Package ‘ptsuite’

April 2, 2019

Title  Tail Index Estimation for Power Law Distributions

Version  1.0.0

Description  Various estimation methods for the shape parameter of Pareto
distributed data. This package contains functions for various estimation
methods such as maximum likelihood
(Newman, 2005)<doi:10.1016/j.cities.2012.03.001>,
Hill's estimator (Hill, 1975)<doi:10.1214/aos/1176343247>,
least squares (Zaher et al., 2014)<doi:10.9734/BJMCS/2014/10890>,
method of moments (Rytgaard, 1990)<doi:10.2143/AST.20.2.2005443>,
percentiles (Bhatti et al., 2018)<doi:10.1371/journal.pone.0196456>,
and weighted least squares (Nair et al., 2019) to estimate the shape
parameter of Pareto distributed data. It also provides both a heuristic
method (Hubert et al., 2013)<doi:10.1016/j.csda.2012.07.011> and a
goodness of fit test
(Gulati and Shapiro, 2008)<doi:10.1007/978-0-8176-4619-6> for testing for
Pareto data as well as a method for generating Pareto distributed data.

Depends  R (>= 3.5.0)

License  GPL-3

LazyData  true

LinkingTo  Rcpp

Imports  Rcpp

RoxygenNote  6.1.1

Suggests  plotly

NeedsCompilation  yes

Author  Ranjiva Munasinghe [aut],
        Pathum Kossinna [cre, aut],
        Dovini Jayasinghe [aut],
        Dilanka Wijeratne [aut]

Maintainer  Pathum Kossinna <pathum@mindlanka.org>

Repository  CRAN

Date/Publication  2019-04-02 15:10:03 UTC
Description

The ptsuite package provides functions to estimate parameters of Type I Pareto data. The details of the methods used and the equations implemented are given in the vignette. The package contains the following functions:

Estimator Functions

- Maximum Likelihood Estimator `alpha_mle`
- Weighted Least Squares Estimator `alpha_wls`
- Hill's Estimator `alpha_hills`
- Method of Moments Estimator `alpha_moment`
- Method of Percentiles Estimator `alpha_percentile`
- Method of Modified Percentiles Estimator `alpha_modified_percentile`
- Method of Geometric Percentiles `alpha_geometric_percentile`
- Least Squares Estimator `alpha_ls`

Other Functions

- Generate Pareto Data `generate_pareto`
- Estimates from all estimators `generate_all_estimates`
- Q-Q Plot to test for Pareto Distribution `pareto_qq_test`
- Pareto Test `pareto_test`
alpha_geometric_percentile

Estimating the Shape Parameter by Geometric Method of Percentiles

Description

This function uses the Geometric Method of Percentiles to estimate the shape parameter of a given set of data. (Bhatti et al. 2018)

Usage

alpha_geometric_percentile(dat)

Arguments

dat vector of observations

Value

A list of the following form:

shape Estimate of the shape parameter of the data
scale Estimate of the scale parameter of the data (which is taken to be the minimum of the data)

References


References


Examples

```r
x <- generate_pareto(10000L, 5, 2)
alpha_geometric_percentile(x)
```

---

alpha_hills

Estimating the Shape Parameter by Hill's Estimator

Description

This function uses the Hill's Estimator to estimate the shape parameter of a given set of data. (Nair et al. 2019; Pokorna 2016; Hill 1975) It is especially useful when the data is known not to follow an exact Pareto distribution but the tail of the data does. Thus, the specification of \( k \), the \( k \)th largest observation, allows to specify the point from where Pareto-like behavior may be seen. It is also possible to specify the value at which the tail begins. When \( k=n \), the Hill’s Estimator returns the same estimate as alpha_mle with a warning notifying the user.

Usage

```r
alpha_hills(dat, k, value = FALSE)
```

Arguments

- **dat**: vector of observations
- **k**: number of observations / value equal to or greater than to consider for tail
- **value**: (TRUE/FALSE) indicating if the value which is specified in "k" (TRUE)

Value

A list of the following form:

- **shape**: Estimate of the shape parameter of the data
- **scale**: Estimate of the scale parameter of the data (which is taken to be the minimum of the data)

References


Examples

```r
x <- generate_pareto(10000, 5, 2)
alpha_hills(x, 400)
```

---

**alpha_ls**  
*Estimating the Shape Parameter by Method of Least Squares*

**Description**

This function uses the Method of Least Squares to estimate the shape parameter of a given set of data. (Zaher et al. 2014; Nair et al. 2019)

**Usage**

```r
alpha_ls(dat)
```

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dat</code></td>
<td>vector of observations</td>
</tr>
</tbody>
</table>

**Value**

A list of the following form:

- **shape**: Estimate of the shape parameter of the data
- **scale**: Estimate of the scale parameter of the data (which is taken to be the minimum of the data)

**References**


**Examples**

```r
x <- generate_pareto(10000, 5, 2)
alpha_ls(x)
```
alpha_mle

Estimating the Shape Parameter by Method of Maximum Likelihood (MLE)

Description

This function can be used to estimate the shape parameter using the Maximum Likelihood Estimator method (Newman 2005). It can be used to obtain biased and unbiased estimates of the shape and scale parameters as well as the confidence interval for the shape parameter for the biased estimates.

Usage

alpha_mle(dat, biased = TRUE, significance = NULL)

Arguments

dat       vector of observations
biased    TRUE/FALSE to indicate biased or unbiased estimates
significance  level of significance

Value

A list of the following form:

shape   Estimate of the shape parameter of the data
lower_bound Upper error bound of the estimate of shape
upper_bound Lower error bound of the estimate of shape
scale   Estimate of the scale parameter of the data (which is taken to be the minimum of the data)

References


Examples

```r
x <- generate_pareto(10000, 5, 2)
alpha_mle(x, TRUE, 0.05)

x <- generate_pareto(10000, 5, 2)
alpha_mle(x, FALSE)
```
alpha_modified_percentile

Estimating the Shape Parameter by Method of Modified Percentiles

Description
This function uses the Method of Modified Percentiles to estimate the shape parameter of a given set of data. (Bhatti et al. 2018)

Usage
alpha_modified_percentile(dat)

Arguments
dat vector of observations

Value
A list of the following form:

shape Estimate of the shape parameter of the data
scale Estimate of the scale parameter of the data (which is taken to be the minimum of the data)

References

Examples
x <- generate_pareto(10000, 5, 2)
alpha_modified_percentile(x)

alpha_moment

Estimating the Shape Parameter by Method of Moments

Description
This function uses the Method of Moments to estimate the shape parameter of a given set of data. (Rytgaard 1990) The method of moments is only accurate if \( \alpha \) (shape parameter) is greater than or equal to 1 (Brazauskas and Serfling 2000). This function issues a warning if it detects that \( \alpha \) may be less than 1.

Usage
alpha_moment(dat)
Arguments

dat            vector of observations

Value

A list of the following form:

shape  Estimate of the shape parameter of the data
scale  Estimate of the scale parameter of the data (which is taken to be the minimum of the data)

References


Examples

x <- generate_pareto(10000L)
alpha_percentile(x)

---

alpha_percentile  Estimating the Shape Parameter by Method of Percentiles

Description

This function uses the Method of Percentiles to estimate the shape parameter of a given set of data.
(Bhatti et al. 2018)

Usage

alpha_percentile(dat)

Arguments

dat            vector of observations

Value

A list of the following form:

shape  Estimate of the shape parameter of the data
scale  Estimate of the scale parameter of the data (which is taken to be the minimum of the data)
References


Examples

```r
x <- generate_pareto(10000L, 5, 2)
alpha_percentile(x)
```

---

**alpha_wls**

*Estimating the Shape Parameter by Weighted Least Squares Method (WLS)*

**Description**

This function uses the Weighted Least Squares Method (WLS) to estimate the shape parameter of a given set of data. (Nair et al. 2019)

**Usage**

```r
alpha_wls(dat)
```

**Arguments**

- `dat` vector of observations

**Value**

A list of the following form:

- `shape` Estimate of the shape parameter of the data
- `scale` Estimate of the scale parameter of the data (which is taken to be the minimum of the data)

**References**


**Examples**

```r
x <- generate_pareto(10000, 5, 2)
alpha_percentile(x)
```
generate_all_estimates

Obtain estimates for Parameters of Pareto Data from all methods

Description

This function combines the results of all the methods (included in this package) provided to estimate the shape and scale parameters of the Pareto data and provides the results in a data frame. Hill’s Estimator is not used in this comparison as it discards a set of observations. We also note here that when considering the entire data set, Hill’s Estimate is equivalent to the MLE.

Usage

generate_all_estimates(dat)

Arguments

dat vector of observations

Value

Dataframe with the following columns:

Method.of.Estimation Name of the method used for estimation
Shape.Parameter Estimates of the shape parameter of the data
Scale.Parameter Estimates of the scale parameter of the data

Examples

x <- generate_pareto(10000L, 5, 2)
generate_all_estimates(x)

generate_pareto Generating data from a Pareto Distribution.

Description

This function is able to generate random Pareto distributed data with the specified shape and scale parameters. The function has been written to be similar in type to the popular runif and rexp type of functions for generating data from a particular distribution.

Usage

generate_pareto(sample_size, shape, scale)
Arguments

- sample_size: number of observations
- shape: shape parameter
- scale: scale parameter

Value

Vector of Pareto distributed data of sample size `sample_size` with shape parameter `shape` and scale parameter `scale`.

Examples

- `generate_pareto(10000L, 5, 2)`
- `generate_pareto(100, 15, 6)`

Description

This function can be used as a first step to identify whether the data is Pareto distributed before estimating the tail index. If most of the data points appear to be distributed along a line, it is possible that the data may be Pareto. Conversely, if most of the data are distributed non-linearly, then the data is most probably not Pareto distributed.

Usage

`pareto_qq_test(dat)`

Arguments

- `dat`: Data to be tested for Pareto distribution

Details

This function plots the quantiles of the standard exponential distribution on the x-axis and the log values of the provided data on the y-axis. If Pareto data was supplied, a log transformation of this data would result in an exponential distribution with mean $\alpha$. These data points would then show up on the QQ-plot as a line with slope $1/\alpha$.

The function makes use of the plotly package if available and installed or if not, defaults to the standard R plot.

Value

A Q-Q plot either using plotly if package is available or else a standard R plot.
Examples

```r
x <- generate_pareto(10000, 5, 2)
pareto_qq_test(x)
```

Description

The `pareto_test` function can be used to identify whether the data is Pareto distributed (Gulati and Shapiro 2008). The test generates a p-value corresponding to the actual distribution of the data and is tested for significance. In the case of Pareto data, the p-value should be greater than the pre-determined significance level (generally taken as 0.05).

Usage

```r
pareto_test(dat)
```

Arguments

- `dat` vector of observations

Value

A list of the following form:

- **p-value** p-value indicating significance of the test

References


Examples

```r
x <- generate_pareto(10000, 5, 2)
pareto_test(x)
```
Index

alpha_geometric_percentile, 2, 3
alpha_hills, 2, 4
alpha_ls, 2, 5
alpha_mle, 2, 6
alpha_modified_percentile, 2, 7
alpha_moment, 2, 7
alpha_percentile, 2, 8
alpha_wls, 2, 9

generate_all_estimates, 2, 10
generate_pareto, 2, 10

pareto_qq_test, 2, 11
pareto_test, 2, 12
ptsuite-package, 2