Package 'EDOIF'

November 10, 2019

Title Empirical Distribution Ordering Inference Framework (EDOIF)

Version 1.0.0

Maintainer C. Amornbunchornvej <grandca@gmail.com>

Description

EDOIF is non-parametric framework based on Estimation Statistics principle. Its main purpose is to infer orders of empirical distributions from different categories base on a probability of finding a value in one distribution that greater than the expectation of another distribution. Given a set of ordered-pair of real-category values the framework is capable of 1) inferring orders of domination of categories and representing orders in the form of a graph; 2) estimating magnitude of difference between a pair of categories in forms of confidence intervals; and 3) visualizing domination orders and magnitudes of difference of categories.

License file LICENSE

URL https://github.com/DarkEyes/EDOIF

BugReports https://github.com/DarkEyes/EDOIF/issues

Language en-US

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0),
boot

Imports distr,
dabestr,
igraph,
ellipsis,
simpleboot,
ggplot2 (>= 3.0)

Suggests knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 6.1.1

R topics documented:

1

bootDiffmeanFunc   *bootDiffmeanFunc function*

Description

bootDiffmeanFunc is a support function for bootstrapping method. Its main task is to infer mean-difference confidence intervals of distributions for all categories except the first category in idx (idx[2],idx[3],...) minus a target category (idx[1]).

Usage

```
bootDiffmeanFunc(Group, Values, idx, reps, ci, methodType)
```

Arguments

Group	is a vector of categories of each real number in Values
Values	is a vector of real-number values
idx	is an order list of categories; idx[1] is a target category while others (idx[2],idx[3],...) are compared against idx[1] in order to compute mean-difference confidence intervals.
reps	is a number of time of sampling with replacement in a bootstrapping method.
ci	is a level of confidence interval inferred.
methodType	is a type of method for inferring confidence intervals. It is a parameter of two.boot function of simpleboot package.

Value

This function returns a list of mean-difference confidence intervals of categories idx[2],idx[3],... minus category idx[1].

result a list of objects that contains mean-difference confidence intervals of pairs of distributions. It contains mean-difference confidence intervals of categories idx[2],idx[3],... minus category idx[1].

---

checkSim3Res *checkSim3Res function*

Description

checkSim3Res is a support function for checking whether an adjacency matrix of inferred a dominant-distribution network adjMat is corrected w.r.t. generator SimNonNormalDist().

Usage

```
checkSim3Res(adjMat, flag)
```

Arguments

adjMat   is an adjacency matrix of inferred a dominant-distribution network.

flag   is a flag of matrix. It should be set only to shift the low of matrix for comparison.

Value

This function returns precision, recall, and F1-score of inferred adjacency matrix.

EDOIF *Empirical Distribution Ordering Inference Framework (EDOIF)*

Description

EDOIF is non-parametric framework based on Estimation Statistics principle. Its main purpose is to infer orders of empirical distributions from different categories base on a probability of finding a value in one distribution that greater than the expectation of another distribution.

Given a set of ordered-pair of real-category values the framework is capable of 1) inferring orders of domination of categories and representing orders in the form of a graph; 2) estimating magnitude of difference between a pair of categories in forms of confidence intervals; and 3) visualizing domination orders and magnitudes of difference of categories.

Usage

```
EDOIF(Values, Group, bootT, alpha, methodType)
```

Arguments

Values   is a vector of real-number values

Group   is a vector of categories of each real number in Values

bootT   is a number of times of sample with replacement for bootstrapping. The default is 1000. It must be above zero

alpha   is a significance level using in both confidence intervals and ordering inference it has the range [0,1]. The default is 0.05.

methodType   is an option for bootstrapping methods:either "perc" or "bca". The "perc" is the default option.

Value

This class constructor returns an object of EDOIF class.

obj an object of EDOIF class that contains the results of ordering inference that can be print in text mode (print(obj)) or graphic mode (plot(obj)).

The obj consists of the following variables

Values, Group	The main inputs of the framework. They are the double and character vectors respectively.
bootT, alpha, methodType	
	The number of bootstrapping, significance level, and bootstrapping method parameters.
sortedGroupList	
	A list of names of categories ascendingly ordered by their means.
sortedmeanList	A list of means of categories that are ascendingly ordered.
MegDiffList[[i]]	
	Mean difference confidence intervals and related information of all categories that have higher means than sortedGroupList[i] category. If methodType = "bca", then MegDiffList[[i]] is an object of dabestr class that can be used to create the Gardner-Altman plot.
confInvsList[i,]	
	A mean confidence interval of sortedGroupList[i] category. confInvsList[i,1] is a lower bound and confInvsList[i,2] is an upper bound.
adjMat[i,j]	An element of adjacency matrix: one if sortedGroupList[j] category dominates sortedGroupList[i] using Mann-Whitney test, otherwise zero.
pValMat[i,j]	A p-value of Mann-Whitney test for adjMat[i,j].
adjDiffMat[i,j]	
	A lower bound of confidence interval of mean difference for sortedGroupList[j] minus sortedGroupList[i] using methodType bootstrap.
adjBootMat[i,j]	
	One if adjDiffMat[i,j] is positive, otherwise, zero.

Author(s)

Chainarong Amornbunchornvej, <chai@ieee.org>

---

getConfInv   *getConfInv function*

Description

getConfInv is a support function for bootstrapping method. Its main purpose is to compute a mean confidence intervals of all distributions.

Usage

```
getConfInv(Values, Group, GroupList, bootT, alpha, methodType)
```

Arguments

Values	is a vector of real-number values
Group	is a vector of categories of each real number in Values
GroupList	is a list of names of categories ascendingly ordered by their means.
bootT	is a number of times of sample with replacement for bootstrapping. The default is 1000. It must be above zero
alpha	is a significance level using in both confidence intervals and ordering inference it has the range [0,1]. The default is 0.05.
methodType	is an option for bootstrapping methods:either "perc" or "bca". The "perc" is the default option.

Value

This function returns a list of mean confidence intervals.

confInvsList[i,]

> The mean confidence interval of sortedGroupList[i] category. confInvsList[i,1] is a lower bound and confInvsList[i,2] is an upper bound.

---

getDominantRADJ	*getDominantRADJ function*

Description

getDominantRADJ is a support function for inferring a dominant-distribution network using mean-difference confidence intervals.

Usage

```
getDominantRADJ(MegDiffList, methodType)
```

Arguments

MegDiffList	is a list of objects that contains mean-difference confidence intervals inferred by getMegDiffConfInv function.
methodType	is an option for bootstrapping methods:either "perc" or "bca".

Value

This function returns an adjacency matrix of a dominant-distribution network adjMat and the corresponding lower-bound of mean difference CIs adjDiffMat.

adjDiffMat[i,j]

> A lower bound of confidence interval of mean difference for j minus i using methodType bootstrap.

adjMat[i,j]   An element of adjacency matrix: One if adjDiffMat[i,j] is positive, otherwise, zero.

---

getMegDiffConfInv   *getMegDiffConfInv function*

Description

getMegDiffConfInv is a support function for bootstrapping method. Its main purpose is to compute a mean-difference confidence intervals between all pair of distributions.

Usage

```
getMegDiffConfInv(Values, Group, GroupList, bootT, alpha, methodType)
```

Arguments

Values	is a vector of real-number values
Group	is a vector of categories of each real number in Values
GroupList	is a list of names of categories ascendingly ordered by their means.
bootT	is a number of times of sample with replacement for bootstrapping. The default is 1000. It must be above zero
alpha	is a significance level using in both confidence intervals and ordering inference it has the range [0,1]. The default is 0.05.
methodType	is an option for bootstrapping methods:either "perc" or "bca". The "perc" is the default option.

Value

This function returns a list of mean-difference confidence intervals.

MegDiffList a list of objects that contains mean-difference confidence intervals of all possible pairs of distributions. It contains MegDiffList[[1]],...,MegDiffList[[length(GroupList)]].

The MegDiffList consists of the following variables

MegDiffList[[i]]

Mean-difference confidence intervals and related information of all categories that have higher means than sortedGroupList[i] category. If methodType = "bca", then MegDiffList[[i]] is an object of dabestr class that can be used to create the Gardner-Altman plot.

---

getOrder   *getOrder function*

Description

getOrder is a support function for inferring a linear order of categories ascendingly sorted by their means.

Usage

```
getOrder(Values, Group)
```

Arguments

Values	is a vector of real-number values
Group	is a vector of categories of each real number in Values

Value

This function returns an order list of categories.

sortedGroupList	
	The list of names of categories ascendingly ordered by their means.
sortedmeanList	The list of means of categories that are ascendingly ordered.

---

getttestDominantRADJ   *getttestDominantRADJ function*

Description

getttestDominantRADJ is a support function for inferring a dominant-distribution network using Student's t-test.

Usage

```
getttestDominantRADJ(Values, Group, GroupList, alpha)
```

Arguments

Values	is a vector of real-number values
Group	is a vector of categories of each real number in Values
GroupList	is a list of names of categories ascendingly ordered by their means.
alpha	is a significance level using in both confidence intervals and ordering inference it has the range [0,1].

Value

This function returns an adjacency matrix of a dominant-distribution network `adjMat` and the corresponding p-values of all category pairs.

adjMat[i,j]	An element of adjacency matrix: one if GroupList[j] category dominates GroupList[i] using Student's t-test, otherwise zero.
pValMat[i,j]	A p-value of Student's t-test for adjMat[i,j].

---

getWilcoxDominantRADJ   *getWilcoxDominantRADJ function*

Description

getWilcoxDominantRADJ is a support function for inferring a dominant-distribution network using Mann-Whitney (Wilcoxon) Test.

Usage

```
getWilcoxDominantRADJ(Values, Group, GroupList, alpha)
```

Arguments

Values	is a vector of real-number values
Group	is a vector of categories of each real number in Values
GroupList	is a list of names of categories ascendingly ordered by their means.
alpha	is a significance level using in both confidence intervals and ordering inference it has the range [0,1].

Value

This function returns an adjacency matrix of a dominant-distribution network adjMat. and the corresponding p-values of all category pairs.

adjMat[i,j]	An element of adjacency matrix: one if GroupList[j] category dominates GroupList[i] using Mann-Whitney test, otherwise zero.
pValMat[i,j]	A p-value of Mann-Whitney test for adjMat[i,j].

---

meanBoot   *meanBoot function*

Description

meanBoot is a support function for bootstrapping method. Its main purpose is to compute a mean of a given samples from data selected by indices.

Usage

```
meanBoot(data, indices)
```

Arguments

data	is a vector of real-number values
indices	is a vector of TRUE/FALSE indices. It allows boot to select samples.

Value

This function returns a mean of values in data that have values TRUE within indices.

---

plot.EDOIF   *plot.EDOIF function*

Description

plot.EDOIF is a support function for printing all plots of EDOIF framework: dominant-distribution network plot, mean CI plot, and mean-difference CI plot.

Usage

```
## S3 method for class 'EDOIF'
plot(x, ..., NList, options, fontSize)
```

Arguments

x	is an object of EDOIF class that contains the results of ordering inference.
...	Signature for S3 generic function.
NList	is a list of based categories users want to have in mean-difference CI plot.
options	is an option of reporting EDOIF plot(s): 0 for reporting all plots, 1 for mean-difference CI plot, 2 for mean CI plot, and 3 for dominant-distribution network plot.
fontSize	is a font size of text for all plots.

---

print.EDOIF   *print.EDOIF function*

Description

print.EDOIF is a support function for printing results of ordering inference in text.

Usage

```
## S3 method for class 'EDOIF'
print(x, ...)
```

Arguments

x	is an object of EDOIF class that contains the results of ordering inference.
...	Signature for S3 generic function.

---

printGraph   *printGraph function*

Description

printGraph is a support function for printing a dominant-distribution network from an adjacency matrix.

Usage

```
printGraph(sortedGroupList, confInvsList, sortedmeanList, adjMat)
```

Arguments

sortedGroupList

is a list of names of categories ascendingly ordered by their means.

confInvsList   is a list of mean confidence intervals of all categories.

sortedmeanList   is a list of means of categories that are ascendingly ordered.

adjMat   is an adjacency matrix of dominant-distribution network

---

SimMixDist   *SimMixDist function*

Description

SimMixDist is a support function for generating samples from mixture distribution. The main purpose of this function is to generate samples from non-normal distribution.

Usage

```
SimMixDist(nInv, mean, std, p1, p2)
```

Arguments

nInv   is a number of samples the function will generate.

mean   is a mean of a normal distribution part of mixture distribution.

std   is a standard deviation of a normal distribution part of mixture distribution.

p1   is a ratio of a normal distribution within a mixture distribution.

p2   is a ratio of a Cauchy distribution within a mixture distribution.

Value

This function returns a vector of samples V generated by a mixture distribution.

SimNonNormalDist	*SimNonNormalDist function*

Description

SimNonNormalDist is a support function for generating samples from mixture distribution. There are five categories. Each categories has `nInv` samples. Categories C1,C2,C3, and C4 are dominated by C5 but none of them dominate each other.

Usage

```
SimNonNormalDist(nInv, noisePer)
```

Arguments

`nInv`	is a number of samples the function will generate for each category.
`noisePer`	is ratio of uniform distribution within a mixture distribution. It is considered as a uniform noise that make an approach to hardly distinguish whether one distribution dominates another.

Details

The main purpose of this function is to generate samples that contains domination relation among categories.

Value

This function returns a vector of samples `Values` and their category `Group` generated by a mixture distribution.

`Values`	A vector of samples generated by a mixture distribution.
`Group`	A list of categories associated with `Values`.
`V1,...,V5`	Vectors of samples separated by categories.

Index