Package ‘Hankel’

February 19, 2015

Version 0.0-1
Date 2014/03/08
Title Univariate non-parametric two-sample test based on empirical Hankel transforms.
Author Daniel Kolbe <Daniel-Kolbe@web.de>
Maintainer Daniel Kolbe <Daniel-Kolbe@web.de>
Description Provides an R routine for a Cramer-von Mises type two sample test which is based on empirical Hankel transforms of the non-negative sample variables. The test is non-parametric and not distribution free. The exact value of the test statistic for univariate data as well as the p-value and the critical value are computed.
License GPL (>= 2)
Repository CRAN
NeedsCompilation yes
Date/Publication 2014-05-18 11:55:19

R topics documented:

hankel.test ................................................................. 1

Index

hankel.test Hankel-Test for the univariate two-sample problem.

Description

Computes the Hankel test statistic and its empirically standardized version in the case that $\nu$ equals the exponential distribution (see details below). Optionally, simple computations of the p-value and the critical value for a selectable significance level and number of replicates are provided.
Usage

```r
hankel.test(x,y,standardized=FALSE,replicates=500,
calcCritVal=FALSE,calcPVal=TRUE,sigLevel=0.95,
probMeasure="exp",params=list(lambda=1.0))
```

Arguments

- **x**
  First set of observations as vector.

- **y**
  Second set of observations as vector.

- **standardized**
  Boolean variable which indicates whether the empirically standardized version of the test statistic is supposed to be applied. The default is `standardized=FALSE`.

- **probMeasure**
  The family of probability measures which is supposed to be applied. However, only the exponential distribution has been implemented so far.

- **params**
  List of parameter which specify the probability measure \( \nu \) which is applied (e.g. the rate parameter \( \lambda \) of the exponential distribution). The default is `list(lambda=1.0)`.

- **sigLevel**
  Significance level of the test. The default is `sigLevel=0.95`.

- **calcCritVal**
  Boolean variable which indicates whether the critical value of the test statistic for the given significance level is supposed to be computed. The default is `calcCritVal=FALSE`.

- **calcPVal**
  Boolean variable which indicates whether the p-value of the statistic corresponding to the given observations is supposed to be computed. The default is `calcPVal=TRUE`.

- **replicates**
  Number of bootstrap replicates for the computation of the critical value and the p-value. The default is `replicates=500`.

Details

The Hankel test statistic is of the Cramer-von Mises type and was proposed by L. Baringhaus, University of Hanover. It is given by

\[
T_{m,n} = \frac{mn}{m + n} \int_{R > 0} \left( \frac{1}{m} \sum_{i=1}^{m} J_0 \left( 2 \sqrt{tX_i} \right) - \frac{1}{n} \sum_{i=1}^{n} J_0 \left( 2 \sqrt{tY_i} \right) \right)^2 d \nu(t),
\]

where \( X_1, \ldots, X_m, Y_1, \ldots, Y_n \) are real nonnegative random variables, \( J_0 \) the Bessel function of the first kind of order zero and \( \nu \) some suitable probability measure on the nonnegative half-line.

In case that \( \nu \) equals the exponential distribution with parameter \( \lambda > 0 \), an alternative expression for \( T_{m,n} \), which is obtained by applying formula 6.615 of Gradstein and Ryshik (1981), is used:

\[
T_{m,n}(\lambda) = \frac{mn}{m + n} \int_{0}^{\infty} \left( \frac{1}{m} \sum_{i=1}^{m} J_0 \left( 2 \sqrt{tX_i} \right) - \frac{1}{n} \sum_{j=1}^{n} J_0 \left( 2 \sqrt{tY_j} \right) \right)^2 \lambda \exp(-\lambda t) dt
\]

\[
= \frac{mn}{m + n} \left[ \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} I_0 \left( 2 \sqrt{X_iX_j}/\lambda \right) \exp(-(X_i + X_j)/\lambda) \right]
\]
\[ + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} I_0(2\sqrt{Y_i Y_j}/\lambda) \exp(-Y_i + Y_j)/\lambda \]

\[ - \frac{2}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I_0(2\sqrt{X_i Y_j}/\lambda) \exp(-X_i + Y_j)/\lambda \]

where \( I_0 \) denotes the Bessel function of the third kind of order 0. This representation is used for the computation of the test statistic.

The empirically standardized version of the test is obtained by replacing the \( X_1, \ldots, X_m \) and \( Y_1, \ldots, Y_n \) by the empirically standardized variables \( U_1 = X_1/\eta_{m,n}, \ldots, U_m = X_m/\eta_{m,n} \) and \( V_1 = Y_1/\eta_{m,n}, \ldots, V_n = Y_n/\eta_{m,n} \), where \( \eta_{m,n} = \frac{1}{m+n} \left( \sum_{i=1}^{m} X_i + \sum_{j=1}^{n} Y_j \right) \), in the representation of \( T_{m,n} (\lambda) \). For further details see Baringhaus and Kolbe (2014).

Value

An object of class "hankeltest" is returned. The following components are retrievable:

- **samplesizes**: A list containing the samplesizes.
- **statistic**: Value of the Hankel test statistic for the given observations.
- **standardized**: Boolean value which indicates whether the standardized version of the test was applied.
- **probMeasure**: String value which indicates which family of probability measures was applied.
- **params**: List of parameter which specify the applied probability measure \( \nu \).
- **sigLevel**: Significance level of the test.
- **pValue**: Bootstrap estimation of the p-value of the test (if computed).
- **critValue**: Bootstrap estimation of the critical value (if computed).
- **replicates**: Number of bootstrap replicates taken for the computation of the critical value and the p-value.

Author(s)

Daniel Kolbe

References


Examples

# comparison of an uniform distribution with an weibull distribution
x <- runif(30)
y <- rweibull(50, 5, 0.75)
highlight(x, y)

# comparison of an uniform distribution with an weibull distribution
# in standardized case with parameter lambda=0.1 and replicates=1000
x <- runif(30)
y <- rweibull(50, 5, 0.75)
hankel.test(x, y, standardized=TRUE, params=list(lambda=0.1), replicates=1000)
Index

*Topic **htest**
  hankel.test, 1

*Topic **nonparametric**
  hankel.test, 1

hankel.test, 1

print.hankeltest(hankel.test), 1