Package 'gnFit'

November 5, 2017

Type Package

Title Goodness of Fit Test for Continuous Distribution Functions

Version 0.1.0

Author Ali Saeb

Maintainer Ali Saeb <ali.saeb@gmail.com>

Description Computes the test statistic and p-value of the Cramer-von Mises and Anderson-Darling test for some continuous distribution functions proposed by Chen and Balakrishnan (1995) <http://asq.org/qic/display-item/index.html?item=11407>. In addition to our classic distribution functions here, we calculate the Goodness of Fit (GoF) test to dataset which follows the extreme value distribution function, without remembering the formula of distribution/density functions.

License GPL

Encoding UTF-8

LazyData true

Imports ismev, rmutil

NeedsCompilation no

Repository CRAN

Date/Publication 2017-11-05 13:39:36 UTC

R topics documented:

gnfit *Goodness of Fit Test for Continuous Distribution Functions*

Description

Computes the test statistic and p-value of the Cramer-von Mises and Anderson-Darling test for some continuous distribution functions proposed by Chen and Balakrishnan (1995). In addition to our classic distribution functions here, we calculate the Goodness of Fit (GoF) test to dataset which follows the extreme value distribution function, without remembering the formula of distribution/density functions.

Usage

```
gnfit(dat, dist, df = 3, pr = NULL, threshold = NULL)
```

Arguments

`dat`	A numeric vector of data values.
`dist`	A named distribution function in R, such as "norm", "t", "laplace", "logis", "gev", "gum", "gpd".
`df`	Degrees of freedom (> 2) for Student-t's distribution. This value is set to 3 by default.
`pr`	An object returned by maximum likelihood estimation of gev.fit, gum.fit or gpd.fit. It is also numeric vector giving the maximum likelihood estimation for the nonstationary model with location, scale and shape parameters for "gev" (generalized extreme value distribution), location and scale parameters in gumbel distribution or scale and shape parameters in case of "gpd" (generalized pareto distribution), resp.
`threshold`	The threshold is a single number. It is allocated by "gpd" with shape and scale parameters.

Details

To test $H_0 : X_1, \ldots, X_n$ is a random sample from a continuous distribution with cumulative distribution function $F(x;\theta)$, where the form of F is known but θ is unknown. We first esitmate θ by θ^* (for eg. maximum likelihood estimation method). Next, we compute $v_i = F(x_i, \theta^*)$, where the x_i's are in ascending order.

The Cramer-von Mises test statistic is

$$W^2 = \sum_{i=1}^{n} (u_i - \frac{(2i-1)}{2n})^2 + \frac{1}{12n},$$

and

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}((2i-1)\log(u_i) + (2n+1-2i)\log(1-u_i));$$

where, $u_i = \Phi((y_i - \bar{y}_i)/s_y)$, $y_i = \Phi^{-1}(v_i)$, and Φ is the standard normal CDF and Φ^{-1} its inverse. Modify W^2 and A^2 into

$$W^* = W^2(1 + 0.5/n),$$

and

$$A^* = A^2(1 + 0.75/n + 2.25/n^2).$$

The p-value is computed from the modified statistic W^* and A^* according to Table 4.9 in Stephens (1986).

Value

The output is an object of the class "htest" for the Cramer-von Mises and Anderson-Darling statistics corresponding to p-values.

References

Stephens (1986, ISBN:0824774876)

Chen and Balakrishnan (1995) <http://asq.org/qic/display-item/index.html?item=11407>

Marsaglia (2004) <doi:10.18637/jss.v009.i02>

See Also

The package of nortest for performing the Anderson-Darling test for normality. ADGofTest implementation of the Anderson-Darling goodness of fit test based on Marsaglia's (2004).

Examples

```
library(rmutil)
r <- rlaplace(1000, m = 1, s = 2)
gnfit(r, "laplace")

library(ismev)
pr <- c(-0.5, 1, 0.2)
r <- gevq(pr, runif(1000, 0, 1))
model <- gev.fit(r)$mle
gnfit(r, "gev", pr = model)

library(ismev)
r <- gum.q(runif(1000, 0, 1), -0.5, 1)
n <- length(r)
time <- matrix(1:n, ncol=1)
model <- gum.fit(r, ydat=time, mul=1)$mle
mle<-dim(2)
mle[1] <- model[1] + model[2] * (n+1)
mle[2] <- model[3]
gnfit(r, "gum", pr = mle)
```

Index