Package ‘GOF’

October 23, 2020

Title Goodness-of-fit tests with estimated parameters

Version 0.0.0.9000

Description This package aims to simplify Goodness-of-Fit tests when the distributions are not entirely specified prior to the analysis. For the most frequent distributions, functions are given to calculate test statistics and obtain critical values that account for shifts introduced by estimation of parameters on the same dataset used for testing GOF.

License MIT

URL https://github.com/alexanderlappe/GOF

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

R topics documented:

`A GOF Package` *A_GOF package*

Description

When dealing with a problem regarding Goodness-of-Fit, one must first check if the distribution under the null hypothesis is fully specified. If all parameters are known prior and need not be estimated, the standard EDF tests may be used. This package does not cover this case, so the reader is referred to the package 'goftest' (available on CRAN). If (some) parameters must be estimated, check if the distribution in question is listed in the table of contents. If so, the corresponding function can be used to estimate the unspecified parameters and subsequently calculate the desired test statistic. The null hypothesis of a sample belonging to the hypothesized distribution is rejected at some level if the test statistic does not lie in the interval given in the test output for that particular level. Some estimation functions and functions for calculating CVM and AD statistics are also documented and may be used, but are rather supplementary to the main test functions. Mathematical derivations of the methods used in this package are mostly given in D'Agostino, Stephens (1986).

Author(s)

Alexander Lappe

References

D'Agostino, Stephens (1986) Goodness-of-Fit Techniques

Choulakian, Stephens (2012) Goodness-of-Fit Tests for the Generalized Pareto Distribution

`A2` *Anderson-Darling test statistic*

Description

Computes the Anderson-Darling test statistic

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)(\log Z_i + \log(1 - Z_{n+1-i})),$$

for values $Z_1, \ldots, Z_n$ where $Z_i = F(x_i)$, F is the assumed CDF and x is the ordered sample.

Usage

```
A2(x)
```

Arguments

`x`	Vector of ordered real numbers

Value

Value of test statistic

Examples

```
A2(c(0.4, 0.5, 0.6))
```

`est_extreme` *Maximum Likelihood Estimator for a Gumbel distribution*

Description

Computes Maximum Likelihood Estimators for any unknown parameters of a Gumbel distribution with CDF

$$F(x) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right),$$

for a sample x.

Usage

```
est_extreme(x, param = c(NA, NA))
```

Arguments

`x`	Vector of real numbers
`param`	Vector of location α and scale $\beta > 0$ of the distribution. For the unknown parameter(s), put NA.

Value

Vector of (estimated) parameters

Examples

```
est_extreme(c(0.5, 0.6, 1.1), c(NA, NA))
```

`est_gamma` *Maximum Likelihood Estimator for a Gamma distribution*

Description

Computes Maximum Likelihood Estimators for any unknown parameters of a Gamma distribution with CDF

$$F(x) = \frac{1}{\beta\Gamma(m)}\left(\frac{x}{\beta}\right)^{m-1}$$

, for a sample x.

Usage

```
est_gamma(x, param = c(NA, NA))
```

Arguments

x Vector of real numbers

param Vector of scale $\beta > 0$ and shape $m > 0$ of the distribution. For the unknown parameter(s), put NA.

Value

Vector of (estimated) parameters

Examples

```
est_gamma(c(0.5, 0.6, 1.1), c(NA, NA))
#' @keywords internal
```

---

est_gpd *Maximum Likelihood Estimator for a Generalized Pareto distribution*

Description

Computes Maximum Likelihood Estimators for any unknown parameters of a Generalized Pareto distribution with CDF

$$F(x) = 1 - \left(1 - \frac{kx}{a}\right)^{\frac{1}{k}},$$

for a sample x.

Usage

```
est_gpd(x, param = c(NA, NA))
```

Arguments

x Vector of real numbers

param Vector of scale $a > 0$ and shape $k < 1$ of the distribution. For the unknown parameter(s), put NA.

Value

Vector of (estimated) parameters

Examples

```
est_gpd(c(0.5, 0.6, 1.1), c(NA, NA))
```

`est_pareto` *Maximum Likelihood Estimator for a Pareto distribution*

Description

Computes Maximum Likelihood Estimators for any unknown parameters of a Pareto distribution with CDF

$$F(x) = 1 - \left(\frac{x_m}{x}\right)^{\alpha},$$

for $(x > x_m)$, for a sample x.

Usage

```
est_pareto(x, param = c(NA, NA))
```

Arguments

`x`	Vector of real numbers
`param`	Vector of scale $x_m > 0$ and shape $\alpha > 0$ of the distribution. For the unknown parameter(s), put NA.

Value

Vector of (estimated) parameters

Examples

```
est_epareto(c(0.5, 0.6, 1.1), c(NA, NA))
```

`F_extreme` *Cumulative distribution function of a Gumbel distribution*

Description

Computes the CDF of a Gumbel distribution given as

$$F(x) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right).$$

Usage

```
F_extreme(x, location, scale)
```

Arguments

`x`	Vector of real numbers
`location`	location parameter α of the distribution
`scale`	scale parameter $\beta > 0$ of the distribution

Value

Vector of values of the distribution function

Examples

```
F_extreme(0.5, 1, 1)
```

`F_gpd` *Cumulative distribution function of a Generalized Pareto distribution*

Description

Computes the CDF of a generalized Pareto distribution given as

$$F(x) = 1 - \left(1 - \frac{kx}{a}\right)^{\frac{1}{k}}.$$

Usage

```
F_gpd(x, scale, shape)
```

Arguments

x	Vector of real numbers
scale	scale parameter a of the distribution
shape	shape parameter k of the distribution

Value

Vector of values of the distribution function at x

Examples

```
F_extreme(c(0.5, 0.6), 1, 1)
```

`gof_exp` *Goodness-of-fit Test for the exponential distribution*

Description

Computes either Cramér-von Mises or Anderson-Darling test for Goodness-of-Fit of the exponential distribution for a sample x. Unknown parameters are first estimated. The CDF is given by

$$F(x) = 1 - \exp(-\lambda x).$$

Usage

```
gof_exp(x, statistic)
```

Arguments

`x`	Vector of real numbers
`statistic`	Either 'cvm' or 'ad'. Specifies which of the two tests to compute

Value

Object of class gof_test

Examples

```
gof_exp(c(0.02, 0.03, 1.1), 'cvm')
```

`gof_extreme` *Goodness-of-fit Test for the Gumbel/Extreme value distribution*

Description

Computes either Cramér-von Mises or Anderson-Darling test for Goodness-of-Fit of the Gumbel/Extreme value distribution for a sample x. Unknown parameters are first estimated. The CDF is given by

$$F(x) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right).$$

Usage

```
gof_extreme(x, statistic, param = c(NA, NA))
```

Arguments

`x`	Vector of real numbers
`statistic`	Either 'cvm' or 'ad'. Specifies which of the two tests to compute
`param`	vector of location α and scale $\beta > 0$ of the distribution. For the unknown parameter(s), put NA.

Value

Object of class gof_test

Examples

```
gof_extremel(c(1, 2, 3), 'ad', c(NA, NA))
```

gof_gamma *Goodness-of-fit Test for the Gamma distribution*

Description

Computes either Cramér-von Mises or Anderson-Darling test for Goodness-of-Fit of the Gamma distribution for a sample x. Unknown parameters are first estimated. The CDF is given as

$$F(x) = \frac{1}{\beta\Gamma(m)} \left(\frac{x}{\beta}\right)^{m-1}.$$

Usage

```
gof_gamma(x, statistic, param = c(NA, NA))
```

Arguments

x	Vector of real numbers
statistic	Either 'cvm' or 'ad'. Specifies which of the two tests to compute
param	vector of scale $\beta > 0$ and shape $m > 0$ of the distribution. For the unknown parameter(s), put NA.

Value

Object of class gof_test

Examples

```
gof_gamma(c(1,2,3), 'ad', c(NA, NA))
```

gof_gpd *Goodness-of-fit Test for the Generalized Pareto distribution*

Description

Computes either Cramér-von Mises or Anderson-Darling test for Goodness-of-Fit of the GPD for a sample x. Unknown parameters are first estimated. The CDF is given as

$$F(x) = 1 - \left(1 - \frac{kx}{a}\right)^{\frac{1}{k}}.$$

Usage

```
gof_gpd(x, statistic, param = c(NA, NA))
```

Arguments

x	Vector of real numbers
statistic	Either 'cvm' or 'ad'. Specifies which of the two tests to compute
param	vector of scale $a > 0$ and shape $k < 1$ of the distribution. For the unknown parameter(s), put NA.

Value

Object of class gof_test

Examples

```
gof_gpd(c(1,2,3), 'ad', c(NA, NA))
```

`gof_normal` *Goodness-of-fit Test for the normal distribution*

Description

Computes either Cramér-von Mises or Anderson-Darling test for Goodness-of-Fit of the normal distribution for a sample x. Unknown parameters are first estimated. The procedure accounts for shifts of critical values due to parameter estimation.

Usage

```
gof_normal(x, statistic, param = c(NA, NA))
```

Arguments

`x`	Vector of real numbers
`statistic`	Either 'cvm' or 'ad'. Specifies which of the two tests to compute
`param`	vector of mean μ and variance σ^2 of the distribution. For the unknown parameter(s), put NA.

Value

Object of class gof_test

Examples

```
sim <- rnorm(100, 0, 1)
gof_normal(sim, 'cvm', c(NA, NA))
```

`gof_pareto` *Goodness-of-fit Test for the Pareto distribution*

Description

Computes either Cramér-von Mises or Anderson-Darling test for Goodness-of-Fit of the Pareto distribution for a sample x. The CDF is given as

$$F(x) = 1 - \left(\frac{x_m}{x}\right)^{\alpha},$$

$(x > x_m)$. The only case that is implemented is the one of known scale x_m and unknown shape α.

Usage

```
gof_pareto(x, param)
```

Arguments

x	Vector of real numbers
param	scale parameter x_m of the distribution
statistic	Either 'cvm' or 'ad'. Specifies which of the two tests to compute

Value

Object of class gof_test

Examples

```
sim <- unif(100, 1, 1)
gof_pareto(sim, 'cvm', 1)
```

---

gof_uniform *Goodness-of-fit Test for the uniform distribution*

Description

Computes the Cramér-von Mises test for Goodness-of-Fit of the uniform distribution for a sample x.

Usage

```
gof_uniform(x)
```

Arguments

x	Vector of real numbers

Value

Object of class gof_test

Examples

```
sim <- runif(100, 0, 1)
gof_uniform(sim)
```

gof_weibull	*Goodness-of-fit Test for the Weibull distribution*

Description

Computes either Cramér-von Mises or Anderson-Darling test for Goodness-of-Fit of the Weibull distribution for a sample x. Unknown parameters are first estimated. The CDF is given by

$$F(x) = 1 - \exp\left(\left(\frac{x}{\beta}\right)^m\right).$$

Usage

```
gof_weibull(x, statistic, param = c(NA, NA))
```

Arguments

`x`	Vector of real numbers
`statistic`	Either 'cvm' or 'ad'. Specifies which of the two tests to compute
`param`	vector of scale $\beta > 0$ and shape $m > 1$ of the distribution. For the unknown parameter(s), put NA.

Value

Object of class gof_test

Examples

```
gof_weibulll(c(1,2,3), 'cvm', c(NA, NA))
```

W2	*Cramer-von Mises test statistic*

Description

Computes the Cramér-von Mises test statistic

$$\sum_{i=1}^{n}\left(Z_i - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n},$$

for values $Z_1, \ldots, Z_n$ where $Z_i = F(x_i)$, F is the assumed CDF and x is the ordered sample.

Usage

```
W2(x)
```

Arguments

`x`	Vector of ordered real numbers

Value

Value of test statistic

Examples

```
W2(c(0.4,0.5,0.6))
```

Index