Package ‘BinOrdNonNor’

March 21, 2021

Type Package

Title Concurrent Generation of Binary, Ordinal and Continuous Data

Version 1.5.2

Date 2021-03-21

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>.

Depends GenOrd, OrdNor

Imports BB, corpcor, Matrix, mvtnorm

NeedsCompilation no

Repository CRAN

Date/Publication 2021-03-21 22:50:10 UTC

R topics documented:

BinOrdNonNor-package *Concurrent generation of binary, ordinal and continuous data*

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:	1.5.2
Date:	2021-03-21
License:	GPL-2 | GPL-3

Author(s)

Hakan Demirtas, Yue Wang, Rawan Allozi, Ran Gao

Maintainer: Ran Gao <rgao8@uic.edu>

References

Demirtas, H. and Hedeker, D. (2011). A practical way for computing approximate lower and upper correlation bounds. The American Statistician, 65(2), 104-109.

Demirtas, H. and Hedeker, D. (2016). Computing the point-biserial correlation under any underlying continuous distribution. Communications in Statistics - Simulation and Computation, 45(8), 2744-2751.

Demirtas, H., Hedeker, D., and Mermelstein, R.J. (2012). Simulation of massive public health data by power polynomials. Statistics in Medicine, 31(27), 3337-3346.

Demirtas, H. and Yavuz Y. (2015). Concurrent generation of ordinal and normal data. Journal of Biopharmaceutical Statistics, 25(4), 635-650.

Fleishman, A.I. (1978). A method for simulating non-normal distributions. Psychometrika, 43(4), 521-532.

Vale, C.D., and Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. Psychometrika, 48(3), 465-471.

`cmat.star.BinOrdNN` *Computes the intermediate correlation matrix*

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

```
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

```
## 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

Demirtas, H., Hedeker, D., and Mermelstein, R.J. (2012). Simulation of massive public health data by power polynomials. Statistics in Medicine, 31(27), 3337-3346.

Fleishman, A.I. (1978). A method for simulating non-normal distributions. Psychometrika, 43(4), 521-532.

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 `cmat.star`, and pre-specified marginal distributions.

Usage

```
genBinOrdNN(n, plist, mean.vec, var.vec, skew.vec, kurto.vec, no.bin, no.ord,
no.NN, cmat.star)
```

Arguments

`n`	Number of rows.
`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.
`mean.vec`	Mean vector for continuous variables.
`var.vec`	Variance vector for continuous variables
`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.
cmat.star	The intermediate correlation matrix obtained from cmat.star.BinOrdNN function.

Value

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

References

Demirtas, H., Hedeker, D., and Mermelstein, R.J. (2012). Simulation of massive public health data by power polynomials. Statistics in Medicine, 31(27), 3337-3346.

Demirtas, H. and Yavuz Y. (2015). Concurrent generation of ordinal and normal data. Journal of Biopharmaceutical Statistics, 25(4), 635-650.

Vale, C.D., and Maurelli, V.A. (1983). Simulating multivariate nonnormal distributions. Psychometrika, 48(3), 465-471.

See Also

cmat.star.BinOrdNN, Fleishman.coef.NN

Examples

```
## 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, 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,
kurto.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, kurto.vec=kurtosis.vec,
no.bin=1, no.ord=1, no.NN=4, cmat.star=corr.mat.star)
```

```
## End(Not run)
```

`IntermediateNonNor` *Computes the intermediate correlations for all continuous pairs*

Description

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

Usage

```
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

Demirtas, H., Hedeker, D., and Mermelstein, R.J. (2012). Simulation of massive public health data by power polynomials. Statistics in Medicine, 31(27), 3337-3346.

Vale, C.D., and Maurelli, V.A.(1983). Simulating multivariate nonnormal distributions. Psychometrika, 48(3), 465-471.

See Also

`IntermediateONN`, `cmat.star.BinOrdNN`

Examples

```
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

```
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

Demirtas, H., Hedeker, D., and Mermelstein, R.J. (2012). Simulation of massive public health data by power polynomials. Statistics in Medicine, 31(27), 3337-3346.

Demirtas, H. and Hedeker, D. (2016). Computing the point-biserial correlation under any underlying continuous distribution. Communications in Statistics - Simulation and Computation, 45(8), 2744-2751.

See Also

`IntermediateNonNor`, `cmat.star.BinOrdNN`

Examples

```
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

```
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

Demirtas, H., Hedeker, D. (2011). A practical way for computing approximate lower and upper correlation bounds. The American Statistician, 65(2), 104-109.

See Also

`Fleishman.coef.NN`

Examples

```
LimitforNN(skew.vec=c(1,2),kurto.vec=c(2,7))
```

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

Demirtas, H., Hedeker, D. (2011). A practical way for computing approximate lower and upper correlation bounds. The American Statistician, 65(2), 104-109.

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

```
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

`valid.limits.BinOrdNN`

Examples

```
Sigma <- diag(4)
Sigma[lower.tri(Sigma)] <- c(0.42, 0.55, 0.29, 0.37, 0.14, 0.26)
Sigma <- Sigma + t(Sigma)
diag(Sigma) <- 1

marginal <- list(0.2, c(0.4, 0.7, 0.9))

validate.target.cormat.BinOrdNN(plist=marginal, skew.vec=c(1,2), kurto.vec=c(2,7),
                                no.bin=1, no.ord=1, no.NN=2, CorrMat=Sigma)
```

Index