Package ‘RFA’

January 2, 2012

Version 0.0-9

Date 2010-01-14

Title Regional Frequency Analysis

Author Mathieu Ribatet <ribatet@lyon.cemagref.fr>

Maintainer Mathieu Ribatet <mathieu.ribatet@epfl.ch>

Depends R (>= 1.8.0), tcltk

Description Several function to perform a Regional Frequency Analysis

Repository CRAN

Date/Publication 2010-01-14 11:58:11

R topics documented:

Fit the GP Distribution

Fitting a POT model through a GP distribution

Description

Maximum Likelihood, Unbiased Probability Weigthed Moments, Biased Probability Weighted Moments and Moments Estimator to fit Peaks Over a Threshold to a GP distribution.

Usage

```
fitgpd(data, threshold, method, ...)
```

Arguments

`data`	A numeric vector.
`threshold`	A numeric value giving the threshold for the GPD
`method`	A string giving the names of the estimator. It can be `'mle'`, `'moments'`, `'pwmu'` and `'pwmb'` for the maximum likelihood, moments, unbiased probability weighted moments, biased probability weigthed moments estimators respectively.
`...`	Other optional arguments to be passed to `gpdmoments`, `gpdpwmu`, `gpdpwmb` and `gpdmle` functions.

Value

This function returns a list with composants:

`estimate`	A vector containing the maximum likelihood estimates.
`std.err`	A vector containing the standard errors.
`fixed`	A vector containing the parameters of the model that have been held fixed.
`param`	A vector containing all parameters (optimized and fixed).
`deviance`	The deviance at the maximum likelihood estimates.
`corr`	The correlation matrix - for the `mle` method.
`convergence, counts, message`	Components taken from the list returned by `optim` - for the `mle` method.
`threshold`	The threshold passed to argument `threshold`.
`nhigh`	The number of exceedences.
`nat, pat`	The number and proportion of exceedences.
`data`	The data passed to the argument `data`.
`exceedances`	The exceedences, or the maxima of the clusters of exceedences.
`scale`	The scale parameter for the fitted generalized Pareto distribution.
`call`	The call of the current function.

Note

The Maximum Likelihood estimator is obtained through a sligthly modified version of Alec Stephenson's `fpot.norm` function in the `evd` package.

`Heterogeneity Statistics`

Compute 3 heterogeneity statistics.

Description

Compute 3 heterogeneity statistics based on the sample L-moments. It is usefull to test if a region could be considered as homogeneous or not, that is if all events from each site could came from only one distribution once normalized with the index flood.

Usage

```
heterogeneity(N.sim,N.site,size,param,Vsite)
```

Arguments

`N.sim`	Numeric. The number of stochastic region generated. It is recommended to set `N.sim` at list 500.
`N.site`	Numeric. The number of site within the region.
`size`	Numeric vector of length equal to `N.site`. The sample size of each site within the region.
`param`	Numeric vector of length 4. The parameter vector of the fitted regional Kappa distribution.
`Vsite`	Numeric vector of length 3. The statistic values for the region.

Details

The code first fits a 4-parameter Kappa distribution from the weigthed mean of the first 4 sample L-moments of each site. Then, it generates a large number of stochastics region with the same properties as the real one. Observations are sampled in the Kappa distribution fitted earlier.

For each stochastic region, 3 statistics are computed and the behaviour of these values are directly compared with the 'real' statistics. Of course, if they have the same behaviour, then the region should be considered as homogeneous.

The general form of the test statistics is

$$H = \frac{V - m}{\sigma}$$

where `V` is one of the 3 statistics `V1, V2, V3` defined later evaluated of the 'real' region, `m` and σ are the mean and standart deviation of the `N` statistics `V` evaluated on each stochastic region.

The 3 statistics V1, V2 and V3 are:

$$V1 = \left(\frac{\sum_{i=1}^{N} n_i \left(\tau_i - \tau_R \right)^2}{\sum_{i=1}^{N} n_i} \right)^{1/2}$$

$$V2 = \frac{\sum_{i=1}^{N} n_i \left\{ \left(\tau_i - \tau_R \right)^2 + \left(\tau_{3,i} - \tau_{3,R} \right)^2 \right\}^{1/2}}{\sum_{i=1}^{N} n_i}$$

$$V3 = \frac{\sum_{i=1}^{N} n_i \left\{ \left(\tau_{4,i} - \tau_{4,R} \right)^2 + \left(\tau_{3,i} - \tau_{3,R} \right)^2 \right\}^{1/2}}{\sum_{i=1}^{N} n_i}$$

where n_i is the sample size of site i, τ_i, tau_R are the at-site i and regional sample L-CV respectively, $\tau_{3,i}, \tau_{3,R}$ are the at-site i and regional sample L-Skewness respectively, $\tau_{4,i}, \tau_{4,R}$ are the at-site i and regional sample L-Kurtosis respectively and N is the number of stochastic region generated.

The last 2 statistics lack power to discriminate between homogeneous and heterogeneous region. They rarely yield values larger than 2 even for grossly heterogeneous regions.

For the first statistic, it is suggest that the region be regarded as *acceptably homogeneous* if H1 < 1, *possibly heterogeneous* if $1 \leq$ H1 < 2, *definitely heterogeneous* if H1 ≥ 2.

Value

The program print value of the three statistics H1, H2 and H3 and return a list of several not so usefull objects.

Author(s)

Ribatet Mathieu

References

Hosking, J. R. M. and Wallis, J. R. (1997) *Regional Frequency Analysis*. Cambridge University Press.

Examples

```
data(northCascades)
lmom <- c(1, 0.1103, 0.0279, 0.1366)
kappaParam <- kappalmom(lmom)
heterogeneity(500, 19, size = northCascades[,1],
kappaParam, c(0.0104, .0339, .0405))
##The heterogeneity statistics given by Hosking for this case
##study are H1 = 0.62, H2 = -1.49 and H3 = -2.37
##Taking into account sample variability, results should be
##consistent
```

`lmomplots` *Generate several L-moments plots*

Description

Produces several plot given the first 3 sample L-moments ratio.

Usage

```
lmomplots(lmom.reg, which = 1:3, ask = nb.fig < length(which) &&
dev.interactive(), pch = 15, draw.dist = TRUE)
```

Arguments

`lmom.reg`	A list. Typically, it is the result of function `lmomreg`.
`which`	A numeric vector which specifies which L-moments plot should be plotted. '1' for L-Skewness - L-CV, '2' for L-Kurtosis - L-CV and '3' for L-Skewness - L-Kurtosis plot. By default, the 3 plots are produced.
`ask`	Logical. If `TRUE` user is asked before a new plot is drawn.
`pch`	Numeric. Specifies the symbol to use for the 'regional point'. See `points`.
`draw.dist`	Logical. If set to `TRUE` - the default, then theoretical lines of some distribution are displayed in the L-Skewness / L-Kurtosis space.

Details

For the L-Skewness / L-Kurtosis plot, theoretical lines of Generalized Logistic, Generalized Extreme Value, Generalized Pareto, Log-Normal a,d Pearson Type III are displayed. OLB is the overall lower bound of τ_4 as a function of τ_3. The triangles U, N, G and E correspond respectively to the Uniform, Normal, Gumbel and Exponential cases.

Value

Produces several plots.

Author(s)

Ribatet Mathieu

lmomreg *L-moments and statistics within a region.*

Description

Compute sample L-moments of each site within a region once normalized with a specified index flood, regional sample L-moments and 3 statistics of interest.

Usage

```
lmomreg(sample.sites,index.flood=mean)
```

Arguments

`sample.sites`	A list giving data from each site.
`index.flood`	A function who computes the index flood from data. The default is to compute the sample mean.

Value

A list with 3 arguments. `'site'`: a matrix containing the first 4 sample L-moments of each site once normalized by the index flood, `'reg'`: the first 4 regional sample L-moments and `'V'`: 3 statistics of interest. For more information, take a look at the references.

Author(s)

Ribatet Mathieu

References

Hosking, J. R. M. and Wallis, J. R. (1997) *Regional Frequency Analysis*. Cambridge University Press.

Examples

```
data(maxWind)
## Not run: lmomreg(sample.site = maxWind)
```

`maxWind` *Data set to perform a Regional Frequency Analysis*

Description

The data set of Hosking LMOMENTS Package.

Usage

```
data(maxWind)
```

Format

A list. Each argument are at-site sample. There is 12 sites.

`northCascades` *Data set to perform a Regional Frequency Analysis*

Description

The data set of Hosking LMOMENTS Package.

Usage

```
data(northCascades)
```

Format

A dataframe which include the names of sites, sample size, and samples L-CV, L-Skewness and L-Kurtosis for each site.

`reganalysis` *Perform a whole Regional Frequency Analysis*

Description

Perform a whole Regional Frequency Analysis via the Index Flood model [Dalrymple, 1960]. The analysis is interactive and the end-user is asked for choices. It produces a test to detect if the region can been regarded as homogeneous or not, L-moment plots for graphical checks and estimation of the regional and at-site distribution.

Usage

```
reganalysis(data)
```

Arguments

`data` a list with each arguments correspond to of a site within the region plus argument 'record': a vector with the record for each site.

Value

A whole regional frequency analysis is performed. That is a heterogeneity test for the selected region - see function `heterogeneity`, L-moments plots - see function `lmomplots`, and estimation of the regional and at-site distribution with eventually return level plots - see function `regdist`, `locdist`.

Author(s)

Mathieu Ribatet

References

Dalrymple, T. (1960) *Flood Frequency Analysis*. Water Supply.

Examples

```
data(region)
## Not run: reganalysis(region)
```

---

regdist *Regional and at-site distribution*

Description

The regional and at-site distribution are estimated via the Index Flood model. This is achieved with the regional L-moments.

Usage

```
regdist(lmom, loc, main, xlab, ylab, xliminf, xlimsup, draw.site =
  TRUE, ...)
  locdist(param.reg, mu, data, main, xlab, ylab, xlimsup, index.flood =
mean, draw.data = TRUE, ...)
```

Arguments

`lmom`	A list. Typically, return of function `lmomreg`.
`param.reg`	Vector of length 3 giving the value of the parameter of the regional distribution.
`mu`	Numeric. The mean number of events per year.
`loc`	Optional numeric. If present, the location parameter is fixed to this value.
`data`	Vector of the at-site sample.

main, xlab, ylab, xliminf, xlimsup
: Optional. Specifies the title, labels for the x and y axis and the lower and upper bound for the x-axis.

index.flood
: A function who compute the index flood in functions of the at-site sample. The default is to compute the sample mean.

draw.site, draw.data
: Logical. Should adimensional distribution of all site and at-site sample be plotted. The default is to plot it.

...
: Optional parameters to be passed to the plot function.

Details

The Index Flood model [Dalrymple, 1960] states that all distribution within a homogeneous region are the same up to a normalisation by a constant - namely, the Index Flood. So, estimate the distribution of a specified site can be performed in two step:

- Estimation of the regional distribution.
- Estimation of the at-site distribution by multiplying the regional one with the at-site index flood.

The regional distribution is estimated with the mean of L-moments of each site weighted with the record length [Hosking and Wallis, 1997].

Value

The program return the parameter estimate of the regional and at-site distribution. Optionally, return level plots are displayed.

Author(s)

Mathieu Ribatet

References

Dalrymple, T. (1960) *Flood Frequency Analysis*. Water Supply Paper.

Hosking, J. R. M. and Wallis, J. R. (1997) *Regional Frequency Analysis*. Cambridge University Press.

---

region *Data set to perform a Regional Frequency Analysis*

Description

A typical homogeneous region.

Usage

```
data(region)
```

Format

A list. Each argument correspond to a sample site.

`samlmu` *Compute sample L-moments*

Description

Compute the sample L-moments - unbiased version.

Usage

```
samlmu(x, nmom = 4, sort.data = TRUE)
```

Arguments

`x`	a vector of data
`nmom`	a numeric value giving the number of sample L-moments to be computed
`sort.data`	a locical which specifies if the vector of data x should be sorted or not.

Value

This function returns a vector of length 'nmom' corresponding to the sample L-moments. Note that for orders greater or equal than 3 it is the L-moments ratio that is sample L-CV, sample L-Skewness, ...

Examples

```
x <- runif(50)
## Not run: samlmu(x, nmom = 5)
```

`Simulation of random number` *Generate random number and fit the 4 parameter Kappa distribution with L-moments.*

Description

Generate random number of the 4 parameter Kappa distribution. This distribution encompass the Generalized Extreme Value, Generalized Pareto and Generalized Logistic distributions.

Usage

```
rkappa(n, loc, scale, shape1, shape2)
rgpd(n, loc = 0, scale = 1, shape = 0)
rgev(n, loc = 0, scale = 1, shape = 0)
qgpd(p, loc = 0, scale = 1, shape = 0, lower.tail = TRUE)
kappalmom(lmom)
```

Arguments

`n`	Numeric. The number of pseudo-random number to be generated.
`p`	Numeric. The probability for which the quantile has to be computed.
`loc, scale, shape, shape1, shape2`	Numerics. The location, scale and shape parameters.
`lower.tail`	Logical. If `TRUE` the default, the quantile associated to probability `p` of non exceedance is computed. Otherwise, this is related to the probability of exceedance.
`lmom`	Numeric vector of length 4. The first 4 sample L-moments, that is the sample mean, L-CV, L-Skewness and L-Kurtosis.

Details

The 4-parameter Kappa distribution has cumulative distribution function

$$F(x) = \left[1 - shape2\left(1 + \frac{shape1\,(x - loc)}{scale}\right)^{-1/shape1}\right]^{1/shape2}$$

For `shape2` = -1, this is the Generalized Logistic distribution, for `shape2` = 0, the Generalized Extreme Value distribution and for `shape2` = 1, the Generalized Pareto distribution.

Function `kappalmom` uses sample L-moments to fit the Kappa distribution. Newtow-Raphson iteration is used to solve the equations that express τ_3 and τ_4 as functions of `shape1` et `shape2`. `loc` and `scale` are calculated as functions of λ_1, τ, `shape1` and `shape2`. To ensure a 1-1 relationship between parameters and L-moments, the parameter space is restricted. See reference for more details.

Value

The program return either a numeric vector containing estimates of the location, scale and shape parameter of the Kappa distribution either a fail flag informing a problem in the optimization algorithm.

`'1'` L-moments invalid

`'2'` (τ_3, τ_4) lies above the generalized-logistic line (suggests that l-moments are not consistent with any kappa distribution with `shape2` > -1)

`'3'` iteration failed to converge

`'4'` unable to make progress from current point in iteration

`'5'` iteration encountered numerical difficulties - overflow would have been likely to occur

`'6'` iteration for the shape parameters converged, but overflow would have occurred when calculating location and scale parameters.

Author(s)

Ribatet Mathieu and Hosking J. R. M. for the fortran code.

References

Hosking, J. R. M. (1994) *The four-parameter kappa distribution.* IBM Journal of Research and Development, **38**, 251-258

Index