Package ‘kolmim’

June 15, 2015

Version 1.0
Date 2015-06-14
Title An Improved Evaluation of Kolmogorov's Distribution
Author Luis Carvalho
Maintainer Luis Carvalho <lexcarvalho@gmail.com>
Description Provides an alternative, more efficient evaluation of extreme
probabilities of Kolmogorov's goodness-of-fit measure, Dn, when compared to
the original implementation of Wang, Marsaglia, and Tsang. These
probabilities are used in Kolmogorov-Smirnov tests when comparing two
samples.
License GPL-2
NeedsCompilation yes
Repository CRAN
Date/Publication 2015-06-15 17:38:00

R topics documented:

ks.test.imp ................................................................. 1
pkolm ................................................................. 3
pkolmim ............................................................. 4

Index 6

ks.test.imp Kolmogorov-Smirnov Tests

Description

Perform a one-sample two-sided exact Kolmogorov-Smirnov test, similarly to ks.test from package stats, but using an improved routine.
Usage

```
ks.test.imp(x, y, ...)  
```

Arguments

- **x**: a numeric vector of data values.
- **y**: either a numeric vector of data values, or a character string naming a cumulative distribution function or an actual cumulative distribution function such as `pnorm`. Only continuous CDFs are valid.
- **...**: parameters of the distribution specified (as a character string) by `y`.

Details

This routine is equivalent to `ks.test(x, y, ..., exact=TRUE)` but uses an improved method based on `pkolmim`. For more details about the arguments, please refer to the documentation for `ks.test`.

Value

A list with class "htest" containing the following components:

- **statistic**: the value of the test statistic.
- **p.value**: the p-value of the test.
- **alternative**: "two-sided".
- **method**: a character string indicating what type of test was performed.
- **data.name**: a character string giving the name(s) of the data.

Source

The two-sided one-sample distribution comes via Carvalho (2015).

References


See Also

- `pkolmim` for the cumulative distribution function of Kolmogorov’s goodness-of-fit measure.

Examples

```
x <- abs(rnorm(100))
p.kt <- ks.test(x, "pexp", exact = TRUE)$p  
p.ktimp <- ks.test.imp(x, "pexp")$p  
abs(p.kt - p.ktimp)
```

# compare execution times
```
x <- abs(rnorm(2000))
```
**pkolm**

Kolmogorov Dn Distribution

**Description**
Cumulative distribution function for Kolmogorov's goodness-of-fit measure.

**Usage**
pkolm(d, n)

**Arguments**
d the argument for the cumulative distribution function of Dn.
n the number of variates.

**Details**
Given an ordered set of n standard uniform variates, \( x_1 < \ldots < x_n \), Kolmogorov suggested \( D_n = \max [D_n^-, D_n^+] \) as a goodness-of-fit measure, where: \( D_n^- = \max_{i=1, \ldots, n} [x_i - (i - 1)/n] \) and \( D_n^+ = \max_{i=1, \ldots, n} [i/n - x_i] \).

pkolm provides the original algorithm proposed by Wang, Tsang, and Marsaglia (2003) to compute the cumulative distribution function \( K(n, d) = P(D_n < d) \). This routine is used by ks.test (package stats) for one-sample two-sided exact tests, and it is implemented in the C routine pkolmogorov2x. pkolm is a simple wrap around pkolmogorov2x.

**Value**
Returns \( K(n, d) = P(D_n < d) \).

**Source**

**References**

**See Also**
pkolmim for an improved routine to compute \( K(n, d) \), and ks.test for the Kolmogorov-Smirnov test.
Examples

```r
n <- 100
x <- 1:100 / 500
plot(x, sapply(x, function (x) pkolm(x, n)), type='l')

# Wang et al. approximation
s <- x ^ 2 * n
ps <- pmax(0, 1 - 2 * exp(-2.000071 / sqrt(n) + 1.409 / n) * s))
lines(x, ps, lty=2)
```

---

**pkolmim**

Kolmogorov Dn Distribution

**Description**

Cumulative distribution function for Kolmogorov's goodness-of-fit measure.

**Usage**

`pkolmim(d, n)`

**Arguments**

- `d` the argument for the cumulative distribution function of Dn.
- `n` the number of variates.

**Details**

Given an ordered set of `n` standard uniform variates, $x_1 < \ldots < x_n$. Kolmogorov suggested $D_n = \max\{D_n^-, D_n^+\}$ as a goodness-of-fit measure, where: $D_n^- = \max_{i=1,\ldots,n}[x_i - (i-1)/n]$ and $D_n^+ = \max_{i=1,\ldots,n}[i/n - x_i]$.

Wang, Tsang, and Marsaglia (2003) have proposed an algorithm to compute the cumulative distribution function $K(n, d) = P(D_n < d)$. `pkolmim` offers an improved implementation that uses less memory and should be more efficient for a range of arguments that are common in practice, while keeping the same precision.

The original algorithm of Wang, Tsang, and Marsaglia is implemented in the C routine `pkolmogorov2x` that is used by `ks.test` (package `stats`) for one-sample two-sided exact tests. Similarly, `pkolmim` is used by `ks.test.imp` in package `kolmim`.

**Value**

Returns $K(n, d) = P(D_n < d)$.

**Source**

The two-sided one-sample distribution comes via Carvalho (2015).
References


See Also

`ks.test.imp` for a Kolmogorov-Smirnov test similar to `ks.test` but that uses `pkmim` for one-sample two-sided exact tests.

Examples

```r
n <- 100
x <- 1:100 / n
plot(x, pkmim(x, n), type='l')

# Wang et al. approximation
s <- x^2 * n
ps <- pmax(0, 1 - 2 * exp(-(2.000071 + .331 / sqrt(n) + 1.409 / n) * s))
lines(x, ps, lty=2)
```
Index

*Topic htest
  ks.test.imp, 1

ks.test, 1–5
ks.test.imp, 1, 4, 5

pkolm, 3
pkolmim, 2, 4