Package 'NPIstats'

February 16, 2021

Type Package

Title Nonparametric Predictive Inference

Version 0.1.0

Description An implementation of the Nonparametric Predictive Inference approach in R. It provides tools for quantifying uncertainty via lower and upper probabilities. It includes useful functions for pairwise and multiple comparisons: comparing two groups with and without terminated tails, selecting the best group, selecting the subset of best groups, selecting the subset including the best group.

License GPL-3

Depends R (>= 3.5.0)

Imports dplyr (>= 1.0.0)

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.1.1

Config/testthat/edition 3

NeedsCompilation no

Author Tahani Coolen-Maturi [cre, aut]

Maintainer Tahani Coolen-Maturi <tahani.maturi@durham.ac.uk>

Repository CRAN

Date/Publication 2021-02-16 09:50:05 UTC

R topics documented:

`best.pair` *NPI for comparing two groups*

Description

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X.

Usage

```
best.pair(X, Y)
```

Arguments

`X`	numeric vector of data values
`Y`	numeric vector of data values, to check if it is the best group

Value

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X.

References

F.P.A. Coolen (1996). Comparing two populations based on low stochastic structure assumptions. Statistics & Probability Letters 29, 297-305.

Examples

```
data(BreakdownTimes)
data2<-split(BreakdownTimes$times, BreakdownTimes$group)
# No terminated tails, complete data
best.pair(data2$X, data2$Y)
```

best.pair.tt *NPI for comparing two groups with terminated tails*

Description

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X. The information available consists of precise measurements of real-valued data only within a specific range, between the cut points, where the numbers of observations to the left and to the right of this range available.

Usage

```
best.pair.tt(X, Y, Lx = -Inf, Ux = Inf, Ly = -Inf, Uy = Inf)
```

Arguments

X	numeric vector of data values
Y	numeric vector of data values, to check if it is the best group
Lx	numeric value, lower cut point for group X, default set to -Inf
Ux	numeric value, upper cut point for group X, default set to Inf
Ly	numeric value, lower cut point for group Y, default set to -Inf
Uy	numeric value, lower cut point for group Y, default set to Inf

Value

NPI lower and upper probabilities for the event that the next future observation from group Y is larger than the next future observation from group X.

References

T.A. Maturi, P. Coolen-Schrijner and F.P.A. Coolen (2009). Nonparametric predictive pairwise comparison with terminated tails. International Journal of Approximate Reasoning, 51(1), 141-150.

Examples

```
data(BreakdownTimes)
data2<-split(BreakdownTimes$times, BreakdownTimes$group)
# No terminated tails, complete data
best.pair.tt(data2$X, data2$Y)
# terminated tails with Ly = 0.5, Uy = 4 and Ux = 10, but as Lx is not given then Lx=-Inf
best.pair.tt(data2$X, data2$Y, Ux = 10, Ly = 0.5, Uy = 4)
```

BirthWeights *BirthWeights data set*

Description

Data set on Birthweights for 12 male and 12 female babies as presented by Dobson (1983, p.14).

Usage

```
data(BirthWeights)
```

Format

An object of class "data.frame"

group male or female

weights BirthWeights for 12 male and 12 female babies

References

Dobson, A.J. (1983). Introduction to Statistical Modelling. Chapman and Hall, London, p.14.

Examples

```
data(BirthWeights)
head(BirthWeights)
```

BreakdownTimes *Breakdown times of units from two groups*

Description

We consider a data set used by Nelson (1982, p.462), which gives the breakdown times of units from 6 different groups. In this data set, only the first two groups are used to illustrate the NPI method for pairwise comparison with tails termination. Both groups consist of 10 observations. The first unit of group X has a reported breakdown time of 0.00, we interpret this as a very small but positive breakdown time.

Usage

```
data(BreakdownTimes)
```

Format

An object of class "data.frame"

group group X or Y

times Breakdown times

References

Nelson W. (1982). Applied Life Data Analysis. New York, Wiley, p.462.

Examples

```
data(BreakdownTimes)
head(BreakdownTimes)
```

ChemicalReaction *Chemical reaction of two methods*

Description

This data set is presented by Box et al. (1978, p. 159), where a chemical reaction was studied by making 10 runs with a standard method X, and 10 runs with a new, supposedly improved method Y.

Usage

```
data(ChemicalReaction)
```

Format

An object of class "data.frame"

method Method X or Y

value Chemical reaction values

References

Box, G.E.P, Hunter, W.G. and Hunter, J.S. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building. New York, Wiley, p.159.

Examples

```
data(ChemicalReaction)
head(ChemicalReaction)
```

FourSources *Four sources*

Description

This data set is used by Coolen and van der Laan (2001) to introduce NPI for multiple comparisons.

Usage

```
data(FourSources)
```

Format

An object of class `"data.frame"`

source Source A, B, C or D

value values given for these sources

References

Coolen F.P.A. and van der Laan P. (2001). Imprecise predictive selection based on low structure assumptions. Journal of Statistical Planning and Inference, 98(1-2), 259–277.

Examples

```
data(FourSources)
head(FourSources)
```

NPIstats *NPIstats: Nonparametric Predictive Inference*

Description

An implementation of the Nonparametric Predictive Inference approach in R. It provides tools for quantifying uncertainty via lower and upper probabilities. It includes useful functions for pairwise and multiple comparisons: comparing two groups with and without terminated tails, selecting the best group, selecting the subset of best groups, selecting the subset including the best group.

Details

Nonparametric Predictive Inference (NPI) is a statistical method which uses few modelling assumptions, enabled by the use of lower and upper probabilities to quantify uncertainty. NPI has been presented for many problems in Statistics, Risk and Reliability and Operations Research. NPI approach is based on Hill's assumption A(n), which gives a direct conditional probability for a future observable random quantity, conditional on observed values of related random quantities. Inferences based on A(n) are predictive and nonparametric, and can be considered suitable if there is hardly any knowledge about the random quantity of interest, other than the n observations, or if one does not want to use such information, e.g. to study effects of additional assumptions underlying other statistical methods. A(n) is not sufficient to derive precise probabilities for many events of interest, but it provides optimal bounds for probabilities for all events of interest involving the next future observation. These bounds are lower and upper probabilities in the theories of imprecise probability and interval probability, and as such they have strong consistency properties. NPI is a framework of statistical theory and methods that use these A(n)-based lower and upper probabilities, and also considers several variations of A(n) which are suitable for different inferences. For more info, visit NPI webpage.

References

Augustin, T. and Coolen, F.P.A. (2004). Nonparametric predictive inference and interval probability. Journal of Statistical Planning and Inference 124, 251-272.

Coolen, F.P.A. (1998). Low structure imprecise predictive inference for Bayes' problem. Statistics & Probability Letters 36, 349-357.

Coolen, F.P.A. and van der Laan, P. (2001). Imprecise predictive selection based on low structure assumptions. Journal of Statistical Planning and Inference 98, 259-277.

Coolen, F.P.A. (1996). Comparing two populations based on low stochastic structure assumptions. Statistics & Probability Letters 29, 297-305.

Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. Journal of the American Statistical Association 63, 677-691.

Weichselberger K. (2000). The theory of interval-probability as a unifying concept for uncertainty. International Journal of Approximate Reasoning, 24(2-3), 149–170.

`select.best.groups` *NPI for selecting the subset of best groups*

Description

NPI lower and upper probabilities for the event that the next future observations from groups S are greater than all future observations from the other groups.

Usage

```
select.best.groups(data, S)
```

Arguments

data	a list of numeric data vectors
S	a vector of indices of the groups in the data list to be considered as the subset of best groups

Value

NPI lower and upper probabilities for the event that the next future observations from groups S are greater than all future observations from the other groups.

Examples

```
# NPI lower and upper probabilities for the event that
# the next future observations from groups 1 and 3 are greater than
# all future observations from the other groups.

data(FourSources)
# Convert the dataframe to a list of groups
data2<-split(FourSources$value,FourSources$source)
select.best.groups(data2,S=c(1,3))
```

---

select.include.best *NPI for selecting the subset including the best group*

Description

NPI lower and upper probabilities for the event that at least one of the next future observations from groups S is greater than all future observations from the other groups.

Usage

```
select.include.best(data, S)
```

Arguments

data	a list of numeric data vectors
S	a vector of indices of the groups in the data list to be considered as the subset of groups that includes the best group.

Value

NPI lower and upper probabilities for the event that at least one of the next future observations from groups S is greater than all future observations from the other groups.

Examples

```
# NPI lower and upper probabilities for the event that  at least one
# of the next future observations from groups S is greater than all
# future observations from the other groups.

data(FourSources)
# Convert the dataframe to a list of groups
data2<-split(FourSources$value,FourSources$source)
select.include.best(data2,S=c(1,3))
```

`select.the.best` *NPI for selecting the best group*

Description

NPI lower and upper probabilities for the event that the next future observation from one (the Sth) group is greater than all future observations from the other groups.

Usage

```
select.the.best(data, S)
```

Arguments

`data`	a list of numeric data vectors
`S`	an index of the group in the data list to be considered as the best group

Value

NPI lower and upper probabilities for the event that the next future observation from the Sth group is greater than all future observations from the other groups.

Examples

```
# NPI lower and upper probabilities for the event that
# the next future observation from group 2 is greater
# than all future observations from the other groups.

data(FourSources)

# Convert the dataframe to a list of groups
data2<-split(FourSources$value,FourSources$source)
select.the.best(data2, 2)
```

Index