

Package ‘gofest’

October 13, 2022

Type Package

Title Classical Goodness-of-Fit Tests for Univariate Distributions

Version 1.2-3

Date 2021-10-07

Depends R (>= 3.3)

Imports stats

Description Cramer-Von Mises and Anderson-Darling tests of goodness-of-fit for continuous univariate distributions, using efficient algorithms.

URL <https://github.com/baddstats/gofest>

BugReports <https://github.com/baddstats/gofest/issues>

License GPL (>= 2)

NeedsCompilation yes

Author Julian Faraway [aut],
George Marsaglia [aut],
John Marsaglia [aut],
Adrian Baddeley [aut, cre]

Maintainer Adrian Baddeley <Adrian.Baddeley@curtin.edu.au>

Repository CRAN

Date/Publication 2021-10-07 09:20:02 UTC

R topics documented:

| | |
|----------------|---|
| gofest-package | 2 |
| ad.test | 3 |
| cvm.test | 4 |
| pAD | 6 |
| pCvM | 7 |
| recogniseCdf | 8 |

| | |
|--------------|-----------|
| Index | 10 |
|--------------|-----------|

Description

Cramér-von Mises and Anderson-Darling tests of goodness-of-fit for continuous univariate distributions, using modern algorithms to compute the null distributions.

Details

The **gofest** package contains implementations of the classical Cramér-von Mises and Anderson-Darling tests of goodness-of-fit for continuous univariate distributions.

The Cramér-von Mises test is performed by `cvm.test`. The cumulative distribution function of the null distribution of the test statistic is computed by `pCvM` using the algorithm of Csörgő and Faraway (1996). The quantiles are computed by `qCvM` by root-finding.

The Anderson-Darling test is performed by `ad.test`. The cumulative distribution function of the null distribution of the test statistic is computed by `pAD` using the algorithm of Marsaglia and Marsaglia (2004). The quantiles are computed by `qAD` by root-finding.

By default, each test assumes that the parameters of the null distribution are known (a *simple* null hypothesis). If the parameters were estimated (calculated from the data) then the user should set `estimated=TRUE` which uses the method of Braun (1980) to adjust for the effect of estimating the parameters from the data.

Author(s)

Adrian Baddeley, Julian Faraway, John Marsaglia, George Marsaglia.

Maintainer: Adrian Baddeley <adrian.baddeley@uwa.edu.au>

References

Braun, H. (1980) A simple method for testing goodness-of-fit in the presence of nuisance parameters. *Journal of the Royal Statistical Society* **42**, 53–63.

Csörgő, S. and Faraway, J.J. (1996) The exact and asymptotic distributions of Cramér-von Mises statistics. *Journal of the Royal Statistical Society, Series B* **58**, 221–234.

Marsaglia, G. and Marsaglia, J. (2004) Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software* **9** (2), 1–5. February 2004. doi: [10.18637/jss.v009.i02](https://doi.org/10.18637/jss.v009.i02)

See Also

[ks.test](#)

Examples

```
x <- rnorm(30, mean=2, sd=1)
# default behaviour: parameters fixed: simple null hypothesis
cvm.test(x, "pnorm", mean=2, sd=1)
ad.test(x, "pnorm", mean=2, sd=1)
# parameters estimated: composite null hypothesis
mu <- mean(x)
sigma <- sd(x)
cvm.test(x, "pnorm", mean=mu, sd=sigma, estimated=TRUE)
ad.test(x, "pnorm", mean=mu, sd=sigma, estimated=TRUE)
```

ad.test

*Anderson-Darling Test of Goodness-of-Fit***Description**

Performs the Anderson-Darling test of goodness-of-fit to a specified continuous univariate probability distribution.

Usage

```
ad.test(x, null = "punif", ..., estimated=FALSE, nullname)
```

Arguments

| | |
|-----------|--|
| x | Numeric vector of data values. |
| null | A function, or a character string giving the name of a function, to compute the cumulative distribution function for the null distribution. |
| ... | Additional arguments for the cumulative distribution function. |
| estimated | Logical value indicating whether the parameters of the distribution were estimated using the data x (composite null hypothesis), or were fixed in advance (simple null hypothesis, the default). |
| nullname | Optional character string describing the null distribution. The default is "uniform distribution". |

Details

This command performs the Anderson-Darling test of goodness-of-fit to the distribution specified by the argument `null`. It is assumed that the values in `x` are independent and identically distributed random values, with some cumulative distribution function F . The null hypothesis is that F is the function specified by the argument `null`, while the alternative hypothesis is that F is some other function.

By default, the test assumes that all the parameters of the null distribution are known in advance (a *simple* null hypothesis). This test does not account for the effect of estimating the parameters.

If the parameters of the distribution were estimated (that is, if they were calculated from the same data x), then this should be indicated by setting the argument `estimated=TRUE`. The test will then use the method of Braun (1980) to adjust for the effect of parameter estimation.

Note that Braun's method involves randomly dividing the data into two equally-sized subsets, so the p -value is not exactly the same if the test is repeated. This technique is expected to work well when the number of observations in x is large.

Value

An object of class "htest" representing the result of the hypothesis test.

Author(s)

Original C code by George Marsaglia and John Marsaglia. R interface by Adrian Baddeley.

References

Anderson, T.W. and Darling, D.A. (1952) Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.

Anderson, T.W. and Darling, D.A. (1954) A test of goodness of fit. *Journal of the American Statistical Association* **49**, 765–769.

Braun, H. (1980) A simple method for testing goodness-of-fit in the presence of nuisance parameters. *Journal of the Royal Statistical Society* **42**, 53–63.

Marsaglia, G. and Marsaglia, J. (2004) Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software* **9** (2), 1–5. February 2004. doi: [10.18637/jss.v009.i02](https://doi.org/10.18637/jss.v009.i02)

See Also

[pAD](#) for the null distribution of the test statistic.

Examples

```
x <- rnorm(10, mean=2, sd=1)
ad.test(x, "pnorm", mean=2, sd=1)
ad.test(x, "pnorm", mean=mean(x), sd=sd(x), estimated=TRUE)
```

cvm.test

Cramer-Von Mises Test of Goodness-of-Fit

Description

Performs the Cramér-von Mises test of goodness-of-fit to a specified continuous univariate probability distribution.

Usage

```
cvm.test(x, null = "punif", ..., estimated=FALSE, nullname)
```

Arguments

| | |
|------------------------|---|
| <code>x</code> | Numeric vector of data values. |
| <code>null</code> | A function, or a character string giving the name of a function, to compute the cumulative distribution function for the null distribution. |
| <code>...</code> | Additional arguments for the cumulative distribution function. |
| <code>estimated</code> | Logical value indicating whether the parameters of the distribution were estimated using the data <code>x</code> (composite null hypothesis), or were fixed in advance (simple null hypothesis, the default). |
| <code>nullname</code> | Optional character string describing the null distribution. The default is "uniform distribution". |

Details

This command performs the Cramér-von Mises test of goodness-of-fit to the distribution specified by the argument `null`. It is assumed that the values in `x` are independent and identically distributed random values, with some cumulative distribution function F . The null hypothesis is that F is the function specified by the argument `null`, while the alternative hypothesis is that F is some other function.

By default, the test assumes that all the parameters of the null distribution are known in advance (a *simple* null hypothesis). This test does not account for the effect of estimating the parameters.

If the parameters of the distribution were estimated (that is, if they were calculated from the same data `x`), then this should be indicated by setting the argument `estimated=TRUE`. The test will then use the method of Braun (1980) to adjust for the effect of parameter estimation.

Note that Braun's method involves randomly dividing the data into two equally-sized subsets, so the p -value is not exactly the same if the test is repeated. This technique is expected to work well when the number of observations in `x` is large.

Value

An object of class "htest" representing the result of the hypothesis test.

Author(s)

Adrian Baddeley.

References

Braun, H. (1980) A simple method for testing goodness-of-fit in the presence of nuisance parameters. *Journal of the Royal Statistical Society* **42**, 53–63.

Csörgő, S. and Faraway, J.J. (1996) The exact and asymptotic distributions of Cramér-von Mises statistics. *Journal of the Royal Statistical Society, Series B* **58**, 221–234.

See Also

[pCVM](#) for the null distribution of the test statistic.

Examples

```
x <- rnorm(10, mean=2, sd=1)
cvm.test(x, "pnorm", mean=2, sd=1)
cvm.test(x, "pnorm", mean=mean(x), sd=sd(x), estimated=TRUE)
```

pAD

*Null Distribution of Anderson-Darling Test Statistic***Description**

pAD computes the cumulative distribution function, and qAD computes the quantile function, of the null distribution of the Anderson-Darling test statistic.

Usage

```
pAD(q, n = Inf, lower.tail = TRUE, fast=TRUE)
qAD(p, n = Inf, lower.tail = TRUE, fast=TRUE)
```

Arguments

| | |
|------------|--|
| q | Numeric vector of quantiles (values for which the cumulative probability is required). |
| p | Numeric vector of probabilities. |
| n | Integer. Sample size for the Anderson-Darling test. |
| lower.tail | Logical. If TRUE (the default), probabilities are $P(X \leq q)$, and otherwise they are $P(X > q)$. |
| fast | Logical value indicating whether to use a fast algorithm or a slower, more accurate algorithm, in the case $n=Inf$. |

Details

pAD uses the algorithms and C code described in Marsaglia and Marsaglia (2004).

qAD uses [uniroot](#) to find the quantiles.

The argument `fast` applies only when $n=Inf$ and determines whether the asymptotic distribution is approximated using the faster algorithm `adinf` (accurate to 4-5 places) or the slower algorithm `ADinf` (accurate to 11 places) described in Marsaglia and Marsaglia (2004).

Value

A numeric vector of the same length as `p` or `q`.

Author(s)

Original C code by G. and J. Marsaglia. R interface by Adrian Baddeley.

References

- Anderson, T.W. and Darling, D.A. (1952) Asymptotic theory of certain 'goodness-of-fit' criteria based on stochastic processes. *Annals of Mathematical Statistics* **23**, 193–212.
- Anderson, T.W. and Darling, D.A. (1954) A test of goodness of fit. *Journal of the American Statistical Association* **49**, 765–769.
- Marsaglia, G. and Marsaglia, J. (2004) Evaluating the Anderson-Darling Distribution. *Journal of Statistical Software* **9** (2), 1–5. February 2004. doi: [10.18637/jss.v009.i02](https://doi.org/10.18637/jss.v009.i02)

See Also

[ad.test](#)

Examples

```
pAD(1.1, n=5)
pAD(1.1)
pAD(1.1, fast=FALSE)

qAD(0.5, n=5)
qAD(0.5)
```

pCvM

Null Distribution of Cramer-von Mises Test Statistic

Description

pCvM computes the cumulative distribution function, and qCvM computes the quantile function, of the null distribution of the Cramér-von Mises test statistic.

Usage

```
pCvM(q, n = Inf, lower.tail = TRUE)
qCvM(p, n = Inf, lower.tail = TRUE)
```

Arguments

- | | |
|------------|---|
| q | Numeric vector of quantiles (values for which the cumulative probability is required). |
| p | Numeric vector of probabilities. |
| n | Integer. Sample size for the Cramér-von Mises test. |
| lower.tail | Logical. If TRUE (the default), probabilities are $P(X \leq q)$, and otherwise they are $P(X > q)$. |

Details

For finite n the cumulative distribution function is approximated by the first order expansion $V(x) + \psi_1(x)/n$, equation (1.8) of Csörgö and Faraway (1996).

qCvM uses [uniroot](#) to find the quantiles.

Value

A numeric vector of the same length as p or q .

Author(s)

Original Matlab code by Julian Faraway, translated to R by Adrian Baddeley.

References

Csörgö, S. and Faraway, J.J. (1996) The exact and asymptotic distributions of Cramér-von Mises statistics. *Journal of the Royal Statistical Society, Series B* **58**, 221–234.

See Also

[cvm.test](#)

Examples

```
pCvM(1.1, n=5)
pCvM(1.1)
```

```
qCvM(0.5, n=5)
qCvM(0.5)
```

recogniseCdf

Explanatory Name of Distribution Function

Description

Recognises many standard cumulative distribution functions and returns a string describing the distribution.

Usage

```
recogniseCdf(s="punif")
```

Arguments

s A single character string giving the *name* of an R function that calculates cumulative probabilities.

Details

The list of recognised distribution functions includes all those available in the **stats** package and in **gofitest**.

By convention, the name of a cumulative distribution function begins with the letter p. For example, `punif` is the cumulative distribution function of the uniform distribution.

The initial letter p can be omitted in the function `recogniseCdf`.

Value

Character string, or NULL if the name is not recognised.

Author(s)

Adrian Baddeley.

See Also

[pAD](#)

Examples

```
recogniseCdf("punif")
recogniseCdf("unif")
recogniseCdf("pt")
```

Index

* **distribution**

pAD, [6](#)

pCvM, [7](#)

recogniseCdf, [8](#)

* **htest**

ad.test, [3](#)

cvm.test, [4](#)

goftest-package, [2](#)

pAD, [6](#)

pCvM, [7](#)

* **package**

goftest-package, [2](#)

ad.test, [2](#), [3](#), [7](#)

cvm.test, [2](#), [4](#), [8](#)

goftest (goftest-package), [2](#)

goftest-package, [2](#)

ks.test, [2](#)

pAD, [2](#), [4](#), [6](#), [9](#)

pCvM, [2](#), [5](#), [7](#)

qAD, [2](#)

qAD (pAD), [6](#)

qCvM, [2](#)

qCvM (pCvM), [7](#)

recogniseCdf, [8](#)

uniroot, [6](#), [8](#)