Package 'fasano.franceschini.test'

September 3, 2021

Title Fasano-Franceschini Test: A 2-D Kolmogorov-Smirnov Two-Sample Test

Version 1.1.0

Description An implementation of the 2-D Kolmogorov-Smirnov (KS) two-sample test as defined by Fasano and Franceschini (Fasano and Franceschini 1987). The 'fasano.franceschini.test' package provides three improvements over the current 2-D KS test on the Comprehensive R Archive Network (CRAN): (i) the Fasano and Franceschini test has been shown to run in O(n^2) versus the Peacock implementation which runs in O(n^3); (ii) the package implements a procedure for handling ties in the data; and (iii) the package implements a parallelized permutation procedure for improved significance testing. Ultimately, the 'fasano.franceschini.test' package presents a robust statistical test for analyzing random samples defined in 2-dimensions.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.1

Imports stats, parallel, methods

Suggests tidyverse, knitr, rmarkdown, bookdown, testthat (>= 3.0.0)

URL https://github.com/nesscoder/fasano.franceschini.test

BugReports https://github.com/nesscoder/fasano.franceschini.test/issues

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Elan Ness-Cohn [aut, cre] (<https://orcid.org/0000-0002-3935-6667>), Rosemary Braun [ctb, ths] (<https://orcid.org/0000-0001-9668-9866>)

Maintainer Elan Ness-Cohn <elanness-cohn2017@u.northwestern.edu>

Repository CRAN

Date/Publication 2021-09-02 23:20:05 UTC

R topics documented:

`fasano.franceschini.test`
Fasano Franceschini Test

Description

Computes the 2-D Kolmogorov-Smirnov two-sample test as described by Fasano and Franceschini (1987).

Usage

```
fasano.franceschini.test(S1, S2, nBootstrap = 0, nPermute = 0, cores = 1)
```

Arguments

`S1`	a [n by 2] `data.frame` of x and y coordinates of sample 1
`S2`	a [n by 2] `data.frame` of x and y coordinates of sample 2
`nBootstrap`	a depreciated `numeric` argument defining the number of bootstrapped samples to be generated for computing the empirical p-value. nBootstrap is set to be replaced by nPermute in the next released version of the package.
`nPermute`	a `numeric` defining the number of permuted samples to be generated for computing the empirical p-value (note this procedure is slow and computationally expensive on the order of nPermute*O(n^2). Default is set to 0. If nPermute is 0, the Fasano Franceschini distributional approximation is used for defining the p-value. See Fasano and Franceschini test (1987) for details.
`cores`	a `numeric` defining the number of cores to use for processing

Details

Code adapted from Press, W. H., Teukolsky, S. A., Vetterling, W. T.,, Flannery, B. P. (2007). Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press. ISBN: 0521880688

Value

the 2-D ks statistic and p-value

References

- Fasano, G., Franceschini, A. (1987) doi: 10.1093/mnras/225.1.155. A multidimensional version of the Kolmogorov-Smirnov test. Monthly Notices of the Royal Astronomical Society 225:155-170.
- Peacock J.A. (1983) doi: 10.1093/mnras/202.3.615. Two-dimensional goodness-of-fit testing in astronomy. Monthly Notices of the Royal Astronomical Society 202:615-627.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T.,, Flannery, B. P. (2007). Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press. ISBN: 0521880688

Examples

```
#Underlying distributions are different
#set seed for reproducible example
set.seed(123)

#create 2-D samples with different underlying distributions
sample1Data <- data.frame(x = rnorm(n = 50, mean = 0, sd = 3), y = rnorm(n = 50,mean = 0, sd = 1))
sample2Data <- data.frame(x = rnorm(n = 50, mean = 0, sd = 1), y = rnorm(n = 50,mean = 0, sd = 3))

fasano.franceschini.test(S1 = sample1Data, S2 = sample2Data)


#Underlying distributions are the same
#set seed for reproducible example
set.seed(123)

#create 2-D samples with the same underlying distributions
sample1Data <- data.frame(x = rnorm(n = 50, mean = 0, sd = 1), y = rnorm(n = 50,mean = 0, sd = 1))
sample2Data <- data.frame(x = rnorm(n = 50, mean = 0, sd = 1), y = rnorm(n = 50,mean = 0, sd = 1))

fasano.franceschini.test(S1 = sample1Data, S2 = sample2Data)
```

`getDstat` *Get KS Stat*

Description

Loop through each row as(i.e. data point) and defines D stat as the largest difference between the quadfrequencies, looping though each point in the sample as the origin

Usage

```
getDstat(originSamples, S1, S2, cores = 1)
```

Arguments

originSamples	a [n by 2] data.frame of x and y coordinates that defines the origins data points.
S1	a [n by 2] data.frame of x and y coordinates of sample 1
S2	a [n by 2] data.frame of x and y coordinates of sample 2
cores	a numeric defining the number of cores to use of processing

Details

Code adapted from Press, W. H., Teukolsky, S. A., Vetterling, W. T.,, Flannery, B. P. (2007). Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press. ISBN: 0521880688

Value

a numeric defining the D stat with the largest difference between the quad frequencies, after checking each point as the origin

---

ksCDF *KS probability*

Description

p-value of getting the specified 2-D KS stat.

Usage

```
ksCDF(lambda)
```

Arguments

lambda	a numeric defining the difference in cumulative distribution function between two data sets

Details

Code adapted from Press, W. H., Teukolsky, S. A., Vetterling, W. T.,, Flannery, B. P. (2007). Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press. ISBN: 0521880688

Value

a numeric defining the p-value of observing the given 2-D KS stat

quadCount *Quad Count*

Description

Counts the frequency of points in the four quadrants - starting from the upper right going counter clockwise. Quadrants defined by the origin points x and y.

Usage

```
quadCount(x_origin, y_origin, x, y)
```

Arguments

x_origin	a numeric defining the x coordinate of the origin defining the 4 quadrants
y_origin	a numeric defining the y coordinate of the origin defining the 4 quadrants
x	a vector of numeric x coordinates
y	a vector of numeric y coordinates

Details

Code adapted from Press, W. H., Teukolsky, S. A., Vetterling, W. T.,, Flannery, B. P. (2007). Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press. ISBN: 0521880688

Value

a vector of frequencies of the number of points in each of the four quadrants defined by the origin point

Index