Package 'FunChisq'

March 4, 2015

Type Package

Title Chi-Square Tests for Non-Parametric Functional Dependencies

Version 2.0.2

Date 2015-03-03

Author Yang Zhang, Hua Zhong, and Joe Song

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

Description Statistical hypothesis testing methods for non-parametric functional dependencies based on chi-square statistics. These tests were designed to reveal causality based on evidence from non-parametric functional dependencies. They include an asymptotic non-constant functional chi-square test, an exact functional test, and a comparative functional chi-square test. The normalized functional chi-square test was used by the Best Performer NMSU-SongLab in DREAM8 Breast Cancer Network Inference Challenges. These tests offer an advantage over regression analysis when a parametric functional form cannot be assumed.

License LGPL (>= 3)

Depends R (>= 3.0.0)

Imports Rcpp

LinkingTo BH, Rcpp, RcppClassic

Suggests Ckmeans.1d.dp, testthat

NeedsCompilation yes

URL http://www.cs.nmsu.edu/~joemsong/software,
http://www.cs.nmsu.edu/~joemsong/publications

Repository CRAN

Date/Publication 2015-03-04 14:57:09

R topics documented:

FunChisq-package   *Chi-Square Tests for Non-Parametric Functional Dependencies*

Description

Statistical hypothesis testing methods for non-parametric functional dependencies based on chi-square statistics. These tests were designed to reveal causality based on evidence from non-parametric functional dependencies. It implements an asymptotic non-constant functional chi-square test (Zhang and Song, 2013; Zhang, 2014), an exact functional test (Zhong, 2014), and a comparative functional chi-square test (Zhang, 2014). It requires data from two or more variables be formatted as a contingency table. Continuous variables must be discretized first using the Ckmeans.1d.dp function. The normalized functional chi-square test was used by the Best Performer NMSUSongLab in DREAM8 Breast Cancer Network Inference Challenges. These tests offer an advantage over regression analysis when a parametric form cannot be reliably assumed for the underlying function.

Details

Package:	FunChisq
Type:	Package
Version:	2.0.2
Date:	2015-03-03
License:	LGPL (>= 3)

Author(s)

Yang Zhang, Hua Zhong and Joe Song.

References

Zhang, Y. (2014) *Nonparametric Statistical Methods for Biological Network Inference.* Unpublished doctoral dissertation, Department of Computer Science, New Mexico State Unviersity, Las Cruces, USA

Zhang, Y. and Song, M. (2013) Deciphering interactions in causal networks without parametric assumptions. *arXiv Molecular Networks*, arXiv:1311.2707, http://arxiv.org/abs/1311.2707

Zhong, H. (2014) *An Exact and Fast Statistical Test for Nonparametric Functional Dependencies.* Unpublished M.S. thesis, Department of Computer Science, New Mexico State University, Las Cruces, USA.

See Also

The Ckmeans.1d.dp package.

cp.fun.chisq.test	*Comparative Chi-Square Test for Non-Parametric Functional Heterogeneity*

Description

Comparative functional chi-square tests on two or more contingency tables.

Usage

```
cp.fun.chisq.test(x, method="default")
```

Arguments

x	a list of at least 2 matrices representing contingency tables of the same dimensions.
method	a character string to specify the method to compute the functional chi-square statistic and its p-value. See Details.

Details

The comparative functional chi-square test determines whether the patterns underlying the contingency tables are heterogeneous in a functional way. Specifically, it evaluates whether the column variable is a changed function of the row variable across the contingency tables.

Two methods are provided to compute the functional chi-square statistic and its p-value. When method = "default", the p-value is computed using the chi-square distribution; when method = "normalized" a normalized functional chi-square is obtained by shifting and scaling the original chi-square 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	functional heterogeneity chi-square (by default), or normalized functional chi-square if method = "normalized".
parameter	degrees of freedom.
p.value	p-value of the comparative functional chi-square test. By default, it is computed by the chi-square distribution. If method = "normalized", it is the p-value of the normalized functional chi-square computed by the standard normal distribution.

Author(s)

Yang Zhang and Joe Song

References

Zhang, Y. (2014) *Nonparametric Statistical Methods for Biological Network Inference.* Unpublished doctoral dissertation, Department of Computer Science, New Mexico State Unviersity, Las Cruces, USA

Zhang, Y. and Song, M. (2013) Deciphering interactions in causal networks without parametric assumptions. *arXiv Molecular Networks*, arXiv:1311.2707. <http://arxiv.org/abs/1311.2707>

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

Examples

```
x <- matrix(c(4,0,4,0,4,0,1,0,1), 3)
y <- t(x)
z <- matrix(c(1,0,1,4,0,4,0,4,0), 3)
data <- list(x,y,z)
cp.fun.chisq.test(data)
cp.fun.chisq.test(data, method="normalized")
```

---

fun.chisq.test	*Chi-Square and Exact Tests for Non-Parametric Functional Dependencies*

Description

Asymptotic chi-square, normalized chi-square or exact tests on contingency tables to determine non-parametric non-constant functional depedencies of the column variable on the row variable.

Usage

```
fun.chisq.test(x, method="default")
```

Arguments

x   a matrix representing a contingency table. The row variable represents the independent variable or all unique combinations of multiple independent variables. The column variable is the dependent variable.

method   a character string to specify the method to compute the functional chi-square test statistic and its p-value. The options are `"default"`, `"normalized"`, or `"exact"`. See Details.

Details

The functional chi-square test determines whether the column variable is a function of the row variable in contingency table x (Zhang and Song, 2013; Zhang, 2014). There are three methods to compute the test statistic or its statistical significance:

When `method = "default"`, the test statistic is computed as described in (Zhang and Song, 2013; Zhang, 2014) and the p-value is computed using the chi-square distribution.

When `method = "normalized"`, a normalized functional chi-square obtained by shifting and scaling the original chi-square (Zhang and Song, 2013; Zhang, 2014) and a p-value is computed using the standard normal distribution (Box et al., 2005). The normalized test, more conservative on the degrees of freedom, was used by the Best Performer NMSUSongLab in DREAM8 Breast Cancer Network Inference Challenges.

When `method = "exact"`, an exact functional test (Zhong, 2014) is performed. It gives more precise p-value when the sample size is small. When the sample size of an input contingency table is large, the exact functional test and the functional chi-square test will return similar p-values. When the expected values for all entries in a contingency table are greater than 5, the asymptotic tests will perform similarly with the exact test. On a 2-by-2 contingency tables, Fisher's exact test (Fisher, 1922) will be applied. In this 2-by-2 case, the other two options for `method` are more desirable when the goal is to test functional depenencies. In addition, if the sample size is larger than 200 or the table size is larger than 5 by 5, the asymptotic functional chi-square test is forced to replace the exact test to for run time speed-up at a comparable precision.

The following is an example suitable for the exact functional test. An epigenetic indicator called CpG island methylator phenotype (CIMP) is strongly associated with liver cancers. Specimens (Shen et al, 2002) to study CIMP were collected and divided into three groups with different CIMP status: negative (no methylation gene), intermediate (1~2 methylated genes) and positive (>2 methylated genes). The following table represents the frequencies of observed tumor protein p53 mutations and CIMP status:

	CpG Island Methylator Phenotype (CIMP)		
	Negative	Intermediate	Positive
p53 mutation			
No	12	26	18
Yes	0	8	12

Example 4 below performs the exact functional test on this table.

Value

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

statistic	the functional chi-square statistic if `method = "default"` or `"exact"`, or the normalized functional chi-square statistic if `method = "normalized"`.
parameter	degrees of freedom for the functional chi-square statistic.
p.value	p-value of the functional chi-square test. If `method = "default"`, it is computed by an asymptotic chi-square distribution; if `method = "normalized"`, it is computed by the standard normal distribution; if `method = "exact"`, it is computed by an exact hypergeometric distribution.

Author(s)

Yang Zhang, Hua Zhong and Joe Song

References

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

Fisher, R. A. (1922). On the interpretation of chi-square from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, 85(1), 87-94.

Pearson, K. (1990). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, 50(302), 157-175.

Shen, L., Ahuja, N., Shen, Y., Habib, N. A., Toyota, M., Rashid, A., and Issa, J.-P. J. (2002). DNA methylation and environmental exposures in human hepatocellular carcinoma. *Journal of the National Cancer Institute*, 94(10), 755-761.

Zhang, Y. and Song, M. (2013) Deciphering interactions in causal networks without parametric assumptions. *arXiv Molecular Networks*, arXiv:1311.2707, http://arxiv.org/abs/1311.2707

Zhang, Y. (2014) *Nonparametric Statistical Methods for Biological Network Inference.* Unpublished doctoral dissertation, Department of Computer Science, New Mexico State Unviersity, Las Cruces, USA

Zhong, H. (2014) *An Exact and Fast Statistical Test for Nonparametric Functional Dependencies.* Unpublished M.S. thesis, Department of Computer Science, New Mexico State University, Las Cruces, USA.

Examples

```
# Example 1. Asymptotic functional chi-square test
x <- matrix(c(20,0,20,0,20,0,5,0,5), 3)
fun.chisq.test(x) # strong functional dependency
fun.chisq.test(t(x)) # weak functional dependency


# Example 2. Normalized functional chi-square test
x <- matrix(c(8,0,8,0,8,0,2,0,2), 3)
fun.chisq.test(x, method="normalized") # strong functional dependency
fun.chisq.test(t(x), method="normalized") # weak functional dependency


# Example 3. Exact functional chi-square test
x <- matrix(c(4,0,4,0,4,0,1,0,1), 3)
fun.chisq.test(x, method="exact") # strong functional dependency
fun.chisq.test(t(x), method="exact") # weak functional dependency


# Example 4. Exact functional chi-square test on a real data set
#           (Shen et al., 2002)
# x is a contingency table with row variable for p53 mutation and
#   column variable for CIMP
x <- matrix(c(12,26,18,0,8,12), nrow=2, ncol=3, byrow=TRUE)

# Test the functional dependency: p53 mutation -> CIMP
```

```
fun.chisq.test(x, method="exact")

# Test the functional dependency CIMP -> p53 mutation
fun.chisq.test(t(x), method="exact")
```

Index