Package ‘BinOrdNonNor’

April 21, 2020

Type Package

Title Concurrent Generation of Binary, Ordinal and Continuous Data

Version 1.5.1

Date 2020-04-14

Author Hakan Demirtas, Yue Wang, Rawan Allozi, Ran Gao

Maintainer Ran Gao <rgao8@uic.edu>

Description Generation of samples from a mix of binary, ordinal and continuous random variables with a pre-specified correlation matrix and marginal distributions. The details of the method are explained in Demirtas et al. (2012) <DOI:10.1002/sim.5362>.

License GPL-2 | GPL-3

Depends GenOrd, OrdNor

Imports BB, corpcor, Matrix, mvtnorm

NeedsCompilation no

Repository CRAN

Date/Publication 2020-04-21 05:20:09 UTC

R topics documented:

BinOrdNonNor-package .................................................. 2
cmat.star.BinOrdNN ................................................... 3
Fleishman.coef.NN ...................................................... 4
genBinOrdNN .......................................................... 5
IntermediateNonNor .................................................... 7
IntermediateONN ......................................................... 8
LimitforNN ............................................................. 9
LimitforONN .......................................................... 10
valid.limits.BinOrdNN .................................................. 11
validate.target.cormat.BinOrdNN .................................... 12

Index 14
Description

This package implements a procedure for generating samples from a mix of binary, ordinal and continuous random variables with a pre-specified correlation matrix and marginal distributions based on the methodology proposed by Demirtas et al. (2012) and its extensions.

This package consists of nine functions. The function `Fleishman.coef.NN` computes the Fleishman coefficients for each continuous variable with pre-specified skewness and kurtosis values. The functions `LimitforNN` and `LimitforONN` return the lower and upper correlation bounds of a pairwise correlation between two continuous variables, and between a binary/ordinal variable and a continuous variable, respectively. The function `valid.limits.BinOrdNN` computes the lower and upper bounds for the correlation entries based on the marginal distributions of the variables. The function `validate.target.cormat.BinOrdNN` checks the validity of the values of pairwise correlations. The function `IntermediateNonNor` and `IntermediateONN` compute the intermediate correlations for continuous pairs, and binary/ordinal-continuous pairs, respectively. The function `cmat.star.BinOrdNN` assembles the intermediate correlation matrix. The engine function `genBinOrdNN` generates mixed data in accordance with a given correlation matrix and marginal distributions.

The key packages and functions that we call in this package include GenOrd, OrdNor, BBsolve, rmvnorm, and nearPD.

Details

- **Package:** BinOrdNonNor
- **Type:** Package
- **Version:** 1.5.1
- **Date:** 2020-04-14
- **License:** GPL-2 | GPL-3

Author(s)

Hakan Demirtas, Yue Wang, Rawan Allozi, Ran Gao
Maintainer: Ran Gao <rgao8@uic.edu>

References


cmat.star.BinOrdNN

Description

The function computes the correlations of intermediate multivariate normal data prior to subsequent dichotomization (for binary variables), ordinalization (for ordinal variables), and transformation (for continuous variables)

Usage

```r
cmat.star.BinOrdNN(plist, skew.vec, kurto.vec, no.bin, no.ord, no.NN, CorrMat)
```

Arguments

- `plist`: A list of probability vectors corresponding to each binary/ordinal variable. The i-th element of `plist` is a vector of the cumulative probabilities defining the marginal distribution of the i-th component of the multivariate variables, which is binary/ordinal. If the i-th variable is binary, the i-th vector of `plist` will contain 1 probability value. If the i-th variable is ordinal with k categories (k > 2), the i-th vector of `plist` will contain (k-1) probability values. The k-th element is implicitly 1.

- `skew.vec`: The skewness vector for continuous variables.

- `kurto.vec`: The kurtosis vector for continuous variables.

- `no.bin`: Number of binary variables.

- `no.ord`: Number of ordinal variables.

- `no.NN`: Number of continuous variables.

- `CorrMat`: The target correlation matrix which must be positive definite and within the valid limits.

Value

An intermediate correlation of size `(no.bin + no.ord + no.NN)*(no.bin + no.ord + no.NN)`
See Also

validate.target.cormat.BinOrdNN, IntermediateNonNor, IntermediateONN

Examples

```r
## Not run:
no.bin <- 1
no.ord <- 2
no.NN <- 4
q <- no.bin + no.ord + no.NN
set.seed(54321)
Sigma <- diag(q)
Sigma[lower.tri(Sigma)] <- runif((q*(q-1)/2),-0.4,0.4)
Sigma <- Sigma + t(Sigma)
diag(Sigma) <- 1
marginal <- list(0.3, cumsum(c(0.30, 0.40) ), cumsum(c(0.4, 0.2, 0.3) ) )
cmat.star <- cmat.star.BinOrdNN(plist=marginal, skew.vec=c(1,2,2,3),
kurto.vec=c(2,7,25,25),no.bin=1, no.ord=2, no.NN=4, CorrMat=Sigma)
## End(Not run)
```

Fleishman.coef.NN

Computes the Fleishman coefficients for each continuous variable

Description

The function checks whether the skewness and kurtosis parameters violates the universal equality given in Demirtas, Hedeker, Mermelstein (2012) and computes the Fleishman coefficients for each continuous variable with pre-specified skewness and kurtosis values by solving the Fleishman’s polynomial equations using BBsolve function in BB package.

Usage

Fleishman.coef.NN(skew.vec, kurto.vec)

Arguments

skew.vec  
The skewness vector for continuous variables.

kurto.vec  
The kurtosis vector for continuous variables.

Value

An matrix with four columns corresponding to the four Fleishman coefficients, and number of rows corresponding to number of continuous variables. The i-th row contains the estimates of the four Fleishman coefficients a, b, c and d for the i-th continuous variable with i-th pre-specified skewness and kurtosis values.
References

Examples
# Consider four continuous variables, which come from
# Exp(1),Beta(4,4),Beta(4,2) and Gamma(10,10), respectively.
# Skewness and kurtosis values of these variables are as follows:

skew.vec <- c(2,0,-0.4677,0.6325)
kurto.vec <- c(6,-0.5455,-0.3750,0.6)
coef.est <- Fleishman.coef.NN(skew.vec, kurto.vec)

---

genBinOrdNN

Generates a data set with binary, ordinal and continuous variables

Description
The function simulates a sample of size \(n\) from a multivariate binary, ordinal and continuous variables with intermediate correlation matrix \(\text{cmat. star}\), and pre-specified marginal distributions.

Usage

\[
\text{genBinOrdNN}(n, \text{plist}, \text{mean.vec}, \text{var.vec}, \text{skew.vec}, \text{kurto.vec}, \text{no.bin}, \text{no.ord}, \text{no.NN}, \text{cmat. star})
\]

Arguments

- \(n\): Number of rows.
- \text{plist}\: A list of probability vectors corresponding to each binary/ordinal variable. The \(i\)-th element of \text{plist} is a vector of the cumulative probabilities defining the marginal distribution of the \(i\)-th component of the multivariate variables, which is binary/ordinal. If the \(i\)-th variable is binary, the \(i\)-th vector of \text{plist} will contain 1 probability value. If the \(i\)-th variable is ordinal with \(k\) categories (\(k > 2\)), the \(i\)-th vector of \text{plist} will contain \((k-1)\) probability values. The \(k\)-th element is implicitly 1.
- \text{mean.vec}\: Mean vector for continuous variables.
- \text{var.vec}\: Variance vector for continuous variables
- \text{skew.vec}\: The skewness vector for continuous variables.
- \text{kurto.vec}\: The kurtosis vector for continuous variables.
- \text{no.bin}\: Number of binary variables.
- \text{no.ord}\: Number of ordinal variables.
no.NN  Number of continuous variables.
cmat.star  The intermediate correlation matrix obtained from `cmat.star.BinOrdNN` function.

**Value**

A matrix of size \( n \times (\text{no.bin} + \text{no.ord} + \text{no.NN}) \), of which the first \text{no.bin} columns are binary variables, the next \text{no.ord} columns are ordinal variables, and the last \text{no.NN} columns are continuous variables.

**References**


**See Also**

`cmat.star.BinOrdNN`, `Fleishman.coef.NN`

**Examples**

```r
## Not run:
set.seed(54321)
no.bin <- 1
no.ord <- 1
no.NN <- 4
q <- no.bin + no.ord + no.NN

marginal <- list(0.4, cumsum(c(0.4, 0.2, 0.3)))

skewness.vec <- c(2,0,-0.4677,0.6325)
kurtosis.vec <- c(6,-0.5455,-0.3750,0.6)

corr.mat <- matrix(c(1.0,-0.3,-0.3,-0.3,-0.3,-0.3,-0.3, 1.0,-0.3,-0.3,-0.3,-0.3,-0.3,-0.3,-0.3, 1.0, 0.4, 0.5, 0.6,-0.3,-0.3, 0.4, 1.0, 0.7, 0.8,-0.3,-0.3, 0.5, 0.7, 1.0, 0.9,-0.3,-0.3, 0.6, 0.8, 0.9, 1.0), q,byrow=TRUE)

corr.mat.star <- cmat.star.BinOrdNN(plist=marginal, skew.vec=skewness.vec, kurt.o.vec=kurtosis.vec, no.bin=1, no.ord=1, no.NN=4, CorrMat=corr.mat)

sim.data <- genBinOrdNN(n=100000, plist=marginal, mean.vec=c(2,3,4,5), var.vec=c(3,5,10,20), skew.vec=skewness.vec, kurt.o.vec=kurtosis.vec, no.bin=1, no.ord=1, no.NN=4, cmat.star=corr.mat.star)
```
IntermediateNonNor

## End(Not run)

### IntermediateNonNor

**Description**

The function computes the intermediate correlation values of pairwise correlations between continuous variables.

**Usage**

```r
IntermediateNonNor(skew.vec, kurto.vec, cormat)
```

**Arguments**

- `skew.vec`: The skewness vector for continuous variables.
- `kurto.vec`: The kurtosis vector for continuous variables.
- `cormat`: A matrix of pairwise target correlation between continuous variables. It is a symmetric square matrix with diagonal elements being 1.

**Value**

A pairwise correlation matrix of intermediate correlation for continuous variables.

**References**


**See Also**

`IntermediateONN`, `cmat.star.BinOrdNN`

**Examples**

```r
IntermediateNonNor(skew.vec=c(1,2), kurto.vec=c(2,7),
                   cormat=matrix(c(1,-0.47,-0.47,1),2,2))
```
IntermediateONN

**Computes the intermediate (biserial/polyserial) correlations given the point-biserial/polyserial correlations for binary/ordinal-continuous pairs prior to dichotomization/ordinalization**

### Description

This function computes the intermediate correlation values of pairwise correlations between binary/ordinal and continuous variables.

### Usage

```r
IntermediateONN(plist, skew.vec, kurto.vec, ONNCorrMat)
```

### Arguments

- **plist**
  A list of probability vectors corresponding to each binary/ordinal variable. The i-th element of `plist` is a vector of the cumulative probabilities defining the marginal distribution of the i-th component of the multivariate variables, which is binary/ordinal. If the i-th variable is binary, the i-th vector of `plist` will contain 1 probability value. If the i-th variable is ordinal with k categories (k > 2), the i-th vector of `plist` will contain (k-1) probability values. The k-th element is implicitly 1.

- **skew.vec**
  The skewness vector for continuous variables.

- **kurto.vec**
  The kurtosis vector for continuous variables.

- **ONNCorrMat**
  A matrix of pairwise target (point-biserial/polyserial) correlations between binary/ordinal and continuous variables. This is a submatrix of the overall correlation matrix, and it is pertinent to the binary/ordinal-continuous part. Hence, the matrix may or may not be square. Even when it is square, it may not be symmetric.

### Value

A pairwise correlation matrix of intermediate correlations, where rows and columns represent continuous and binary/ordinal variables, respectively.

### References


### See Also

`IntermediateNonNor`, `cmat.star.BinOrdNN`
LimitforNN

Examples

```r
no.bin <- 1
no.ord <- 2
no.NN <- 4
q <- no.bin + no.ord + no.NN
set.seed(54321)

Sigma <- diag(q)
Sigma[lower.tri(Sigma)] <- runif((q*(q-1)/2),-0.4,0.4)
Sigma <- Sigma + t(Sigma)
diag(Sigma) <- 1

marginal <- list(0.3, cumsum( c(0.30, 0.40) ), cumsum(c(0.4, 0.2, 0.3) ) )
ONNCorrMat <- Sigma[4:7, 1:3]
IntermediateONN(marginal, skew.vec=c(1,2,2,3), kurto.vec=c(2,7,25,25), ONNCorrMat)
```

LimitforNN

Finds the feasible correlation range for a pair of continuous variables

Description

The function computes the lower and upper correlation bounds of a pairwise correlation between two continuous variables using generate, sort, and correlate (GSC) algorithm in Demirtas and Hedeker (2011).

Usage

```r
LimitforNN(skew.vec, kurto.vec)
Limit_forNN(skew.vec, kurto.vec) #Deprecated
```

Arguments

- `skew.vec`: The skewness vector for continuous variables.
- `kurto.vec`: The kurtosis vector for continuous variables.

Value

A vector of two elements. The first element is the lower bound and the second element is the upper bound.

References


See Also

`Fleishman.coef.NN`
LimitforONN

Finds the feasible correlation range for a pair of binary/ordinal and continuous variables

Description

The function computes the lower and upper correlation bounds of a pairwise correlation between a binary/ordinal variable and a continuous variable using GSC algorithm in Demirtas and Hedeker (2011).

Usage

LimitforONN(pvec1, skew1, kurto1)
Limit_forONN(pvec1, skew1, kurto1) #Deprecated

Arguments

- pvec1: A vector of the cumulative probabilities defining the marginal distribution for the binary/ordinal variable of the pair. If the variable is binary, the probability vector will contain only 1 probability value. If the variable is ordinal with k categories (k > 2), the probability vector will contain (k-1) values. The k-th element is implicitly 1.
- skew1: The skewness value for continuous variable of the pair.
- kurto1: The kurtosis value for continuous variable of the pair.

Value

A vector of two elements. The first element is the lower correlation bound and the second element is the upper correlation bound.

References


See Also

Fleishman.coef.NN

Examples

LimitforONN(pvec1=c(0.2, 0.5), skew1=1, kurto1=2)
valid.limits.BinOrdNN

Computes the lower and upper bounds of correlation in the form of two matrices

Description

The function computes the lower and upper bounds for the correlation entries based on the marginal distributions of the variables.

Usage

valid.limits.BinOrdNN(plist, skew.vec, kurto.vec, no.bin, no.ord, no.NN)

Arguments

- **plist**
  A list of probability vectors corresponding to each binary/ordinal variable. The i-th element of `plist` is a vector of the cumulative probabilities defining the marginal distribution of the i-th component of the multivariate variables, which is binary/ordinal. If the i-th variable is binary, the i-th vector of `plist` will contain 1 probability value. If the i-th variable is ordinal with k categories (k > 2), the i-th vector of `plist` will contain (k-1) probability values. The k-th element is implicitly 1.

- **skew.vec**
  The skewness vector for continuous variables.

- **kurto.vec**
  The kurtosis vector for continuous variables.

- **no.bin**
  Number of binary variables.

- **no.ord**
  Number of ordinal variables.

- **no.NN**
  Number of continuous variables.

Value

A list of two matrices. The one named `lower` contains the lower bounds and the other named `upper` contains the upper bounds of the feasible correlations.

See Also

LimitforNN, LimitforONN

Examples

```r
marginal <- list(0.2, c(0.4, 0.7, 0.9))
valid.limits.BinOrdNN(plist=marginal, skew.vec=c(1,2), kurto.vec=c(2,7),
no.bin=1, no.ord=1, no.NN=2)
```
validate.target.cormat.BinOrdNN

Checks the validity of the target correlation matrix

Description

The function checks the validity of pairwise correlations. In addition, it checks positive definiteness, symmetry, and correct dimensions.

Usage

validate.target.cormat.BinOrdNN(plist, skew.vec, kurto.vec, no.bin, no.ord, no.NN, CorrMat)

Arguments

- **plist**: A list of probability vectors corresponding to each binary/ordinal variable. The i-th element of `plist` is a vector of the cumulative probabilities defining the marginal distribution of the i-th component of the multivariate variables, which is binary/ordinal. If the i-th variable is binary, the i-th vector of `plist` will contain 1 probability value. If the i-th variable is ordinal with k categories (k > 2), the i-th vector of `plist` will contain (k-1) probability values. The k-th element is implicitly 1.
- **skew.vec**: The skewness vector for continuous variables.
- **kurto.vec**: The kurtosis vector for continuous variables.
- **no.bin**: Number of binary variables.
- **no.ord**: Number of ordinal variables.
- **no.NN**: Number of continuous variables.
- **CorrMat**: The target correlation matrix which must be positive definite and within the valid limits.

Value

In addition to being positive definite and symmetric, the values of pairwise correlations in the target correlation matrix must also fall within the limits imposed by the marginal distributions of the variables. The function ensures that the supplied correlation matrix is valid for simulation. If a violation occurs, an error message is displayed that identifies the violation. The function returns a logical value `TRUE` when no such violation occurs.

See Also

validate.limits.BinOrdNN
Examples

\[
\begin{align*}
\text{Sigma} & \leftarrow \text{diag}(4) \\
\text{Sigma}[\text{lower.tri(Sigma)}] & \leftarrow c(0.42, 0.55, 0.29, 0.37, 0.14, 0.26) \\
\text{Sigma} & \leftarrow \text{Sigma} + \text{t(Sigma)} \\
\text{diag(Sigma)} & \leftarrow 1
\end{align*}
\]

\[
\text{marginal} \leftarrow \text{list}(0.2, c(0.4, 0.7, 0.9))
\]

\[
\text{validate.target.cormat.BinOrdNN}(\text{plist=marginal, skew.vec=c(1,2), kurto.vec=c(2,7), no.bin=1, no.ord=1, no.NN=2, CorrMat=Sigma})
\]
Index

BinOrdNonNor (BinOrdNonNor-package), 2
BinOrdNonNor-package, 2

cmat.star.BinOrdNN, 2, 3, 6–8
Fleishman.coef.NN, 2, 4, 6, 9, 10

genBinOrdNN, 2, 5
IntermediateNonNor, 2, 4, 7, 8
IntermediateONN, 2, 4, 7, 8

Limit_forNN (LimitforNN), 9
Limit_forONN (LimitforONN), 10
LimitforNN, 2, 9, 11
LimitforONN, 2, 10, 11

valid.limits.BinOrdNN, 2, 11, 12
validate.target.cormat.BinOrdNN, 2, 4, 12