-validation score with a indirect kernel. If the score is strictly increasing or decreasing then a warning will be shown together with the selected bandwidth (in this case one of the extremes in grid.b, adjusted by the rescaling constant, which is 0.5371 for the Epanechnikov kernel, and 0.5874 for the sextic kernel).

The score is defined with two different weighting functions. This is controlled with the parameter `wei`. By default `wei="exposure"` that means that only areas where the exposure is significant contribute to the criterion. Specify `wei="same"` to allow all time points contribute the same to the criterion (see Gamiz et al. 2017).

Value

`bbo`	The onesided cross-validated bandwidth.
`ind.bo`	The position of the best one-sided cross-validated bandwidth into "grid.b".
`cvbo.values`	The values of the cross-validation score for each bandwidth in `grid.b`.
`b.grid`	The grid of bandwidths where the score has been evaluated.

Author(s)

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

See Also

`hazard.LL`,`b.OSCV`,`b.CV`

Examples

```
data(Iceland)
Oi<-Iceland$D
Ei<-Iceland$E
ti<-40:110  # time is age and it goes from 40 to 110 years
my.bs<-seq(20,40,length.out=30)
res.bo<-b.BO(grid.b=my.bs,K="sextic",type.bo="Oi",xi=ti,Oi=Oi,Ei=Ei,wei="same")
bbo<-res.bo$bbo
cvs<-res.bo$cvbo.values
plot(my.bs,cvs,main="BO-validation score",xlab="Bandwidth")
print(paste("The best one-sided cross-validated bandwidth is:", bbo,sep=" "))
```

b.BO.MBC *Best One-Sided Cross-Validation for Multiplicative Bias Corrected Hazard Estimators*

Description

Bandwidth selection for multiplicatively biased corrected local linear hazard estimation using best one-sided cross-validation

Usage

```
b.BO.MBC(grid.b, nb , K = "sextic", type.bo = "Oi", xi, Oi, Ei, wei = "same")
```

Arguments

grid.b	Optional. A vector of bandwidths to minimise the cross-validation score. If not specified it will be considered an equally-spaced grid of nb bandwidths between "amp/(M+1)" and "amp/2" for "amp" being the range of xi and "M" its length.
nb	Optional. The number of bandwidths used to minimise the cross-validation score. If grid.b is provided then the argument nb will be ignored (if specified).
K	Indicates the kernel function to be considered in the hazard. Choose between values "epa" (for the Epanechnikov kernel) or "sextic" (see details of hazard.MBC for the definition).
type.bo	Choose between "Oi" or "Ei" to find the best side using the occurrences or the exposures, respectively.
xi	Vector of time points where the count data are given.
Oi	Vector with the number (counts) of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
wei	Indicates the weights used in the cross-validation score. Choose between the value "exposure" or "same". See details below.

Details

The BO-validated bandwidth is calculated as the minimizer of a cross-validation score with a indirect kernel. If the score is strictly increasing or decreasing then a warning will be shown together with the selected bandwidth (in this case one of the extremes in grid.b, adjusted by the rescaling constant, which is 0.5948 for the Epanechnikov kernel, and 0.6501 for the sextic kernel).

The score is defined with two different weighting functions. This is controlled with the parameter wei. By default wei="exposure" that means that only areas where the exposure is significant contribute to the criterion. Specify wei="same" to allow all time points contribute the same to the criterion (see Gamiz et al. 2017).

Value

bbo	The onesided cross-validated bandwidth.
ind.bo	The position of the best one-sided cross-validated bandwidth into "grid.b".
cvbo.values	The values of the cross-validation score for each bandwidth in grid.b.
b.grid	The grid of bandwidths where the score has been evaluated.

Author(s)

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

See Also

hazard.MBC,b.OSCV.MBC,b.CV.MBC

Examples

```
data(Iceland)
Oi<-Iceland$D
Ei<-Iceland$E
ti<-40:110  # time is age and it goes from 40 to 110 years
my.bs<-seq(50,80,length.out=30)
res.bo<-b.BO.MBC(grid.b=my.bs,K="sextic",type.bo = "Oi",xi=ti,Oi=Oi,Ei=Ei,wei="same")
bbo<-res.bo$bbo
cvs<-res.bo$cvbo.values
plot(my.bs,cvs,main="BO-validation score",xlab="Bandwidth")
print(paste("The best one-sided cross-validated bandwidth is:", bbo,sep=" "))
```

---

b.CV *Least Squares Cross-Validation for Local Linear Hazards*

Description

Bandwidth selection for local linear hazard estimation using least squares cross-validation

Usage

```
b.CV(grid.b, nb , K = "epa", xi, Oi, Ei, wei = "exposure")
```

Arguments

grid.b	Optional. A vector of bandwidths to minimise the cross-validation score. If not specified it will be considered an equally-spaced grid of nb bandwidths between "amp/(M+1)" and "amp/2" for "amp" being the range of xi and "M" its length.
nb	Optional. The number of bandwidths used to minimise the cross-validation score. If grid.b is provided then the argument nb will be ignored (if specified).
K	Indicates the kernel function to be considered in the local linear hazard estimator. Choose between values "epa" (for the epanechnikov kernel) or "sextic" (see details of hazard.LL for the definition).
xi	Vector of time points where the count data are given.
Oi	Vector with the number (counts) of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
wei	Indicates the weights used in the cross-validation score. Choose between the value "exposure" or "same". See details below.

Details

It is assumed that the data are given as count data i.e. number of occurrences and exposures.

If the cross-validation score is strictly increasing or decreasing then a warning will be shown together with the cross-validated bandwidth (in this case one of the extremes in grid.b).

The cross-validation score is defined with two different weighting functions. This is controlled with the parameter wei. By default wei="exposure" that means that only areas where the exposure is significant contribute to the criterion. Specify wei="same" to allow all time points contribute the same to the criterion (see Gamiz et al. 2017).

Value

bcv	The cross-validated bandwidth.
ind.cv	The position of the cross-validated bandwidth into grid.b.
cv.values	The values of the cross-validation score for each bandwidth in grid.b.
b.grid	The grid of bandwidths where the cross-validation score has been evaluated.

Author(s)

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics, 28, 675-698.

See Also

hazard.LL,b.OSCV,b.BO

Examples

```
data(UK)
Oi<-UK$D
Ei<-UK$E
ti<-40:110  # time is age and it goes from 40 to 110 years
M<-length(ti)
my.bs<-seq(1,5,length=50)
res.cv<-b.CV(grid.b=my.bs,K="sextic",xi=ti,Oi=Oi,Ei=Ei)
bcv<-res.cv$bcv
cv.values<-res.cv$cv.values
plot(my.bs,cv.values,main="Cross-validation score",xlab="Bandwidth")
print(paste("The cross-validated bandwidth is:", bcv,sep=" "))
```

---

b.CV.MBC — *Least Squares Cross-Validation for Multiplicative Bias Corrected Hazard Estimators*

Description

Bandwidth selection for multiplicatively bias corrected local linear hazard estimation using least squares cross-validation

Usage

```
b.CV.MBC(grid.b, nb , K = "sextic", xi, Oi, Ei, wei = "same")
```

Arguments

grid.b	Optional. A vector of bandwidths to minimise the cross-validation score. If not specified it will be considered an equally-spaced grid of nb bandwidths between "amp/(M+1)" and "amp/2" for "amp" being the range of xi and "M" its length.
nb	Optional. The number of bandwidths used to minimise the cross-validation score. If "grid.b" is provided then the argument "nb" will be ignored (if specified).
K	Indicates the kernel function to be considered in the local linear hazard estimator. Choose between values "epa" (for the epanechnikov kernel) or "sextic" (see details of hazard.MBC for the definition).
xi	Vector of time points where the count data are given.
Oi	Vector with the number (counts) of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
wei	Indicates the weights used in the cross-validation score. Choose between the value "exposure" or "same". See details below.

Details

It is assumed that the data are given as count data i.e. number of occurrences and exposures.

If the cross-validation score is strictly increasing or decreasing then a warning will be shown together with the cross-validated bandwidth (in this case one of the extremes in "grid.b").

The cross-validation score is defined with two different weighting functions. This is controlled with the parameter `wei`. By default `wei="exposure"` that means that only areas where the exposure is significant contribute to the criterion. Specify `wei="same"` to allow all time points contribute the same to the criterion (see Gamiz et al. 2017).

Value

`bcv`	The cross-validated bandwidth.
`ind.cv`	The position of the cross-validated bandwidth into "grid.b".
`cv.values`	The values of the cross-validation score for each bandwidth in "grid.b".
`b.grid`	The grid of bandwidths where the cross-validation score has been evaluated.

Author(s)

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics, 28, 675-698.

See Also

`hazard.MBC`,`b.BO.MBC`,`b.CV`

Examples

```
data(Iceland)
Oi<-Iceland$D
Ei<-Iceland$E
ti<-40:110  # time is age and it goes from 40 to 110 years
my.bs<-seq(50,80,length.out=30)
res.cv<-b.CV.MBC(grid.b=my.bs,K="sextic",xi=ti,Oi=Oi,Ei=Ei,wei="same")
bcv<-res.cv$bcv
cv.values<-res.cv$cv.values
plot(my.bs,cv.values,main="Cross-validation score",xlab="Bandwidth")
print(paste("The cross-validated bandwidth is:", bcv,sep=" "))
```

b.OSCV *DO-Validation for Local Linear Hazards*

Description

Bandwidth selection for local linear hazard estimation using DO-validation and one-sided (left or right) cross-validation

Usage

```
b.OSCV(grid.b, nb , K = "epa", Ktype = "left", xi, Oi, Ei, wei = "exposure")
```

Arguments

grid.b	Optional. A vector of bandwidths to minimise the cross-validation score. If not specified it will be considered an equally-spaced grid of nb bandwidths between "amp/(M+1)" and "amp/2" for "amp" being the range of xi and "M" its length.
nb	Optional. The number of bandwidths used to minimise the cross-validation score. If grid.b is provided then the argument nb will be ignored (if specified).
K	Indicates the kernel function to be considered in the local linear hazard estimator. Choose between values "epa" (for the Epanechnikov kernel) or "sextic" (see details of hazard.LL for the definition).
Ktype	Choose between "left" or "right" for left- or right- sided cross-validation respectively.
xi	Vector of time points where the count data are given.
Oi	Vector with the number (counts) of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
wei	Indicates the weights used in the cross-validation score. Choose between the value "exposure" or "same". See details below.

Details

It is assumed that the data are given as count data i.e. number of occurrences and exposures.

The DO-validated bandwidth is calculated as the average of left- and right- sided cross-validation (see example below).

If the one-sided cross-validation score is strictly increasing or decreasing then a warning will be shown together with the onesided cross-validated bandwidth (in this case one of the extremes in grid.b, adjusted by the rescaling constant, which is 0.5371 for the Epanechnikov kernel, and 0.5874 for the sextic kernel).

The score is defined with two different weighting functions. This is controlled with the parameter wei. By default wei="exposure" that means that only areas where the exposure is significant contribute to the criterion. Specify wei="same" to allow all time points contribute the same to the criterion (see Gamiz et al. 2017).

Value

boscv	The onesided cross-validated bandwidth.
ind.oscv	The position of the one-sided cross-validated bandwidth into grid.b.
oscv.values	The values of the one-sided cross-validation score for each bandwidth in grid.b.
b.grid	The grid of bandwidths where the one-sided cross-validation score has been evaluated.

Author(s)

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

See Also

hazard.LL,b.BO,b.CV

Examples

```
data(UK)
Oi<-UK$D
Ei<-UK$E
ti<-40:110  # time is age and it goes from 40 to 110 years
M<-length(ti)
my.bs<-seq(3,10,length=50)
## The left one-sided cross-validated bandwidth
res.left<-b.OSCV(grid.b=my.bs,K="sextic",Ktype="left",xi=ti,Oi=Oi,Ei=Ei)
bleft<-res.left$boscv
## The right one-sided cross-validated bandwidth
res.right<-b.OSCV(grid.b=my.bs,K="sextic",Ktype="right",xi=ti,Oi=Oi,Ei=Ei)
bright<-res.right$boscv
## The DO-validated bandwidth
bdo<-(bleft+bright)/2
print(paste("DO-validated bandwidth= ", bdo, sep=""))
```

b.OSCV.MBC *DO-Validation for Multiplicative Bias Corrected Hazard Estimators*

Description

Bandwidth selection for the multiplicatively bias corrected local linear hazard estimation using DO-validation and one-sided (left or right) cross-validation

Usage

```
b.OSCV.MBC(grid.b, nb , K = "sextic", Ktype = "left", xi, Oi, Ei, wei = "same")
```

Arguments

grid.b	Optional. A vector of bandwidths to minimise the cross-validation score. If not specified it will be considered an equally-spaced grid of nb bandwidths between "amp/(M+1)" and "amp/2" for "amp" being the range of xi and "M" its length.
nb	Optional. The number of bandwidths used to minimise the cross-validation score. If grid.b is provided then the argument nb will be ignored (if specified).
K	Indicates the kernel function to be considered in the hazard estimator. Choose between values "epa" (for the Epanechnikov kernel) or "sextic" (see details of hazard.MBC for the definition).
Ktype	Choose between "left" or "right" for left- or right- sided cross-validation, respectively.
xi	Vector of time points where the count data are given.
Oi	Vector with the number (counts) of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
wei	Indicates the weights used in the cross-validation score. Choose between the value "exposure" or "same". See details below.

Details

It is assumed that the data are given as count data i.e. number of occurrences and exposures.

The DO-validated bandwidth is calculated as the average of left- and right- sided cross-validation (see example below).

If the one-sided cross-validation score is strictly increasing or decreasing then a warning will be shown together with the onesided cross-validated bandwidth (in this case one of the extremes in grid.b, adjusted by the rescaling constant, which is 0.5948 for Epanechnikov kernel, and 0.6501 for the sextic kernel).

The score is defined with two different weighting functions. This is controlled with the parameter wei. By default wei="exposure" that means that only areas where the exposure is significant contribute to the criterion. Specify wei="same" to allow all time points contribute the same to the criterion (see Gamiz et al. 2017).

Value

boscv	The one-sided cross-validated bandwidth.
ind.oscv	The position of the one-sided cross-validated bandwidth into grid.b.
oscv.values	The values of the one-sided cross-validation score for each bandwidth in grid.b.
b.grid	The grid of bandwidths where the one-sided cross-validation score has been evaluated.

Author(s)

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

See Also

hazard.MBC,b.CV.MBC,b.BO.MBC

Examples

```
data(Iceland)
Oi<-Iceland$D
Ei<-Iceland$E
ti<-40:110  # time is age and it goes from 40 to 110 years
my.bs<-seq(50,80,length.out=30)
## The left one-sided cross-validated bandwidth
res.left<-b.OSCV.MBC(grid.b=my.bs,K="sextic",Ktype="left",
    xi=ti,Oi=Oi,Ei=Ei,wei="same")
bleft<-res.left$boscv
## The right one-sided cross-validated bandwidth
res.right<-b.OSCV.MBC(grid.b=my.bs,K="sextic",Ktype="right",
    xi=ti,Oi=Oi,Ei=Ei,wei="same")
bright<-res.right$boscv
## The do-validated bandwidth
bdo<-(bleft+bright)/2
print(paste("DO-validated bandwidth= ", bdo, sep=""))
```

`discretise.data` *Aggregate data in the form of occurrences and exposures*

Description

Aggregate data in the form of occurrences and exposures from individual survival data (possibly right censored and/or left truncated).

Usage

```
discretise.data(Li, Zi, deltai, xi, M)
```

Arguments

`Li`	Vector of truncation levels: the datum is registered only if the life time is greater than the truncation level.
`Zi`	Vector of observed life times (Zi=min(Ti,Ci) with Ci censoring value and Ti the true life time).
`deltai`	Vector with non-censoring indicator values (0 if datum is censored, 1 otherwise).
`xi`	Optional. Vector with the grid of time points where the occurrences and exposures should be calculated. If not provide the grid is calculated automatically.
`M`	Optional. A positive scalar used as the grid size. If not provided it is chosen automatically.

Details

The hazard estimators and bandwidth selectors available in the DOvalidation package work from data aggregated in the form of occurrences and exposures. This function can be used to work with individual survival data in the form (Li,Zi,deltai) – left-truncation level (Li), observed time (Zi) and non-censoring indicator (deltai). If data are not truncated then Li can be chosen as 0.

Value

`xi`	Vector with the time grid points.
`Oi`	Vector with the calculated occurrences at the time grid points.
`Ei`	Vector with the calculated exposures at the time grid points.

Author(s)

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

See Also

hazard.LL

Examples

```
## A simple example with right censored data
Zi<-c(3,6,7,7,8,10,11,11,11,12,13,13,14,16,20,20,22,32,34,36)
n<-length(Zi)
Li<-deltai<-rep(0,n)
ind.cens<-c(1,3,4,8,9,13,14,15,16)
deltai[-ind.cens]<-1
## Obtain the occurrences and exposures for a grid of 6 time points
res<-discretise.data(Li,Zi,deltai,M=6)
## Now calculate the local linear hazard estimator
hazard.LL(res$xi,res$Oi,res$Ei,res$xi,b=10)
```

DK *Denmark Female Mortality Data*

Description

Mortality data of women in the calendar year 2006 from Denmark. The data were obtained from the Humam Mortality Database. Only ages from 40 to 110 have been included.

Usage

```
data(DK)
```

Format

This data frame contains 71 rows and the following 2 columns.

D Death counts for women of ages between 40 and 110 during the calendar year 2006.

E "Person-years" lived in the female population during the year 2006 for each age-group (from 40 to 110)

Source

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Spreeuw, J., Nielsen, J.P. and Jarner, S.F. (2013). A visual test of mixed hazard models, SORT, 37, 149-170.

Examples

```
data(DK)
```

`hazard.LL` *Local Linear Hazard Estimator (Natural Weighting)*

Description

Local linear estimator of the unidimensional hazard (or hazard rate) with natural weighting introduced by Nielsen and Tanggaard (2001).

Usage

```
hazard.LL(xi, Oi, Ei, x, b, K="epa", Ktype="symmetric" , CI=FALSE)
```

Arguments

`xi`	Vector of time points where the counts data are given.
`Oi`	Vector with the number of occurrences observed at each time point (`xi`).
`Ei`	Vector with the observed exposure at each time point (`xi`).
`x`	Vector (or scalar) with the (time) grid points where the hazard estimator will be evaluated.
`b`	A positive scalar used as the bandwidth.
`K`	Indicates the kernel function to be considered in the estimator. Choose between values `"epa"` (for the Epanechnikov kernel) or `"sextic"` (see details for its expression).
`Ktype`	Indicates the type of kernel to be used. Choose among `"symmetric"` for the usual kernel definition (chosen in the argument `K`), `"left"` for the left-sided version of the kernel, or `"right"` for the right-sided version. See details below.
`CI`	Logical. If `TRUE` then 95% pointwise confidence intervals are provided for the hazard function.

Details

The estimator is calculated assuming that the data are given as count data i.e. number of occurrences and exposures. The function allows to consider two different kernels using the argument `K`. These are: Epanechnikov, K(u)=.75*(1-u^2)*(abs(u)<1), and sextic K(u)=(3003/2048)*(1-(u)^2)^6)*(abs(u)<1). The argument `Ktype` will define the usual estimator with whole support kernel as it is defined by `K` or the one-sided versions using left-sided kernel, 2*K(u)*(u<0), or right-sided kernel 2*K(u)*(u>0). See more details in Gamiz et al. (2016).

Value

x	Vector (or scalar) with the (time) grid points where the hazard estimator has been evaluated.
OLL	Vector with the smoothed occurrences (using the local linear kernel).
ELL	Vector with the smoothed exposures (using the local linear kernel).
hLL	Vector (or scalar) with the resulting hazard estimates at grid points x.
OLL.norm	Vector with the normalized smoothed occurrences (the smoothing weights are defined as for O.LL but normalized to add up one.
ELL.norm	Vector with the normalized smoothed exposures (the smoothing weights are defined as for E.LL but normalized to add up one.
CI.inf	Vector with the lower limits for the 95% confidence intervals. If CI=FALSE then NA values are provided.
CI.sup	Vector with the upper limits for the 95% confidence intervals. If CI=FALSE then NA values are provided.

Author(s)

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

See Also

hazard.LL.RH,hazard.MBC

Examples

```
## Calculation of the local linear hazard estimator with do-validated bandwidth.
## The hazard estimator is shown and decomposed into smoothed occurrences and exposures.
## This example is described in Gamiz et al. (2016).
data(UK)
Oi<-UK$D
Ei<-UK$E
ti<-40:110  # time is age and it goes from 40 to 110 years
M<-length(ti)
country<-'UK'
bdo<-5.11
resLL.do<-hazard.LL(xi=ti,Oi=Oi,Ei=Ei,x=ti,b=bdo,K="sextic",Ktype="symmetric",CI=TRUE)

## The local linear hazard estimate is hLL.do below
hLL.do<-resLL.do$hLL

## The smoothed occurrences and exposures are:
```

```
ELL.norm.do<-resLL.do$ELL.norm
OLL.norm.do<-resLL.do$OLL.norm

## The 95% pointwise confidence intervals based on the asymptotics are
hLL.do.inf<-resLL.do$CI.inf
hLL.do.sup<-resLL.do$CI.sup

# Now we plot the hazard estimator with confidence intervals
old.par<-par(mar=c(3,1.5,1.5,1.5),oma=c(2,0.5,0.5,0.2),
mgp=c(1.5,0.5,0),cex.axis=1,cex.main=1.5,mfrow=c(3,2))

#hazard estimate
tit<-paste(country,"Hazard estimate",sep= ' - ' )
yy<-range(c(hLL.do.inf,hLL.do.sup),na.rm=TRUE)
plot(ti,hLL.do,main=tit,xlab='age',ylab='',type='l',lwd=2,ylim=yy)
# the confidence bands
x1<-ti;x2<-ti[M:1]
y1<-hLL.do.sup;y2<-hLL.do.inf[M:1]
polygon(c(x1,x2,x1[1]),c(y1,y2,y1[1]),col=gray(0.7),border=FALSE)
lines(ti,hLL.do,lty=1,lwd=2,col=1)

## Zooming at the old mortality
ind.ages<- -c(1:60)  ## only women with ages 100 or higher
ti2<-ti[ind.ages];M2<-length(ti2)
yy2<-range(c(hLL.do.inf[ind.ages],hLL.do.sup[ind.ages]),na.rm=TRUE)
plot(ti2,hLL.do[ind.ages],main=tit,xlab='age',ylab='',type='l',
lwd=2,ylim=yy2)
# the confidence intervals
x1<-ti2;x2<-ti2[M2:1]
y1<-hLL.do.sup[ind.ages];hLL.do.inf2<-hLL.do.inf[ind.ages]
y2<-hLL.do.inf2[M2:1]
polygon(c(x1,x2,x1[1]),c(y1,y2,y1[1]),col=gray(0.7),border=FALSE)
lines(ti2,hLL.do[ind.ages],lty=1,lwd=2,col=1)

## We decompose the estimator in the smooth occurrences and exposures
#   The occurrences with a zoom at old-age mortality
yy<-range(OLL.norm.do,na.rm=TRUE)
plot(ti,OLL.norm.do,main="Smoothed occurrences",xlab='age',ylab='',type='l',
lwd=2,ylim=yy)
yy2<-range(OLL.norm.do[ind.ages],na.rm=TRUE)
plot(ti2,OLL.norm.do[ind.ages],main="Smoothed occurrences",xlab='age',
ylab='',type='l',lwd=2,ylim=yy2)

#   The exposures with a zoom at old-age mortality
yy<-range(ELL.norm.do,na.rm=TRUE)
plot(ti,ELL.norm.do,main="Smoothed exposures",xlab='age',ylab='',type='l',
lwd=2,ylim=yy)
yy2<-range(ELL.norm.do[ind.ages],na.rm=TRUE)
plot(ti2,ELL.norm.do[ind.ages],main="Smoothed exposures",xlab='age',ylab='',
type='l',lwd=2,ylim=yy2)

# Revert the changes made in the graphics options
par(old.par)
```

hazard.LL.RH *Local Linear Hazard Estimator (Ramlau-Hansen Weighting)*

Description

Local linear estimator of the unidimensional hazard (or hazard rate) with Ramlau-Hansen weighting as was defined by Nielsen and Tanggaard (2001).

Usage

```
hazard.LL.RH(xi , Oi , Ei , x , b , K="epa")
```

Arguments

xi	Vector of time points where the counts data are given.
Oi	Vector with the number (counts) of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
x	Vector (or scalar) with the (time) grid points where the hazard estimator will be evaluated.
b	A positive scalar used as the bandwidth.
K	Indicates the kernel function to be considered in the estimator. Choose between values "epa" (for the Epanechnikov kernel) or "sextic" (see details for its expression).

Details

The estimator is calculated assuming that the data are given as count data i.e. number of occurrences and exposures. The function allows to consider two different kernels using the argument K. These are: Epanechnikov, K(u)=.75*(1-u^2)*(abs(u)<1), and sextic K(u)=(3003/2048)*(1-(u)^2)^6)*(abs(u)<1).

Value

x	Vector (or scalar) with the (time) grid points where the hazard estimator has been evaluated.
hLL	Vector (or scalar) with the resulting hazard estimates at grid points x.

Author(s)

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

See Also

hazard.LL

Examples

```
## This example was described by Gamiz et al. (2016) to analyse the exposure robustness of
## local linear hazards with natural weigthing and Ramlau-Hansen weighting
data(Iceland)
Ei<-Iceland$E
Oi<-Iceland$D
xi<-40:110
n<-length(xi)
x<-seq(xi[1],xi[n],length=100)

## Hazard estimates with the original data
b0<-11.9899
alphaIC17<-hazard.LL.RH(xi,Oi,Ei,x,b=b0,K="sextic")$hLL
alLL17<-hazard.LL(xi,Oi,Ei,x,b=b0,K="sextic",Ktype="symmetric")$hLL
hi<-Oi/Ei;hi[Ei==0]<-0
print(round(hi[60:71],3))
## Hazard estimates with the modified data (one change in the exposure)
Ei2<-Ei; Ei2[67]<-2/365
alphaIC005<-hazard.LL.RH(xi,Oi,Ei2,x,b=b0,K="sextic")$hLL
alLL005<-hazard.LL(xi,Oi,Ei2,x,b=b0,K="sextic",Ktype="symmetric")$hLL

## Figure: Exposure robustness
old.par<-par(mfrow=c(2,2))
plot(x[73:100],alphaIC17[73:100],lwd=2,type='l',main='Exposure: 0.17',
xlab='',ylab='Ramlau-Hansen weighting')
plot(x[73:100],alphaIC005[73:100],lwd=2,type='l',main='Exposure: 0.005',
xlab='',ylab='Ramlau-Hansen weighting')
plot(x[73:100],alLL17[73:100],lwd=2,type='l',main='Exposure: 0.17',
xlab='',ylab='Natural  weighting')
plot(x[73:100],alLL005[73:100],lwd=2,type='l',main='Exposure: 0.005',
xlab='',ylab='Natural weighting')

par(old.par)
```

---

hazard.MBC *Multiplicative Bias Corrected Hazard Estimator*

Description

Multiplicatively bias corrected local linear estimator of the unidimensional hazard with natural weighting introduced by Nielsen and Tanggaard (2001).

Usage

```
hazard.MBC(xi, Oi, Ei, x, b, K="sextic", Ktype="symmetric")
```

Arguments

xi	Vector of time points where the counts data are given.
Oi	Vector with the number of occurrences observed at each time point (xi).
Ei	Vector with the observed exposure at each time point (xi).
x	Vector (or scalar) with the (time) grid points where the hazard estimator will be evaluated.
b	A positive scalar used as the bandwidth.
K	Indicates the kernel function to be considered in the estimator. Choose between values "epa" (for the Epanechnikov kernel) or "sextic" (see details for its expression).
Ktype	Indicates the type of kernel to be used. Choose among "symmetric" for the usual kernel definition (chosen in the argument K), "left" for the left-sided version of the kernel, or "right" for the right-sided version. See details below.

Details

The estimator is calculated assuming that the data are given as count data i.e. number of occurrences and exposures. The function allows to consider two different kernels using the argument K. These are: Epanechnikov, K(u)=.75*(1-u^2)*(abs(u)<1), and sextic K(u)=(3003/2048)*(1-(u)^2)^6)*(abs(u)<1). The argument Ktype will define the usual estimator with whole support kernel as it is defined by K or the one-sided versions using left-sided kernel, 2*K(u)*(u<0), or right-sided kernel 2*K(u)*(u>0). See more details in Gamiz et al. (2017).

Value

x	Vector (or scalar) with the (time) grid points where the hazard estimator has been evaluated.
hMBC	Vector (or scalar) with the resulting hazard estimates at grid points x.

Author(s)

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Martinez-Miranda, M.D. and Nielsen, J.P. (2017). Multiplicative local linear hazard estimation and best one-sided cross-validation. Available at http://arxiv.org/abs/1710.05575

Nielsen, J.P. and Tanggaard, C. (2001). Boundary and bias correction in kernel hazard estimation. Scandinavian Journal of Statistics,28, 675-698.

See Also

hazard.LL

Examples

```
data(Iceland)
Oi<-Iceland$D
Ei<-Iceland$E
ti<-40:110  # time is age and it goes from 40 to 110 years
res<-hazard.MBC(xi=ti,Oi=Oi,Ei=Ei,x=ti,b=48.7)
plot(ti,res$hMBC,main='Hazard estimate',xlab='age',ylab='',type='l',lwd=2)
```

Iceland *Iceland Female Mortality Data*

Description

Mortality data of women in the calendar year 2006 from Iceland. The data were obtained from the Humam Mortality Database. Only ages from 40 to 110 have been included.

Usage

```
data(Iceland)
```

Format

This data frame contains 71 rows and the following 2 columns.

`D` Death counts for women of ages between 40 and 110 during the calendar year 2006

`E` "Person-years" lived in the female population during the year 2006 for each age-group (from 40 to 110).

Source

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Spreeuw, J., Nielsen, J. P. and Jarner, S. F. (2013). A visual test of mixed hazard models, SORT, 37, 149-??170.

Examples

```
data(Iceland)
```

K.epa *Epanechnikov Kernel*

Description

Evaluation of the Epanechnikov kernel function

Usage

```
K.epa(u)
```

Arguments

u	A vector (or scalar) with the evaluation point(s).

Value

The value of the kernel function at u.

Author(s)

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

See Also

K.sextic

Examples

```
curve(K.epa,-1.5,1.5,main="Epanechnikov kernel",ylab="K(u)",xlab="u")
# The left onesided
K.epa.left<-function(u) return(2*K.epa(u)*(u<0))
curve(K.epa.left,-1.5,1.5,main="Left onesided Epanechnikov kernel",ylab="K(u)",xlab="u")
```

K.sextic *Sextic Kernel*

Description

Evaluation of the Sextic kernel function

Usage

```
K.sextic(u)
```

Arguments

u	A vector (or scalar) with the evaluation point(s).

Value

The value of the kernel function at u.

Author(s)

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

See Also

K.epa

Examples

```
curve(K.sextic,-1.5,1.5,main="Sextic kernel",ylab="K(u)",xlab="u")
# The left onesided
K.sextic.left<-function(u) return(2*K.sextic(u)*(u<0))
curve(K.sextic.left,-1.5,1.5,main="Left onesided sextic kernel",ylab="K(u)",xlab="u")
```

UK *UK Female Mortality Data*

Description

Mortality data of women in the calendar year 2006 from United Kingdom. The data were obtained from the Humam Mortality Database. Only ages from 40 to 110 have been included.

Usage

```
data(UK)
```

Format

This data frame contains 71 rows and the following 2 columns.

`D` Death counts for women of ages between 40 and 110 during the calendar year 2006.

`E` "Person-years" lived in the female population during the year 2006 for each age-group (from 40 to 110)

Source

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Spreeuw, J., Nielsen, J.P. and Jarner, S.F. (2013). A visual test of mixed hazard models, SORT, 37, 149-170.

Examples

```
data(UK)
```

US *US Female Mortality Data*

Description

Mortality data of women in the calendar year 2006 from United States. The data were obtained from the Humam Mortality Database. Only ages from 40 to 110 have been included.

Usage

```
data(US)
```

Format

This data frame contains 71 rows and the following 2 columns.

`D` Death counts for women of ages between 40 and 110 during the calendar year 2006. Some of these numbers are estimates (of population size or numbers of deaths), not actual counts, and therefore may be expressed as non-integers.

`E` "Person-years" lived in the female population during the year 2006 for each age-group (from 40 to 110)

Source

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de

References

Gamiz, M.L., Mammen, E., Martinez-Miranda, M.D. and Nielsen, J.P.(2016). Double one-sided cross-validation of local linear hazards. Journal of the Royal Statistical Society B, 78, 755-779.

Spreeuw, J., Nielsen, J.P. and Jarner, S.F. (2013). A visual test of mixed hazard models, SORT, 37, 149-170.

Examples

```
data(US)
```

Index

∗ **datasets**
DK, 15
Iceland, 22
UK, 25
US, 26
∗ **distribution**
K.epa, 23
K.sextic, 24
∗ **nonparametric**
b.BO, 3
b.BO.MBC, 5
b.CV, 6
b.CV.MBC, 8
b.OSCV, 10
b.OSCV.MBC, 12
discretise.data, 14
DOvalidation-package, 2
hazard.LL, 16
hazard.LL.RH, 19
hazard.MBC, 20
∗ **package**
DOvalidation-package, 2
∗ **survival**
b.BO, 3
b.BO.MBC, 5
b.CV, 6
b.CV.MBC, 8
b.OSCV.MBC, 12
discretise.data, 14
DOvalidation-package, 2
hazard.LL, 16
hazard.LL.RH, 19
hazard.MBC, 20

b.BO, 3, *8*, *11*
b.BO.MBC, 5, *9*, *13*
b.CV, *4*, 6, *9*, *11*
b.CV.MBC, *6*, 8, *13*
b.OSCV, *4*, *8*, 10
b.OSCV.MBC, *6*, 12
discretise.data, 14
DK, 15
DOvalidation *(*DOvalidation-package*)*, 2
DOvalidation-package, 2

hazard.LL, *3*, *4*, *7*, *8*, *10*, *11*, *15*, 16, *20*, *22*
hazard.LL.RH, *17*, 19
hazard.MBC, *5*, *6*, *8*, *9*, *12*, *13*, *17*, 20

Iceland, 22

K.epa, 23, *24*
K.sextic, *23*, 24

UK, 25
US, 26