

Package ‘SmoothWin’

July 27, 2019

Type Package

Version 3.0.0

Date 2019-07-26

Author Hamed Haselimashhadi <hamedhm@ebi.ac.uk>

Maintainer Hamed Haselimashhadi <hamedhm@ebi.ac.uk>

Depends R (>= 3.5)

Description The main function in the package utilizes a windowing function in the form of an exponential weighting function to linear models. The bandwidth and sharpness of the window are controlled by two parameters. Then, a series of tests are used to identify the right parameters of the window (see Hamed Haselimashhadi et al (2019) <<https://www.biorxiv.org/content/10.1101/656678v1>>).

Title Soft Windowing on Linear Regression

License LGPL (>= 2)

Imports nlme, Rfast

URL <http://hamedhaseli.webs.com>, <https://www.mousephenotype.org/>

NeedsCompilation no

Repository CRAN

Date/Publication 2019-07-27 21:40:10 UTC

R topics documented:

expWeight	1
plot.SmoothWin	4
SmoothWin	5

 expWeight

This function computes the smooth windowing weights

Description

The symmetric weight generating function (SWGf). This function computes the exponential weights/kernel (soft windowing weights) for different shapes (k) and bandwidth (l) and plots the weights.

Usage

```
expWeight (
  t           ,
  k           ,
  l           ,
  m = 0       ,
  direction = c(1, 1) ,
  plot = FALSE ,
  zeroCompensation = 0 ,
  cdf = plogis ,
  progress   = FALSE ,
  ...
)
```

Arguments

t	Vector of numeric time. A vector of positive continuous values for time
k	A single positive value for sharpness
l	A single non-negative value for bandwidth
m	Vector of indices. The index of the modes on 't' (modes are the peak of the windows)
direction	Vector of two numeric values. A vector of the form on (Left,right). The first element specifies the speed of expansion of the window(s) from the left and the second value for the right expansion. Setting to (0,1) and (1,0) lead to right and left expansions of the windows respectively. Default (1,1) that is the window(s) expand symmetrically from both sides.
plot	Logical flag. Setting to TRUE shows a plot of the weights
zeroCompensation	Single non-negative value. Setting to any non-negative value would replace all (weights \leq zeroCompensation) with zeroCompensation. Default 0 (zero)
cdf	A cdf function preferably symmetric. The cdf function is used for the (window) weight generating function. The function must have two parameters precisely a location such as mean and a scale. Standard cdf functions such as pnorm, pcauchy and plogis (default) can be used. For an example of custom made function we define uniform function as below:

*punif0 = function(x, mean = 0.5, sd = sqrt(1/12)) a = mean - sqrt(3) * sd; b = mean + sqrt(3)*

progress Logical flag. Setting to TRUE shows the progress of the function

... Other parameters that can be passed to the 'plot()' function such as pch, colour etc.

Value

A numeric vector of weights

Author(s)

Hamed Haselimashhadi <hamedhm@ebi.ac.uk>

See Also

SmoothWin

Examples

```
par(mfrow = c(4, 1))
#####
# Example 1 - no merging happens between windows
#####
weight = expWeight(
  t = 1:100           ,
  k = 5               ,
  l = 10              ,
  m = c(25, 50, 75)  ,
  plot = TRUE        ,
  ### Passed parameters to the plot function
  type = 'l'         ,
  lty = 2            ,
  lwd = 3            ,
  main = '1. If windows do not intersect, then wont merge! (l=10, k=5)'
)

#####
# Example 2 - merging in windows
#####
weight = expWeight(
  t = 1:100           ,
  k = 5               ,
  l = 15              ,
  m = c(25, 50, 75)  ,
  plot = TRUE        ,
  ### Passed parameters to the plot function
  type = 'l'         ,
  lty = 2            ,
  lwd = 3            ,
```

```

    main = '2. If windows intersect, then merge! (l=15, k=5)'
  )

#####
# Example 3.1 - partial merging in windows
#####
weight = expWeight(
  t = 1:100                                ,
  k = 1                                    ,
  l = 12                                    ,
  m = c(25, 50, 75)                        ,
  plot = TRUE                              ,
  ### Passed parameters to the plot function
  type = 'l'                               ,
  lty = 2                                  ,
  lwd = 3                                  ,
  main = '3.1 If windows intersect with small k, then partially merge! (l=12, k=1)'
)

#####
# Example 3.2 - partial merging in windows
#####
weight = expWeight(
  t = 1:100                                ,
  k = .1                                    ,
  l = 12                                    ,
  m = c(25, 50, 75)                        ,
  plot = TRUE                              ,
  ### Passed parameters to the plot function
  type = 'l'                               ,
  lty = 2                                  ,
  lwd = 3                                  ,
  main = '3.2 If windows intersect with small k, then partially merge! (l=12, k=0.1)'
)

```

plot.SmoothWin

Plot function for the SmoothWin object

Description

This function plots a SmoothWin object

Usage

```

## S3 method for class 'SmoothWin'
plot(x,
      ylab = 'Response'                ,
      xlab = 'Time (continuous)'      ,
      sub  = NULL                      ,

```

```

col      = NULL      ,
digits  = 2          ,
...
)

```

Arguments

<code>x</code>	SmoothWin object
<code>ylab</code>	Label on the y axis. Default 'Response'
<code>xlab</code>	Label on the x axis. Default 'Time (continuous)'
<code>sub</code>	See the 'sub' parameter in 'plot()' function. If left NULL then some information about the final window will be shown. Default NULL
<code>col</code>	Colour parameter for the points. Set to NULL to use the default colouring (spectrum colouring). Default NULL
<code>digits</code>	The number of visible digits for l, k and SWS. Default 2
<code>...</code>	Optional parameters that can be passed to the 'plot'/'qqPlot' function. See 'car' package for the qqPlot function

Author(s)

Hamed Haselimashhadi <hamedhm@ebi.ac.uk>

See Also

SmoothWin

Examples

```
example(SmoothWin)
```

SmoothWin

Implementation of the soft windowing for linear models

Description

Implementation of the (symmetric) soft windowing on a range of methods/models by imposing weights on the model.

- The function accepts a model fit, such as 'lm', 'lme', 'glm' etc., as the input and fits a window to it.
- The parameters "k" and "l" control the shape and bandwidth of the windowing function respectively.
- There are several other parameters to cope with the different scenarios/models/window shapes.
- The default settings of the function is adapted to International Mouse Phenotyping Consortium (IMPC) statistical pipeline

Usage

```

SmoothWin(object
  data
  t
  m
  l = function(ignore.me.in.default) {
    r = SmoothWin:::lseq(
      from = 1
      to = max(abs(t[m] - min(t, na.rm = TRUE))
        abs(t[m] - max(t, na.rm = TRUE)), 1)
      length.out = min(500, max(1, diff(range(
        t, na.rm = TRUE
      ))))
    )
    r = unique(round(r))
    return(r)
  }
  k = SmoothWin:::lseq(from = .5
    to = 10
    length.out = 50)
  min.obs = function(ignore.me.in.default) {
    lutm = length(unique(t[m]))
    r = ifelse(lutm > 1, 35, max(pi * sqrt(length(t)), 35))
    r = max(r * lutm, length(m), na.rm = TRUE)
    r = min(r, length(t), na.rm = TRUE)
    return(r)
  }
  direction = c(1, 1)
  weightFUN = function(x) {
    x
  }
  residFun = function(x) {
    resid(x)
  }
  predictFun = function(x) {
    predict(x)
  }
  weightORthreshold = 'weight'
  cdf = plogis
  check = 2
  sensitivity = c(1, 1, 1, 0)
  pvalThreshold = c(0, 0, 0, 0)
  threshold = sqrt(.Machine$double.eps) * 10
  zeroCompensation = 0
  messages = TRUE
  seed = NULL
  simple.output = FALSE
  debug = FALSE

```

...)

Arguments

<code>object</code>	The fitted model. The object must support 'update(weights =)'. See examples
<code>data</code>	data.frame. Input data that is used to fit the initial model
<code>t</code>	Vector of (numeric) time values.
<code>m</code>	Vector of integers (peaks). Mode indices on the time component. For example 10, 11, 12. Note that it is different from <code>t[10]</code> , <code>t[11]</code> , <code>t[12]</code>
<code>l</code>	Vector of numeric values for the bandwidth parameter, <code>l</code> . The default uses the maximum distance of the modes (<code>t[m]</code>) from the time boundaries, $\max(\max(t) - t[m], t[m] - \min(t))$ split on 500 points on the logarithmic scale.
<code>k</code>	Vector of numeric values for the shape parameter, <code>k</code> . The default uses 50 splits of the values from 0.5 to 10 on the logarithmic scale.
<code>min.obs</code>	<p>Single value. The minimum observations/sum weight scores (SWS) that must be in the total window(s). The default uses the following steps.</p> <ol style="list-style-type: none"> 1. If there are more than one modes (peaks) in the data, then: $35 * (\text{the number of the unique modes})$ 2. If there is a single mode in the data, then: $\max(\pi * \sqrt{\text{length}(t)}, 35)$ <p>* <code>min.obs</code> must not be less than the total number of observations in the mode time(s). For example, it can not be less than the number of mutant animals in the IMPC application.</p> <p>** to function properly, <code>min.obs</code> should be less than the total number of observations</p> <p>*** <code>min.obs</code> is applied on the total number of observations on all windows NOT each single window</p> <p>**** if <code>weightORthreshold='weight'</code> then <code>min.obs</code> will be evaluated against SWS</p>
<code>direction</code>	Vector of two non-negative values. A non-negative vector of the form <code>c(Left,right)</code> , for example <code>c(1,1)</code> [default] or <code>c(0.5,0.5)</code> or <code>c(0,1)</code> . The first element specifies the speed of expansion of the window(s) from the left and the second value for the right expansion. Setting to <code>c(0,1)</code> and <code>c(1,0)</code> lead to right and left expansions of the window(s) respectively. Default <code>c(1,1)</code> that is the window(s) expand symmetrically from both sides.
<code>weightFUN</code>	Weight function. By default, a vector of weights called "ModelWeight" is passed to this function. See the examples.

residFun	Residual computation function. The default is 'resid()'. However, the user can define its own function. Note that the input of the function is the model object. The default is <code>residFun = function(object){resid(object)}</code>
predictFun	Similar to residFun but instead defines the 'predict()' function. The default is <code>predictFun = function(object){predict(object)}</code>
weightORthreshold	select between 'weight' (default) or 'threshold'. If set to 'weight' then the sum of weights (Sum Weight Score (SWS)) would be used as the total number of (active) observations in the window, otherwise, total number of weights (count of weights) that are greater than a threshold (see 'threshold' below) (<code>count weights >= threshold</code>) would be used for the total number of samples in the window (see 'threshold').
cdf	A cdf function preferably symmetric. The cdf function is used for the (window) weight generating function (WGF). The cdf function must have two parameters precisely a location such as mean and a scale. Standard cdf functions such as 'pnorm', 'pcauchy' and 'plogis' (default) can be used. For an example of custom made function we define uniform function as below: <i>puni</i> f0 = <i>function</i> (<i>x</i> , <i>mean</i> = .5, <i>sd</i> = <i>sqr</i> t(1/12)) <i>a</i> = <i>mean</i> - <i>sqr</i> t(3) * <i>sd</i> ; <i>b</i> = <i>mean</i> + <i>sqr</i> t(3)
check	Single integer in {0,1,2}: - check=1, the function selects the times (t) with more than one observations. Further, the function only selects the values with weights greater than the 'threshold' (see threshold below). Mostly useful in fitting linear mixed model - check=2 (default), the function only selects the values with weights greater than the 'threshold' parameter - check=0, disables all checks
sensitivity	(Default window selection criteria) Vector of four values (m, v, m*v, normality_test). For example (default) c(1,1,1,0) specifies the same weights for mean, variance, mean*variance interaction and zero weight for the test of normality (shapiro.test) in determining the optimal (final) window. We should stress that the window size is calculated by detecting the changes amongst the consecutive means (two sample t.test) and variances (two sample var.test) (as well as the normality of the first set) from each of predictFun() and residFun() combined together. For example, (m, v, m*v, normality_test) is calculated for predictFun() and the same for residFun(), then two means are combined under the 'sensitivity'[1]; and the same for variance, interactions and the normality.
pvalThreshold	Vector of four values. It would be used as the significant level for the mean, variation and normality tests (for more details see 'sensitivity' above). If all zero (default) or (all) negative (<code><= 0</code>) then the internal adaptive method (sensitivity - see above) would be used.
threshold	Single positive value. The minimum value for weights before removing the corresponded samples, given check=1 or check=2 and also in 'weightORthreshold'. Default <code>sqrt(.Machine\$double.eps)*10 ~ 10^-7</code>


```

zeroCompensation      Single non-negative value. Setting to any non-negative value would replace
                      all (weights =< zeroCompensation) with 'zeroCompensation'. Useful for al-
                      gorithms that have difficulties with zero. Default 0.

messages              Logical value. Set to TRUE (default) to see the errors and warnings

seed                  seed. Default NULL

simple.output          Logical flag. Setting to TRUE leads to not exporting the list of models for l and
                      k. Useful for preventing memory overflow. Default FALSE

debug                 Logical flag. Setting to TRUE will show some plots for the parameter selection
                      step. Useful for debugging. Default FALSE

...                   Other parameters that can be passed to the weightFUN()

```

Value

```

final.k, final.l      Final values for k and l

model.l, model.k, finalModel  List of models for l, k and the final model.

others                The input parameters such as x, y, t and so on

```

Author(s)

Hamed Haselimashhadi <hamedhm@ebi.ac.uk>

See Also

expWeight

Examples

```

#####
##### Example in the manuscript
#####
set.seed(1234)
par(mfrow = c(3, 1))
#####
# Simulating data
#####
n = 60
t = 1:n
sd = 1
m = n / 2
x = t
y = c(0 * x[t <= n / 3]
      x[t < 2 * n / 3 & t > n / 3] * 1
      0 * x[t >= 2 * n / 3]) + rnorm(n, 0, sd)
# True weights
w = weights = expWeight(
  t = t

```

```

k = 5
l = n/6
m = m
plot = 0
)
#####
# Fitting and plotting data and models
#####
l = lm(y ~ x, weights = w)
plot(
  x
  y
  ylim = c(min(y), max(y) * 1.5)
  col = t %in% seq(n / 3 + 1, 2 * n / 3 - 1) + 1,
  cex = 1.5
  pch = 16
  xlab = 'Time'
  main = 'Simulated data'
)
abline(v = x[c(n / 3 + 1, 2 * n / 3 - 1)],
       lty = 2,
       lwd = 4,
       col = 'gray')
abline(l, col = 2, lty = 2, lwd = 4)
abline(lm(y ~ x),
       col = 3,
       lty = 3,
       lwd = 4)
plot(
  t,
  w,
  type = 'b',
  main = 'True weights',
  ylab = 'Weight',
  xlab = 'Time'
)
#####
# Fitting the Windowing model
#####
r = SmoothWin(
  object = l,
  data = data.frame(y = y, x = x),
  t = t,
  m = m,
  min.obs = 4,
  debug = FALSE
)
#####
# Plot fitted (windowed) model
#####
plot(r, main = 'Estimated weights from WGF')

```

```
#####
##### Other examples
#####
# All examples import the Orthodont dataset from the nlme package
library(nlme)
# Sort the data on the time component (age)
Orthodont = Orthodont[order(Orthodont$age), ]
#####
# Modes
#####
mode = which(Orthodont$age %in% c(12))
#####
# Time component
#####
time = Orthodont$age
f = formula(distance ~ Sex)

#####
##### Examples #####
#####
### Example 1. Linear model
#####
# Method 1 (recommanded)
#####
lm = do.call('lm', list(formula = f, data = Orthodont))
rm(f)

#####
# Method 2 (can cause error if you pass the formula to the lm function)
# lm = lm(distance ~ Sex, data = Orthodont)
#####

lm.result = SmoothWin(
  object = lm,
  data = Orthodont,
  t = time,
  m = mode,
  check = 0,
  weightFUN = function(x) {
    x
  },
  debug = TRUE
)
plot(
  lm.result,
  col = Orthodont$Sex,
  pch = as.integer(Orthodont$Sex),
  main = 'Simple liner model'
)

#####
#### Example 2. Linear Model Using Generalized Least Squares
```

```

# Method 1 (recommanded)
#####
f = formula(distance ~ Sex)
gls = do.call('gls', list(model = f, data = Orthodont))
rm(f)

#####
# Method 2 (can cause error if you pass the formula to the gls function)
# gls = gls(distance ~ Sex, data = Orthodont)
#####
gls.result = SmoothWin(
  object = gls,
  data = Orthodont,
  t = time,
  m = mode,
  check = 2,
  weightFUN = function(ignore.me) {
    varFixed(~ 1 / ModelWeight) #nlme package uses the inverse weights
  },
  debug = TRUE
)
plot(
  gls.result,
  col = Orthodont$Sex,
  pch = as.integer(Orthodont$Sex),
  main = 'Linear model using GLS'
)

#####
### Example 3. Linear mixed model
#####
# Method 1 (recommanded)
#####
fixed = formula(distance ~ Sex)
random = formula(~ 1 | Subject)
lme = do.call('lme', list(
  fixed = fixed,
  random = random,
  data = Orthodont
))
rm(fixed, random)

#####
# Method 2 (can cause error if you pass the formula to the lme function)
# lme = lme(fixed = distance ~ Sex, random=~1|Subject , data = Orthodont)
#####
lme.result = SmoothWin(
  object = lme,
  data = Orthodont,
  t = time,
  m = mode,
  # Remove zero weights as well as single observation dates
  check = 1,

```

```
weightFUN = function(ignore.me) {  
  varFixed(~ 1 / ModelWeight)  
},  
debug = TRUE  
)  
plot(  
  lme.result,  
  col = Orthodont$Sex,  
  pch = as.integer(Orthodont$Sex),  
  main = 'Linear mixed model'  
)
```