Package 'ExcessMass'

January 20, 2025

Type Package

Title Excess Mass Calculation and Plots

Version 1.0.1

Date 2022-06-13

Author Marc-Daniel Mildenberger

Maintainer Marc-Daniel Mildenberger <mildenberger.stat@web.de>

Description Implementation of a function which calculates the empirical excess mass for given \eqn{\lambda} and given maximal number of modes (excessm()). Offering powerful plot features to visualize empirical excess mass (exmplot()). This includes the possibility of drawing several plots (with different maximal number of modes / cut off values) in a single graph.

License LGPL

Suggests MASS

NeedsCompilation no

Repository CRAN

Date/Publication 2022-06-13 22:00:22 UTC

Contents

`excessm` *Excess Mass Function*

Description

Algorithm which calculates the empirical excess mass for a given λ and given maximal number of modes.

Usage

```
excessm(x, lambda, M = 1, UpToM = FALSE)
```

Arguments

`x`	data in form of a vector
`lambda`	λ of interest
`M`	maximal number of modes
`UpToM`	if true, the intervals for modes up to M are returned

Value

`intervals`	Matrix containing the empirical λ-clusters. Line i consists of information about the $i.-\lambda$-cluster, where $i \leq M$. First (Second) column contains the corresponding starting (ending) sorted observation index. Third (Fourth) column the starting (ending) value. In case $UpToM$ is true and $M > 1$, a list is returned with the intervals for all modes $\leq M$
`excess_mass`	returns a vector with excess masses, the $i.$ entry is the excess mass achieved with i modes

Note

Please note that an allowance for M modes does not necessarily result in M λ-clusters. Hence, the number of intervals returned can be smaller than M. In this case a warning will be displayed. The vector $excess_mass$ does have less than M entries.

Author(s)

Marc-Daniel Mildenberger <mildenberger.stat@web.de>, based on earlier code from Dr. Guenther Sawitzki <gs@statlab.uni-heidelberg.de>

References

Muller, D. W. and Sawitzki, G., 09.1991, "Excess Mass Estimates and Tests for Multimodality", *Journal of the American Statistical Association* **, Vol. 86, No. 415, pp. 738–746**, http://www.jstor.org/stable/2290406

See Also

exmplot, exmsilhouette, mexmsilhouette

Examples

```
library(MASS)
attach(geyser)

##calculating excess mass for duration of 'Old Faithful Geyser' for lambda=0.2 allowing for one mode
excessm(duration, lambda=0.2)

##same as above, but allowing for up to three modes
excessm(duration, lambda=0.2, M=3)

#returns the intervals for modes 1,2 and 3
excessm(duration, lambda=0.2, M=3, UpToM=TRUE)
```

ExcessMass *Excess Mass Calculation and Plots*

Description

Implementation of a function which calculates the empirical excess mass for a given λ and given maximal number of modes (excessm). Offering powerful plot features to visualize empirical excess mass (exmsilhouette). This includes the possibility of drawing several plots (with different maximal number of modes / cut off values) in a single graph. Furthermore, plotting the empirical excess mass against lambda is implemented (exmplot).

Details

Package:	ExcessMass
Type:	Package
Version:	1.0.1
Date:	2017-05-17
License:	GPL

Author(s)

Marc-Daniel Mildenberger <mildenberger.stat@web.de>, based on earlier code from Dr. Guenther Sawitzki <gs@statlab.uni-heidelberg.de>

References

Muller, D. W. and Sawitzki, G., 09.1991, "Excess Mass Estimates and Tests for Multimodality", *Journal of the American Statistical Association* **, Vol. 86, No. 415, pp. 738–746**, http://www.jstor.org/stable/2290406

Muller, D. W., 12.1992, "The Excess Mass Approach in Statistics", *Beitraege zur Statistik – StatLab Heidelberg*, http://archiv.ub.uni-heidelberg.de/volltextserver/21357/

Examples

```
library(MASS)
attach(geyser)
excessm(duration, lambda=0.2)

x <- rnorm(1000)
exmsilhouette(x, M=2, CutOff=0.5)

mexmsilhouette(duration, CutOff=c(1,2), steps=60)
```

`exmplot` *Excess Mass Lambda Plot*

Description

Produces an excess mass lambda plot and calculates the maximal excess mass difference achieved by allowing for an additional mode.

Usage

```
exmplot(xdata, M=1, CutOff=1, steps=50, Lambda=NULL)
```

Arguments

`xdata`	data in form of a vector
`M`	the maximal number of modes
`CutOff`	determines the cut off value and hence the level up to which the λs are calculated
`steps`	number of different λs which are calculated / drawn
`Lambda`	allows specifying an own vector of λs

Details

$CutOff$ should not be set too small or too large, as this results in meaningless graphs. The excess mass for several λs can be calculated by specifying the Lambda.

Value

An Excess Mass Lambda plot is produced. The lines in the plot are sorted by the maximal number of modes from left to right, due to the monotonicity of the excess mass in M.

`max_dist` The i. entry is the maximal distance of the excess mass by allowing for up to i+1 instead of i modes

`max_dist_Lambda` Shows the λ at which the maximal excess mass difference is achieved

Author(s)

Marc-Daniel Mildenberger <mildenberger.stat@web.de>, based on earlier code from Dr. Guenther Sawitzki <gs@statlab.uni-heidelberg.de>

References

Muller, D. W. and Sawitzki, G., 09.1991, "Excess Mass Estimates and Tests for Multimodality", *Journal of the American Statistical Association* , **Vol. 86, No. 415, pp. 738–746**, http://www.jstor.org/stable/2290406

See Also

excessm, exmsilhouette, mexmsilhouette

Examples

```
library(MASS)
attach(geyser)

##calculating the maximal excess mass difference for duration of 'Old Faithful Geyser' for M=3
exmplot(duration, M=3)

##Plotting the excess mass against lambda for modes 1-5,
##increase CutOff value, double the number of steps
exmplot(duration, M=5, CutOff=1.2, steps=100)

##Specifying Lambda
Lambda=seq(.0,0.5,0.005)
exmplot(duration, M=7, Lambda=Lambda)
```

---

exmsilhouette *(Single) Excess Mass Silhouette*

Description

Produces an excess mass plot and the corresponding numerical values if required.

Usage

```
exmsilhouette(xdata, M = 1, CutOff = 1,steps = 50,rug = TRUE,
Lambda = NULL,col = FALSE,rdata = FALSE,label = TRUE)
```

Arguments

xdata	data in form of vector
M	the maximal number of modes
CutOff	determines the cut off value and hence the appearance of the graph
steps	number of different λs which are calculated / drawn
rug	draws a rug plot at the bottom of the graph
Lambda	allows to specify an own vector of λs which are drawn in the graph
col	lines get colored in purple ($em > 0.75$), green ($0.75 \geq em > 0.5$), turquoise ($0.5 \geq em > 0.25$), blue ($0.25 \geq em > 0.05$) and black ($0.05 \geq em$)
rdata	a numerical output is returned
label	allows to reduce labeling

Details

$CutOff$ should not be set too small or too large, as this results in meaningless graphs. The excess mass for several λs can be calculated by specifying the Lambda.

Value

A plot is always produced. By setting $rdata = TRUE$ numerical results are returned in form of a two-dimensional list. The first argument specifies λ. This means that if $steps = 50$ with 1 (50) you get access to the numerical results for the smallest (largest) λ.

In case no λ-vector is used for each λ, the following information is displayed:

[, 1]	value of λ,
[, 2]	calculated λ-clusters
[, 3]	excess mass vector

The last two components are presented in the way known from the excess mass function. In case $Lambda$ was set manually the value of λ is not returned, as it is known.

Author(s)

Marc-Daniel Mildenberger <mildenberger.stat@web.de>, based on earlier code from Dr. Guenther Sawitzki <gs@statlab.uni-heidelberg.de>

References

Muller, D. W. and Sawitzki, G., 09.1991, "Excess Mass Estimates and Tests for Multimodality", *Journal of the American Statistical Association* **, Vol. 86, No. 415, pp. 738–746**, http://www.jstor.org/stable/2290406

See Also

excessm, mexmsilhouette, exmplot

Examples

```
library(MASS)
attach(geyser)

##Plot allowing for up to two modes and reduced CutOff value
exmsilhouette(duration, M=2, CutOff=1.25)

##Plot with twice the default number of steps, omitting rug plot,
##colorizing the graph and asking for numerical output
res <- exmsilhouette(duration, M=2, CutOff=1.25, steps=100, rug=FALSE, col=TRUE, rdata=TRUE)

##Specifying Lambda and requesting numerical output
L=seq(.01,0.25,0.005)
res <- exmsilhouette(duration, M=3, Lambda=L, col=TRUE, rdata=TRUE)
```

---

mexmsilhouette *(Multiple) Excess Mass Silhouette*

Description

Produces a graph with several excess mass plots allowing for different maximal numbers of modes/ cut off values.

Usage

```
mexmsilhouette(xdata, M = 1:3, CutOff = c(1,2,5), steps = 30,
Lambda = NULL, col = FALSE, rug = TRUE, rdata = FALSE)
```

Arguments

xdata	data in form of a vector
M	vector containing the max. number of modes
CutOff	vector which determines the cut off values and hence the appearance of the graph
steps	number of different λs which are calculated / drawn
Lambda	allows to specify an own vector of λs which is drawn in the graph
col	lines get colored in purple ($em > 0.75$), green ($0.75 \geq em > 0.5$), turquoise ($0.5 \geq em > 0.25$), blue ($0.25 \geq em > 0.05$) and black ($0.05 \geq em$)
rug	draws a rug plot at the bottom of the graph
rdata	a numerical output is returned

Details

$CutOff$ should not be set too small or too large, as this results in meaningless graphs.

Value

Always a graph with multiple plots is produced. Each column contains another maximal number of modes and each row another CutOff factor.

Setting rdata=TRUE numerical results are returned in form of a list. If the number of modes and the CutOff parameter contain just one element, the output of "mexmsilhouette" and "exmplot" are equal.

Otherwise we can distinguish between two cases. First $Lambda$ is not specified, hence the list is four-dimensional. The first element determines the CutOff value of the data by using the sorted CutOff vector (using the plot, this means the row in which the graph is shown). The second element specifies the maximal number of modes by using the sorted mode vector (again using the plot, this means the column of the plot). The third element selects the λ of the graph. For each plot and each λ, the following information is stored: the value of λ, the λ-clusters and the excess mass vector. Using the default setting $[2, 2, 5, 2]$ shows the λ-clusters of the fifth smallest λ of the $CutOff = 2 - M = 2$-plot.

If $Lambda$ is declared manually, the list is three-dimensional. Hence, the first argument denotes the maximal number of modes (the column of the graph). The second argument indicates the λ by the position held by it in the $Lambda$ vector. As in "exmplot" only two information are shown. The λ-clusters ($[, , 1]$) and the vector of excess mass ($[, , 2]$), as the value of λ is known.

Author(s)

Marc-Daniel Mildenberger <mildenberger.stat@web.de>, based on earlier code from Dr. Guenther Sawitzki <gs@statlab.uni-heidelberg.de>

References

Muller, D. W. and Sawitzki, G., 09.1991, "Excess Mass Estimates and Tests for Multimodality", *Journal of the American Statistical Association* **, Vol. 86, No. 415, pp. 738–746**, http://www.jstor.org/stable/2290406

See Also

excessm, exmplot, exmsilhouette

Examples

```
library(MASS)
attach(geyser)

##calculating excess mass plots for duration of 'Old Faithful Geyser',
##specifying CutOff and number of steps manually
mexmsilhouette(duration, CutOff=c(1,2), steps=60)

##Allowing for three different maximal number of modes
##and CutOff factors as well as color.
##The rug plot is omitted and numerical data is requested.
res=mexmsilhouette(duration, M=c(2,3,7), CutOff=c(0.8,1,2), col=TRUE, rug=FALSE, rdata=TRUE)

##Lambda is specified, color is set to true, numerical data is requested
```

```
L=seq(.01,.25,0.005)
res=mexmsilhouette(duration, M=c(2,3,4), Lambda=L, col=TRUE, rdata=TRUE)
```

`searchMaxLambda` *Searching For A Maximal Lambda*

Description

Function which gives a rough approximation of maximal λ.

Usage

```
searchMaxLambda(x, limcount = 4, step = 1.05, trylambda = 0.01)
```

Arguments

`x`	data in form of a vector
`limcount`	divided by the square root of the number of data points. The result determines the cut off value.
`step`	determines step size
`trylambda`	initial λ

Details

Excess mass is calculated for $trylambda$. In case the resulting excess mass is larger (smaller) than the cut off value, $trylambda$ is set as $trylambda * step$ (respectively $trylambda/step$) and excess mass is calculated again until it is smaller (larger) than the cut off value. The corresponding $trylambda$ is returned.

The approximation is done allowing only for one λ-cluster, as scans including more λ-clusters have high computational costs due to the recursive structure of the algorithm.

Value

$trylambda$ calculated as described in Details.

Author(s)

Marc-Daniel Mildenberger <mildenberger.stat@web.de>, based on earlier code from Dr. Guenther Sawitzki <gs@statlab.uni-heidelberg.de>

References

Muller, D. W. and Sawitzki, G., 09.1991, "Excess Mass Estimates and Tests for Multimodality", *Journal of the American Statistical Association* **, Vol. 86, No. 415, pp. 738–746**, http://www.jstor.org/stable/2290406

See Also

excessm, exmplot, exmsilhouette, mexmsilhouette

Examples

```
library(MASS)
attach(geyser)

#Calculating Lambda using standard settings
searchMaxLambda(duration)

#Calculating Lambda, reducing cut off value and step. Setting another initial lambda
searchMaxLambda(duration, limcount = 5, step = 1.01, trylambda = 1)
```

Index