Package 'ercv'

October 15, 2019

Type Package

Title Fitting Tails by the Empirical Residual Coefficient of Variation

Version 1.0.1

Date 2019-10-09

Encoding UTF-8

Author Joan del Castillo, David Moriña Soler and Isabel Serra

Maintainer Isabel Serra <iserra@crm.cat>

Description Provides a methodology simple and trustworthy for the analysis of extreme values and multiple threshold tests for a generalized Pareto distribution, together with an automatic threshold selection algorithm. See del Castillo, J, Daoudi, J and Lockhart, R (2014) <doi:10.1111/sjos.12037>.

Depends R (>= 3.5.0)

Suggests poweRlaw, evir

License GPL (>= 2)

NeedsCompilation no

LazyData true

Repository CRAN

Date/Publication 2019-10-15 15:30:02 UTC

R topics documented:

`ercv-package` *Empirical residual coefficient of variation*

Description

Fitting tails by the empirical residual coefficient of variation.

Details

Package:	ercv
Type:	Package
Version:	1.0.1
Date:	2019-09-19
License:	GPL version 2 or newer
LazyLoad:	yes

The package provides a methodology simple and trustworthy for the analysis of extreme values. The package contains functions for visualizing, fitting and validating the distribution of tails. Moreover, it also provides multiple threshold tests for a generalized Pareto distribution, together with an automatic threshold selection algorithm.

Author(s)

Joan del Castillo (Universitat Autònoma de Barcelona), David Moriña Soler (Catalan Institute of Oncology (ICO)-IDIBELL) and Isabel Serra (Centre de Recerca Matemàtica)

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, evicv, fitpot, ppot, qpot, tdata, thrselect, Tm

---

BIFP *EEMBC AutoBench suite (Benchmark 3)*

Description

This data corresponds to 1000 observations sampled from the third benchmark of the well-known suite for real-time systems EEMBC AutoBench suite (Poovey, 2007), including a number of of programs used in automotive embedded systems. It corresponds to the basic integer and floating point (BIFP) algorithm.

Usage

```
BIFP
```

Format

A numeric vector.

References

Abella J., Padilla, M.,del Castillo, J. & Cazorla, F. (2017). Measurement-Based Worst-Case Execution Time Estimation Using the Coefficient of Variation". ACM Transactions on Design Automation of Electronic Systems (TODAES), **22**(4).

Poovey, J. (2007). Characterization of the EEMBC Benchmark Suite. North Carolina State University.

---

bilbao *Bilbao waves data set*

Description

This data corresponds to the Bilbao waves data set, firstly analysed by Castillo and Hadi (1997) and in del Castillo and Serra (2015) from the MLE point of view.

Usage

```
bilbao
```

Format

A numeric vector.

References

Castillo, E. and Hadi, A. S. (1997). Fitting the Generalized Pareto Distribution to Data. Journal of the American Statistical Association, **92**, 1609-1620. del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

`ccdfplot` *Plot of complementary empirical distribution function and the complementary distribution function*

Description

Plot of complementary empirical distribution function of a sample and the complementary distribution function from peaks-over-threshold model.

Usage

```
ccdfplot(data, pars=NA, log="y", from=NA, ci=FALSE, main="Complementary cdf",
xlab="data", ylab="ccdf", ...)
```

Arguments

`data`	a numeric vector.
`pars`	a list with the set of parameters of peaks-over-threshold model.
`log`	a character string which contains `x` if the x axis is to be logarithmic, `y` if the y axis is to be logarithmic and `xy` or `yx` if both axes are to be logarithmic.
`from`	the origen of x-axis in the plot.
`ci`	should confidence bands be plotted. Defaults to `FALSE`.
`main`	an overall title for the plot.
`xlab`	horizontal axis label. Defaults to `data`.
`ylab`	vertical axis label. Defaults to `ccdf`.
`...`	usual graphic parameters.

Value

Plot of complementary empirical distribution function and the complementary distribution function.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, cvevi, cvplot, evicv, fitpot, ppot, qpot, tdata, thrselect, Tm

Examples

```
data(iFFT)
ccdfplot(iFFT)
```

---

cievi *Confidence interval for extreme value index*

Description

Confidence interval for extreme value index estimation by Tm method.

Usage

```
cievi(nextremes, evi=0, conf.level=0.90, m=10, nsim=100)
```

Arguments

nextremes	the number of upper extremes to be used.
evi	extreme value index. In particular, the shape parammeter of a generalized Pareto distribution.
conf.level	confidence level of the interval.
m	number of thresholds to do multiplicial test.
nsim	number of simulation.

Value

A numerical vector with two elements, containing the limits of the interval.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, evicv, fitpot, ppot, qpot, tdata, thrselect, Tm

Examples

```
cievi(70, evi=0)
```

---

cvevi *Coefficient of variation for a given extreme value index*

Description

The coefficient of variation for a given extreme value index in the generalized Pareto distribution.

Usage

```
cvevi(evi)
```

Arguments

evi	extreme value index. In particular, the shape parameter of a generalized Pareto distribution. It has to satisfy evi < 1/2.

Value

A numerical value containing the coefficient of variation for the given extreme value index.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, evicv, fitpot, ppot, qpot, tdata, thrselect, Tm

Examples

```
cvevi(-1)
```

---

cvplot *Exploratory empirical residual coefficient of variation*

Description

Exploratory empirical residual coefficient of variation for extreme value analysis.

Usage

```
cvplot(data, threshold = NA, nextremes = NA, omit=4, evi=0, main="CVplot",
       conf.level=0.90, xlab="Excluded sample size",
       ylab="Coefficient of variation", col="blue", ...)
```

Arguments

data	a numeric vector.
threshold	a threshold value (either this or nextremes must be given but not both).
nextremes	the number of upper extremes to be used (either this or threshold must be given but not both).
omit	the minimum required number of upper extremes for computing residual statistics.
evi	extreme value index. In particular, the shape parammeter of a generalized Pareto distribution.
main	an overall title for the plot.
conf.level	confidence level of the interval (defaults to 0.90).
xlab	horizontal axis label. Defaults to Excluded sample size.
ylab	vertical axis label. Defaults to Coefficient of variation.
col	plot color. Defaults to blue.
...	Usual graphic parameters.

Value

Plot of the empirical residual CV and confidence intervals.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, evicv, fitpot, ppot, qpot, tdata, thrselect, Tm

Examples

```
data("moby", package = "poweRlaw")
cvplot(moby, main="MobyDick")

data(iFFT)
cvplot(iFFT, threshold=median(iFFT), main="iFFT")
```

EURUSD *Euro/Dollar daily exchange rates*

Description

This data corresponds to the euro/dollar daily exchange rates between 1999 and 2016, including the financial crisis of 2007-2008, which has been generated from the package quantmod (Ryan, 2016).

Usage

```
EURUSD
```

Format

A data frame with 6575 rows and 1 column.

References

Ryan, J. A. (2016). quantmod: Quantitative Financial Modelling Framework. R package version 0.4-7. https://CRAN.R-project.org/package=quantmod

evicv *Extreme value index*

Description

The extreme value index for a given coefficient of variation in the generalized Pareto distribution.

Usage

```
evicv(cv)
```

Arguments

cv	coefficient of variation. It has to satisfy cv > 0.

Value

The extreme value index for a given coefficient of variation in the generalized Pareto distribution as a numerical value.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, fitpot, ppot, qpot, tdata, thrselect, Tm

Examples

```
evicv(2)
```

FFT *EEMBC AutoBench suite (Benchmark 2)*

Description

This data corresponds to 1000 observations sampled from the second benchmark of the well-known suite for real-time systems EEMBC AutoBench suite (Poovey, 2007), including a number of of programs used in automotive embedded systems. It corresponds to the fast fourier transform (FFT) algorithm.

Usage

```
FFT
```

Format

A numeric vector.

References

Abella J., Padilla, M.,del Castillo, J. & Cazorla, F. (2017). Measurement-Based Worst-Case Execution Time Estimation Using the Coefficient of Variation". ACM Transactions on Design Automation of Electronic Systems (TODAES), **22**(4).

Poovey, J. (2007). Characterization of the EEMBC Benchmark Suite. North Carolina State University.

fitpot *Fits peaks-over-threshold model of a sample*

Description

Fits peaks-over-threshold model of a sample.

Usage

```
fitpot(data, threshold=NA, nextremes=NA, evi=NA)
```

Arguments

`data`	a numeric vector.
`threshold`	a threshold value (either this or `nextremes` must be given but not both).
`nextremes`	the number of upper extremes to be used (either this or `threshold` must be given but not both).
`evi`	extreme value index. In particular, the shape parammeter of a generalized Pareto distribution.

Value

A `data.frame` with the following columns:

- evi extreme value index. In particular, the shape parammeter of a generalized Pareto distribution.
- psi the scale parameter of a generalized Pareto distribution.
- threshold a threshold value where peaks-over-threshold is applied.
- prob proportion of size of data corresponding to the upper extremes modelled with generalized pareto distribution.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

`ercv-package`, `cievi`, `ccdfplot`, `cvevi`, `cvplot`, `evicv`, `ppot`, `qpot`, `tdata`, `thrselect`, `Tm`

Examples

```
data("nidd.thresh", package = "evir")
fitpot(nidd.thresh)
```

---

iFFT *EEMBC AutoBench suite (Benchmark 1)*

Description

This data corresponds to 1000 observations sampled from the first benchmark of the well-known suite for real-time systems EEMBC AutoBench suite (Poovey, 2007), including a number of of programs used in automotive embedded systems. It corresponds to the inverse fast fourier transform (iFFT) algorithm.

Usage

```
iFFT
```

Format

A numeric vector.

References

Abella J., Padilla, M.,del Castillo, J. & Cazorla, F. (2017). Measurement-Based Worst-Case Execution Time Estimation Using the Coefficient of Variation". ACM Transactions on Design Automation of Electronic Systems (TODAES), **22**(4).

Poovey, J. (2007). Characterization of the EEMBC Benchmark Suite. North Carolina State University.

MA *EEMBC AutoBench suite (Benchmark 4)*

Description

This data corresponds to 1000 observations sampled from the fourth benchmark of the well-known suite for real-time systems EEMBC AutoBench suite (Poovey, 2007), including a number of of programs used in automotive embedded systems. It corresponds to the matrix arithmetic (MA) algorithm.

Usage

```
MA
```

Format

A numeric vector.

References

Abella J., Padilla, M.,del Castillo, J. & Cazorla, F. (2017). Measurement-Based Worst-Case Execution Time Estimation Using the Coefficient of Variation". ACM Transactions on Design Automation of Electronic Systems (TODAES), **22**(4).

Poovey, J. (2007). Characterization of the EEMBC Benchmark Suite. North Carolina State University.

ppot *Cumulative distribution function*

Description

Cumulative distribution function from the peaks-over-threshold model.

Usage

```
ppot(q, pars, lower.tail=TRUE, log.p=FALSE)
```

Arguments

q	vector of quantiles.
pars	a numeric vector with the set of parameters of peaks-over-threshold model. The names of the elements have to be evi, psi, threshold, prob.
lower.tail	logical; if TRUE (default), probabilities are $P[X \leq x]$ otherwise, $P[X > x]$.
log.p	logical; if TRUE probabilities are given as log(p).

Value

Cumulated probability function as a numerical value.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, evicv, fitpot, qpot, tdata, thrselect, Tm

Examples

```
ppot(1.9, c(evi=0.1, psi=0.2, threshold=0.3, prob=0.4), lower.tail=FALSE)

x<-runif(10000)
x<-c(x^-1,x)
pars<-fitpot(x,1)
ppot(10,pars$coeff,lower.tail=FALSE) #the true value is 0.5/10
```

qpot *Quantile function*

Description

Quantile function from the peaks-over-threshold model.

Usage

```
qpot(p, pars, lower.tail=TRUE, log.p=FALSE)
```

Arguments

p	vector of probabilities.
pars	a numeric vector with the set of parameters of peaks-over-threshold model. The names of the elements have to be evi, psi, threshold, prob.
lower.tail	logical; if TRUE (default), probabilities are $P[X \leq x]$ otherwise, $P[X > x]$.
log.p	logical; if TRUE probabilities are given as log(p).

Value

Quantile function as a numerical value.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, evicv, fitpot, ppot, tdata, thrselect, Tm

Examples

```
qpot(0.1, c(evi=0.1, psi=0.2, threshold=0.3, prob=0.4), lower.tail=FALSE)

x<-runif(10000)
x<-c(x^-1,x)
pars<-fitpot(x,1)
qpot(0.5/10,pars$coeff,lower.tail=FALSE) #the true value is 10
```

tdata *Transforms a heavy-tailed sampled to non-heavy tailed*

Description

Transformation of a sample with assumption of heavy-tail to a sample with non-heavy tail.

Usage

```
tdata(data, threshold = NA, nextremes = NA, sigma=NA)
```

Arguments

data	a numeric vector.
threshold	a threshold value (either this or nextremes must be given but not both).
nextremes	the number of upper extremes to be used (either this or threshold must be given but not both).
sigma	the scale parammeter divided by shape parameter in generalized Pareto distribution.

Value

The transformed data as a numerical vector.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, evicv, fitpot, ppot, qpot, thrselect, Tm

Examples

```
data("danish", package = "evir")
tdata(danish)
```

`thrselect` *Threshold selection algorithm*

Description

Threshold selection algorithm.

Usage

```
thrselect(data, threshold=NA, nextremes=NA, omit=16, evi=NA, m=10, nsim=100,
          conf.level=0.90, oprint=TRUE)
```

Arguments

`data`	a numeric vector.
`threshold`	a threshold value (either this or `nextremes` must be given but not both).
`nextremes`	the number of upper extremes to be used (either this or `threshold` must be given but not both).
`omit`	the minimum required number of upper extremes for computing residual statistics.
`evi`	extreme value index. In particular, the shape parammeter of a generalized Pareto distribution.
`m`	number of thresholds to do multiplicial test.
`nsim`	number of simulations.
`conf.level`	confidence level of the interval.
`oprint`	logical. If `TRUE` (default), the single solution is printed. In any case, the full solution is the output of the function.

Value

A list including two `data.frame` (solution and options). Each of the `data.frame` contains the following columns:

- m number of thresholds for testing tail index.
- nextremes number of thresholds for testing tail index.
- threshold the threshold value
- rcv residual coefficient of variation for selected threshold.
- cvopt optimal coefficient of variation for the tail.
- evi the corresponding tail index for optimal coefficient of variation if `evi` parameter is `NA`.
- tms the statistic of the tail index test.
- pvalue p-value associated to `tms`.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

ercv-package, cievi, ccdfplot, cvevi, cvplot, evicv, fitpot, ppot, qpot, tdata, Tm

Examples

```
data("nidd.thresh", package = "evir")
thrselect(nidd.thresh, nsim=500)
```

Tm *Multiple threshold test for a GPD*

Description

Multiple threshold test for a GPD.

Usage

```
Tm(data, threshold = NA, nextremes = NA, omit = 16, evi = NA, m = 10, nsim = 100)
```

Arguments

data	a numeric vector.
threshold	a threshold value (either this or nextremes must be given but not both).
nextremes	the number of upper extremes to be used (either this or threshold must be given but not both).
omit	the minimum required number of upper extremes for computing residual statistics.
evi	extreme value index. In particular, the shape parammeter of a generalized Pareto distribution.
m	number of thresholds to do multiplicial test.
nsim	number of simulations.

Value

A `data.frame` containing the following columns:

- nextremes the number of upper extremes to be used.
- cvopt optimal coefficient of variation for the tail.
- evi the corresponding tail index for optimal coefficient of variation if `evi` parameter is `NA`.
- tms the statistic of the tail index test.
- pvalue p-value associated to `tms`.

Author(s)

Joan del Castillo, David Moriña Soler and Isabel Serra

References

del Castillo, J. and Padilla, M. (2016). Modeling extreme values by the residual coefficient of variation. SORT Statist. Oper. Res. Trans. **40**(2), 303-320.

del Castillo, J. and Serra, I. (2015). Likelihood inference for Generalized Pareto Distribution. Computational Statistics and Data Analysis, **83**, 116-128.

del Castillo, J., Daoudi, J. and Lockhart, R. (2014). Methods to Distinguish Between Polynomial and Exponential Tails. Scandinavian Journal of Statistics, **41**, 382-393.

See Also

`ercv-package`, `cievi`, `ccdfplot`, `cvevi`, `cvplot`, `evicv`, `fitpot`, `ppot`, `qpot`, `tdata`, `thrselect`

Examples

```
data("nidd.thresh",package = "evir")
Tm(nidd.thresh,evi=0, nextremes = 75)
```

Index