Package ‘gnFit’

June 7, 2018

Type Package

Title Goodness of Fit Test for Continuous Distribution Functions

Version 0.2.0

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. Calculates the Value at Risk (VaR) and Average VaR are another important risk factors which are estimated by using well-known distribution functions. Pflug and Romisch (2007, ISBN: 9812707409) is a good reference to study the properties of risk measures.

License GPL

Encoding UTF-8

LazyData true

Imports ismev, rmutil

NeedsCompilation no

Repository CRAN

Date/Publication 2018-06-07 19:00:29 UTC

R topics documented:

   gnfit ................................................................. 2
   rskFac ............................................................. 4

Index 6
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 = NULL, 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 NULL 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 \( \theta \) is unknown. We first estimate \( \theta \) by \( \theta^* \) (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 \( \Phi \) is the standard normal CDF and \( \Phi^{-1} \) 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**


**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)
library(ismev)
library(ismeth)

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

gnfit(r, "gev", pr = model)

gnfit(r, "gum", pr = mle)
```
**rskFac**

**Risk Factors**

**Description**

The Value at Risk (VaR) of level $\alpha$ ($\alpha$-quantile) of an event is a number attempting to summarize the risk of that event and define the worst expected loss of the event over a period of time. The Average VaR is another important measure of the risk at a given confidence level, which calculated by using the function of "rskFac".

**Usage**

```
rskFac(dat, alpha = 0.1, dist = \"norm\", df = NULL)
```

**Arguments**

- `dat` A numeric vector of object data.
- `alpha` Confidence level $0 < \alpha < 0.5$.
- `dist` A named of distribution function which should be fitted to data values. The distribution function is selected by the name of "laplace", "logis", "gum", "t" and "norm".
- `df` degrees of freedom from a specified distribution function.

**Details**

Suppose $X$ is random variable (rv) has distribution function (df) $F$. Given a confidence level $\alpha \in (0, 1)$, Value at Risk (VaR) of the underlying $X$ at the confidence level $\alpha$ is the smallest number $x$ such that the probability that the underlying $X$ exceeds $x$ is at least $1 - \alpha$. In other word, if $X$ is a rv with symmetric distribution function $F$ (e.g., the return value of a portfolio), then $VaR_\alpha$ is the negative of the $\alpha$ quantile, i.e.,

$$VaR_\alpha(X) = Q(\alpha) = \inf x \in \text{Real} : Pr(X \leq x) \leq \alpha.$$

where, $Q(.) = F^{-1}(.)$.

Since, the $VaR_\alpha(X)$ is the negative of $\alpha$ quantile in the left tail, $-VaR_{1-\alpha}(-X)$ is positive value of VaR in right tail.

The average $VaR_\alpha$, $(AVaR_\alpha)$ for $0 < \alpha \leq 1$ of $X$ is defined as

$$AVaR_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR(x) dx,$$

The AVaR is known under the names of conditional VaR (CVaR), tail VaR (TVaR) and expected shortfall.

Pflug and Romisch (2007, ISBN: 9812707409) shows the AVaR may be represented as the optimal value of the following optimization problem

$$AVaR_\alpha(X) = VaR_\alpha(X) - \frac{1}{\alpha} E((X - VaR_\alpha(X))^{-}).$$
where, \((y)^- = \min(y, 0)\). To approximate the integral, it is given by

\[
AVaR_\alpha(X) = VaR_\alpha(X) + \frac{1}{t\alpha} \sum_{i=1}^{t} \max(VaR_\alpha(X) - X, 0),
\]

where, \(t\) is number of observations. By considering the rv \(-X\), the \(-AVaR_{1-\alpha}\) in right tail is obtainable.

**Value**

The values of output are "VaR", "AVaR_n" and "AVaR_p" correspond to the VaR, Average VaR in left tail, Average VaR in right tail.

**References**


**Examples**

```r
library(rmutil)
r <- rlaplace(1000L, m = 1, s = 2)
rskFac(r, dist = "laplace", alpha = 0.1)
```
Index

gnfit, 2
rskFac, 4