

Package ‘DiffXTables’

March 20, 2020

Title Pattern Heterogeneity via Distributional Differences Across Contingency Tables

Version 0.1.0

Date 2020-03-19

Author Ruby Sharma [aut],
Joe Song [aut, cre] (<<https://orcid.org/0000-0002-6883-6547>>)

Maintainer Joe Song <joemsong@cs.nmsu.edu>

Description Statistical hypothesis testing of pattern heterogeneity via differences in underlying distributions across two or more contingency tables. Three tests are included: the comparative chi-squared test (Song et al, 2014) <[doi:10.1093/nar/gku086](https://doi.org/10.1093/nar/gku086)> (Zhang et al, 2015) <[doi:10.1093/nar/gkv358](https://doi.org/10.1093/nar/gkv358)>, the Sharma-Song test, and the heterogeneity test. Under the null hypothesis that row and column variables are statistically independent and joint distributions are equal, their test statistics all follow an asymptotically chi-squared distribution. These options test for heterogeneous patterns that differ in either the first order (marginal) or the second order (joint distribution deviation from product of marginals). Second-order differences may reveal more fundamental changes than first-order differences across heterogeneous patterns.

Depends R (>= 3.0)

Suggests Ckmeans.1d.dp, FunChisq, knitr, rmarkdown, testthat

License LGPL (>= 3)

Encoding UTF-8

NeedsCompilation no

LazyData true

VignetteBuilder knitr

RoxygenNote 7.1.0

Repository CRAN

Date/Publication 2020-03-20 07:50:03 UTC

R topics documented:

cp.chisq.test	2
heterogeneity.test	4
sharma.song.test	5
simulate_diff_tables	7

Index	9
--------------	----------

cp.chisq.test	<i>Comparative Chi-Squared Test for Distributional Differences Across Contingency Tables</i>
---------------	--

Description

Across given contingency tables, the test admits any type of differences in either the joint or marginal distributions underlying the tables.

Usage

```
cp.chisq.test(
  tables, method=c("chisq", "nchisq", "default", "normalized"),
  log.p = FALSE
)
```

Arguments

tables	a list of at least two nonnegative matrices or data frames representing contingency tables of the same dimensions.
method	a character string to specify the method to compute the chi-squared statistic and its p-value. The default is "chisq". See Details. Note: "default" and "normalized" are deprecated.
log.p	logical; if TRUE, the p-value is given as $\log(p)$. Taking the log improves the accuracy when p-value is close to zero. The default is FALSE.

Details

The comparative chi-squared test determines whether the patterns underlying multiple contingency tables are heterogeneous. Its null test statistic is proved to asymptotically follow the chi-squared distribution (Song et al., 2014; Zhang et al., 2015). This test is different from the heterogeneity test (Zar, 2010).

Two methods are provided to compute the chi-squared statistic and its p-value. When method = "chisq" (or "default"), the p-value is computed using the chi-squared distribution; when method = "nchisq" (or "normalized") a normalized statistic is obtained by shifting and scaling the original chi-squared test statistic and a p-value is computed using the standard normal distribution (Box et al., 2005). The normalized test is more conservative on the degrees of freedom.

Value

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

statistic	chi-squared test statistic if method = "chisq" (equivalent to "default"), or normalized test statistic if method = "nchisq" (equivalent to "normalized").
parameter	degrees of freedom of the chi-squared statistic.
p.value	p-value of the comparative chi-squared test. By default, it is computed by the chi-squared distribution (method = "chisq" or "default"). If method = "nchisq" (or "normalized"), it is the p-value of the normalized chi-squared statistic using the standard normal distribution.

Author(s)

Joe Song

References

Box, G. E., Hunter, J. S. and Hunter, W. G. (2005) *Statistics for Experimenters: Design, Innovation and Discovery*, 2nd Ed., New York: Wiley-Interscience.

Song M., Zhang Y., Kataroff A. J., Edgar B. A. and Buttitta L. (2014) Hunting complex differential gene interaction patterns across molecular contexts. *Nucleic Acids Research* **42**(7), e57. Retrieved from <https://doi.org/10.1093/nar/gku086>

Zar, J. H. (2010) *Biostatistical Analysis*, 5th Ed., New Jersey: Prentice Hall.

Zhang, Y., Liu, Z. L. and Song, M. (2015) ChiNet uncovers rewired transcription subnetworks in tolerant yeast for advanced biofuels conversion. *Nucleic Acids Research* **43**(9), 4393–4407. Retrieved from <https://doi.org/10.1093/nar/gkv358>

See Also

The Sharma-Song test [sharma.song.test](#).

The heterogeneity test [heterogeneity.test](#).

Examples

```
# Two second-order differential tables:
tables <- list(
  matrix(c(4,0,0,
           0,4,0,
           0,0,4), nrow=3),
  matrix(c(0,4,4,
           4,0,4,
           4,4,0), nrow=3)
)
cp.chisq.test(tables)

# Three tables differ in the first-order but not second-order:
tables <- list(
  matrix(c(2, 4, 6, 8,
```

```

      3, 6, 9, 12,
      4, 8, 12, 16), nrow=4),
matrix(c( 2, 1, 3, 7,
         2, 1, 3, 7,
         10, 5, 15, 35), nrow=4),
matrix(c(40, 16, 72, 16,
         45, 18, 81, 18,
         25, 10, 45, 10), nrow=4)
)
cp.chisq.test(tables)

```

heterogeneity.test	<i>Heterogeneity Test for Distributional Differences Across Contingency Tables</i>
--------------------	--

Description

Across given contingency tables, the test admits any type of differences in either the joint or marginal distributions of the tables.

Usage

```
heterogeneity.test(tables)
```

Arguments

tables	a list of at least two non-negative matrices or data frames representing contingency tables of the same dimensions.
--------	---

Details

The heterogeneity test determines whether the patterns underlying multiple contingency tables are heterogeneous or differential. The chi-squared distribution is used for the null distribution of its test statistic (Zar, 2010).

Value

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

statistic	heterogeneity test statistic.
parameter	degrees of freedom of used for the null distribution of the heterogeneity test statistic.
p.value	p-value of the heterogeneity test, computed using the chi-squared distribution.

References

Zar, J. H. (2010) *Biostatistical Analysis*, 5th Ed., New Jersey: Prentice Hall.

See Also

The comparative chi-squared test [cp.chisq.test](#).

The Sharma-Song test [sharma.song.test](#).

Examples

```
# Two second-order differential tables:
tables <- list(
  matrix(c(4,0,0,
           0,4,0,
           0,0,4), nrow=3),
  matrix(c(0,4,4,
           4,0,4,
           4,4,0), nrow=3)
)
heterogeneity.test(tables)

# Three tables differ in the first-order but not second-order:
tables <- list(
  matrix(c(2, 4, 6, 8,
           3, 6, 9, 12,
           4, 8, 12, 16), nrow=4),
  matrix(c( 2, 1, 3, 7,
           2, 1, 3, 7,
           10, 5, 15, 35), nrow=4),
  matrix(c(40, 16, 72, 16,
           45, 18, 81, 18,
           25, 10, 45, 10), nrow=4)
)
heterogeneity.test(tables)
```

sharma.song.test

Sharma-Song Test for Second-Order Distributional Differences Across Contingency Tables

Description

Across given contingency tables, the test admits only second-order difference in the joint distributions underlying the tables.

Usage

```
sharma.song.test(tables)
```

Arguments

`tables` a list of at least two non-negative matrices or data frames representing contingency tables of the same dimensions.

Details

The Sharma-Song test determines whether the patterns underlying multiple input contingency tables are second-order differential. Its null test statistic is proved to asymptotically follow the chi-squared distribution.

Value

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

statistic	the Sharma-Song chi-squared test statistic.
parameter	degrees of freedom of the chi-squared test statistic.
p.value	p-value of the Sharma-Song test, computed using the chi-squared distribution.

Author(s)

Ruby Sharma and Joe Song

See Also

The comparative chi-squared test [cp.chisq.test](#).

The heterogeneity test [heterogeneity.test](#).

Examples

```
# Two second-order differential tables:
tables <- list(
  matrix(c(4,0,0,
           0,4,0,
           0,0,4), nrow=3),
  matrix(c(0,4,4,
           4,0,4,
           4,4,0), nrow=3)
)
sharma.song.test(tables)

# Three tables differ in the first-order but not second-order:
tables <- list(
  matrix(c(2, 4, 6, 8,
           3, 6, 9, 12,
           4, 8, 12, 16), nrow=4),
  matrix(c( 2, 1, 3, 7,
           2, 1, 3, 7,
           10, 5, 15, 35), nrow=4),
  matrix(c(40, 16, 72, 16,
           45, 18, 81, 18,
           25, 10, 45, 10), nrow=4)
)
sharma.song.test(tables)
```

simulate_diff_tables *Simulating Contingency Tables that are Differential in Joint Distribution*

Description

Generate contingency tables that are first-order, second-order or full-order differential in the joint distribution of row and column variables.

Usage

```
simulate_diff_tables(  
  K = 2, nrow = 3, ncol = 3, n = 100, B = 100,  
  type = c("second-order", "first-order", "full-order")  
)
```

Arguments

K	the number of tables that are differential. It must be an integer greater than one.
nrow	the number of rows for all tables to be generated. It must be an integer greater than one.
ncol	the number of columns for all tables to be generated. It must be an integer greater than one.
n	the sample size for each table to be generated. It must be a positive integer.
B	the number of iterations indicating the level of differentiability. It must be a positive integer. The greater the value, the stronger the differentiability across tables.
type	the type of differential tables to be generated. Options are "first-order", "second-order" (default), and "full-order". See Details.

Details

The function randomly generates contingency tables differential in the joint distribution of the row and column variables. Specifically, three types of differential contingency tables can be simulated:

First-order differential contingency tables only differ in row or column marginal distribution. Such tables are differential in joint distribution, but different from second-order differential tables.

Second-order differential contingency tables differ in joint distribution. The difference is not attributed to row or column marginal distributions.

Full-order differential contingency tables are tables that are both first-order and second-order differential.

The simulation starts with randomly generated probability tables where row and column variables are independent. The probability tables are modified to K tables such that they represent specific distributions that strictly satisfy the type requirement. Finally, contingency tables are generated using multinomial distribution using these probability tables parameters and the required sample size.

Value

A list containing the following components:

contingency.tables

a list of K contingency tables that are differential in joint distribution according to the type argument. They contain non-negative integers following the multinomial distribution with probability parameters from probability.tables.

probability.tables

a list of K tables representing randomly generated differential joint probabilities that reflect the specified type.

method

a string that specifies the type of the differential tables.

Author(s)

Ruby Sharma and Joe Song

See Also

Differential tables are simulated to evaluate the following tests for comparing contingency tables:

The Sharma-Song test [sharma.song.test](#)

The comparative chi-squared test [cp.chisq.test](#)

The heterogeneity test [heterogeneity.test](#)

The [simulate_tables](#) function in package **FunChisq** can generate a variety of tables.

Examples

```
# Three first-order differential tables:
res <- simulate_diff_tables(K = 3, nrow = 4, ncol = 3, n = 150, B = 200, type = "first-order")
print(res)

# Two second-order differential tables:
res <- simulate_diff_tables(K = 2, nrow = 2, ncol = 5, n = 100, B = 100, type = "second-order")
print(res)

# Four full-order differential tables:
res <- simulate_diff_tables(K = 4, nrow = 3, ncol = 4, n = 250, B = 200, type = "full-order")
print(res)
```


Index

*Topic **datagen**
 simulate_diff_tables, 7

*Topic **distribution**
 sharma.song.test, 5

*Topic **htest**
 cp.chisq.test, 2
 heterogeneity.test, 4
 sharma.song.test, 5

*Topic **nonparametric**
 cp.chisq.test, 2
 heterogeneity.test, 4
 sharma.song.test, 5

cp.chisq.test, 2, 5, 6, 8

heterogeneity.test, 3, 4, 6, 8

sharma.song.test, 3, 5, 5, 8

simulate_diff_tables, 7

simulate_tables, 8