Package 'XICOR'

May 7, 2026

Type Package

Title Association Measurement Through Cross Rank Increments

Version 0.4.1

Author Susan Holmes [aut,cre], Sourav Chatterjee [aut]

Maintainer Susan Holmes <sp.holmes@gmail.com>

Description Computes robust association measures that do not presuppose linearity. The xi correlation (xicor) is based on cross correlation between ranked increments. The reference for the methods implemented here is Chatterjee, Sourav (2020) <doi:10.48550/arXiv.1909.10140> This package includes the Galton peas example.

Depends R (>= 3.5.0)

License Apache License (>= 2)

Date 2023-04-07

Encoding UTF-8

Imports psychTools, stats

Suggests testthat (>= 2.1.0), ggplot2

RoxygenNote 7.2.3

NeedsCompilation no

Repository CRAN

Date/Publication 2023-04-21 15:52:35 UTC

Contents

`backdec` *Inverse function to wholebinary returns the number from its expansion*

Description

Inverse function to wholebinary returns the number from its expansion

Usage

```
backdec(rmat, sgn)
```

Arguments

`rmat`	is a matrix of two rows, the first row of the matrix is the expansion of the integer part the second row is the binary expansion of the fractional part.
`sgn`	is the sign

Note

It may be necessary to make a new version of this using special functions for large integers.

`borelmerge` *Auxiliary function that takes avector and produces a single number through a Borel isomorphism using the wholebinary and backdec functions.*

Description

Auxiliary function that takes avector and produces a single number through a Borel isomorphism using the wholebinary and backdec functions.

Usage

```
borelmerge(xvec)
```

Arguments

`xvec`	is a vector of real numbers

Value

produces a single real number by converting each element

`calculateXI` *Compute the cross rank coefficient xi on two vectors.*

Description

This function computes the xi coefficient between two vectors x and y.

Usage

```
calculateXI(xvec, yvec, simple = TRUE)
```

Arguments

`xvec`	Vector of numeric values in the first coordinate.
`yvec`	Vector of numeric values in the second coordinate.
`simple`	Whether auxiliary information is kept to pass on.

Value

In the case simple = TRUE, function returns the value of the xi coefficient, If simple = FALSE is chosen, the function returns a list:

xi The xi coefficient

fr rearranged rank of yvec

CU `mean(gr*(1-gr))`

Note

Auxiliary function with no checks for NA, etc.

Author(s)

Sourav Chatterjee, Susan Holmes

References

Chatterjee, S. (2020) A New Coefficient Of Correlation, <arXiv:1909.10140>.

See Also

xicor

Examples

```
# Compute one of the coefficients
library("psychTools")
data(peas)
calculateXI(peas$parent,peas$child)
calculateXI(peas$child,peas$parent)
```

`fracbinary` *Take fractionary part and make its binary expansion Auxiliary function used in expanding real numbers*

Description

Take fractionary part and make its binary expansion Auxiliary function used in expanding real numbers

Usage

```
fracbinary(x)
```

Arguments

`x`	is a number between 0 and 1

Value

Binary expansion of length 31 of the decimal input

Note

this implementation uses the built-in function intToBits

`FRpredcor` *Compute the FR coefficient on two vectors based exactly on Gamma2.*

Description

This function computes the unidimensional graph prediction coefficient between two vectors xvec and yvec.

Usage

```
FRpredcor(xvec, yvec, tiemethod = "average")
```

Arguments

`xvec`	Vector of numeric values in the first coordinate.
`yvec`	Vector of numeric values in the second coordinate.
`tiemethod`	Choice of treatment for ties, default is the "average"

Value

In the case simple = TRUE, function returns the value of the FR standardized coefficient.

Note

Auxiliary function with no checks for NA, etc.

Author(s)

Sourav Chatterjee, Susan Holmes

References

Chatterjee, S. and Holmes, S (2020) Practical observations and applications of the robust prediction coefficient.

See Also

xicor FRpredcorhalf

Examples

```
# Compute  the coefficient and compare to the xi coefficient
simulCompare <- function(n = 20, B = 1000)
{
 diffs<- rep(0,B)
 xvec <- 1:n
 for (i in 1:B)
 {
   yvec <- runif(n)
   diffs[i] <- FRpredcor(xvec, yvec) - xicor(xvec, yvec)
 }
 return(diffs)
 }

 simulcompare1K <- simulCompare()
 summary(simulcompare1K)
```

`FRpredcorhalf` *Compute the FR half coefficient on two vectors based on half Gamma 2.*

Description

This function computes the unidimensional ranked half graph prediction coefficient between two vectors xvec and yvec.

Usage

```
FRpredcorhalf(xvec, yvec, tiemethod = "average")
```

Arguments

xvec	Vector of numeric values in the first coordinate.
yvec	Vector of numeric values in the second coordinate.
tiemethod	Choice of treatment for ties, default is the "average"

Value

In the case simple = TRUE, function returns the value of the FR standardized coefficient.

Note

Auxiliary function with no checks for NA, etc.

Author(s)

Sourav Chatterjee, Susan Holmes

References

Chatterjee, S. and Holmes, S (2020) Practical observations and applications of the robust prediction coefficient.

See Also

xicor FRpredcor

Examples

```
# Compute  the coefficient and compare to the xi coefficient
simulCompare <- function(n = 20, B = 1000)
{
 diffsim <- rep(0,B)
 xvec <- 1:n
 for (i in 1:B)
 {
   yvec <- sample(n,n)
   diffsim[i] <- FRpredcorhalf(xvec,yvec)-xicor(xvec,yvec)
 }
 return(diffsim)
 }

 compare1K <- simulCompare()
 summary(compare1K)
```

genxicor *Compute the generalized cross rank increment correlation coefficient gxi.*

Description

This function computes the generalized xi coefficient between two matrices xmat and ymat. There is a limitation on the size of the matrices, for the time being, xmat and ymat can only have 31 columns. If they are wider than 31, there is the option of using a dimension reduction technique to bring the number of columns down to 31, the first 31 components are then used. The function encodes the data using a binary expansion and then calls xicor on the vectors, so some of the arguments relevant for xicor can be specified, such as pvalue.

Usage

```
genxicor(xmat, ymat)
```

Arguments

xmat	Matrix of numeric values in the first argument.
ymat	Matrix of numeric values in the second argument.

Value

Function returns the value of the genxi coefficient. Since by default the option pvalue=TRUE is chosen, the function returns a list:

xi The value of the xi coefficient.

sd The standard deviation.

pval The test p-value.

Note

This version does not use a seed as argument, if reproducibility is an issue, set a seed before calling the function.

The p-value of rejecting independence is set to TRUE.

Author(s)

Sourav Chatterjee, Susan Holmes

References

Chatterjee, S. (2022) <arXiv:2211.04702>

Examples

```
example_joint_calc = function(n,x=runif(n),y=runif(n),ep=runif(n)) {
u = (x + y + ep) %% 1
v = ((x + y)/2 + ep) %% 1
w = (4*x/3 + 2*y/3 + ep) %% 1
z = (2*x/3 + y/3 + ep) %% 1
q = cbind(u,v,w,z)
p = cbind(x,y)
c1 = genxicor(u, p)
c2 = genxicor(v, p)
c3 = genxicor(w, p)
c4 = genxicor(z, p)
c5 = genxicor(q, p)
return(list(marg1 = c1$xi, marg2 = c2$xi, marg3 = c3$xi,
marg4 = c4$xi, joint = c5$xi, p1 = c1$pval, p2 = c2$pval, p3 = c3$pval,
p4 = c4$pval, p5 = c5$pval))
}
result1 <- example_joint_calc(n=10)
```

---

numbinary *Computes the binary expansion of a number*

Description

If the argument x is a real number the decimal portion is dropped.

Usage

```
numbinary(x)
```

Arguments

x	is a real or integer number

Value

the output is a binary vector of length 31

weave — *Take a matrix of two numbers given in their binary expansion one in each of the two rows and return the interleaving of the two numbers*

Description

Take a matrix of two numbers given in their binary expansion one in each of the two rows and return the interleaving of the two numbers

Usage

```
weave(rmat, sgn)
```

Arguments

rmat	a matrix with two times m rows corresponding to the the expansions of the m numbers to be interleaved.
sgn	is the sign vector associated to the numbers to be weaved

wholebinary — *Encodes a number as a two row binary matrix and its sign*

Description

Auxiliary function used for generating expansion of a number, the binary expansion of length nc of the integer part is the first row and the binary expansion of length nc of the fractional part is the second row of the matrix. The sign as appended into the final list object which the function returns.

Usage

```
wholebinary(x, nc = 31)
```

Arguments

x	is a decimal number
nc	is the length of the binary expansion and defines the number of columns of the output matrix

Value

This function generates a list with a binary matrix rmat with two rows and the sign sgn in a separate entry of the list.

`xicor` *Compute the cross rank increment correlation coefficient xi.*

Description

This function computes the xi coefficient between two vectors x and y, possibly all coefficients for a matrix. If only one coefficient is computed it can be used to test independence using a Monte Carlo permutation test or through an asymptotic approximation test.

Usage

```
xicor(
  x,
  y = NULL,
  pvalue = FALSE,
  ties = TRUE,
  method = "asymptotic",
  nperm = 1000,
  factor = FALSE
)
```

Arguments

`x`	Vector of numeric values in the first coordinate.
`y`	Vector of numeric values in the second coordinate.
`pvalue`	Whether or not to return the p-value of rejecting independence, if TRUE the function also returns the standard deviation of xi.
`ties`	Do we need to handle ties? If ties=TRUE the algorithm assumes that the data has ties and employs the more elaborated theory for calculating s.d. and P-value. Otherwise, it uses the simpler theory. There is no harm in putting ties = TRUE even if there are no ties.
`method`	If method = "asymptotic" the function returns P-values computed by the asymptotic theory. If method = "permutation", a permutation test with nperm permutations is employed to estimate the P-value. Usually, there is no need for the permutation test. The asymptotic theory is good enough.
`nperm`	In the case of a permutation test, `nperm` is the number of permutations to do.
`factor`	Whether to transform integers into factors, the default is to leave them alone.

Value

In the case pvalue=FALSE, function returns the value of the xi coefficient, if the input is a matrix, a matrix of coefficients is returned. In the case pvalue=TRUE is chosen, the function returns a list:

xi The value of the xi coefficient.

sd The standard deviation.

pval The test p-value.

Note

Dataset peas no longer available in psych, we are now using psychTools.

This version does not use a seed as argument, if reproducibility is an issue, set a seed before calling the function.

Author(s)

Sourav Chatterjee, Susan Holmes

References

Chatterjee, S. (2020) <arXiv:1909.10140>.

See Also

dcov

Examples

```
##---- Should be DIRECTLY executable !! ----
library("psychTools")
data(peas)
# Visualize       the peas data
library(ggplot2)
ggplot(peas,aes(parent,child)) +
geom_count() + scale_radius(range=c(0,5)) +
       xlim(c(13.5,24))+ylim(c(13.5,24))+       coord_fixed() +
       theme(legend.position="bottom")
# Compute one of the coefficients
xicor(peas$parent,peas$child,pvalue=TRUE)
xicor(peas$child,peas$parent)
# Compute all the coefficients
xicor(peas)
```

Index