Package ‘PoisBinOrdNor’

March 28, 2017

Type  Package
Title  Data Generation with Poisson, Binary, Ordinal and Normal Components
Version  1.4
Date  2017-03-28
Author  Hakan Demirtas, Yiran Hu, Rawan Allozi
Maintainer  Rawan Allozi <ralloz2@uic.edu>
Description  Generation of multiple count, binary, ordinal and normal variables simultaneously given the marginal characteristics and association structure.
License  GPL-2 | GPL-3
Depends  Matrix, corpcor, mvtnorm, psych, GenOrd
NeedsCompilation  no
Repository  CRAN
Date/Publication  2017-03-28 16:20:55 UTC

R topics documented:

PoisBinOrdNor-package .................................................. 2
corr.nn4bb .............................................................. 3
corr.nn4bn .............................................................. 4
corr.nn4on .............................................................. 4
corr.nn4pbo ............................................................. 5
corr.nn4pn .............................................................. 6
corr.nn4pp .............................................................. 7
genPBONdata ............................................................. 7
intermat ................................................................. 9
validation_specs ......................................................... 11

Index  14
Data Generation with Count, Binary, Ordinal and Normal Components

Description

Generation of multiple count, binary, ordinal and normal variables simultaneously given the marginal characteristics and association structure based on the methodologies proposed in Demirtas et al. (2012), Demirtas and Yavuz (2015), Amatya and Demirtas (2016), Demirtas and Hedeker (2016).

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>PoisBinOrdNor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>1.4</td>
</tr>
<tr>
<td>Date</td>
<td>2017-03-28</td>
</tr>
<tr>
<td>License</td>
<td>GPL-2</td>
</tr>
</tbody>
</table>

PoisBinOrdNor package consists of ten functions. The function validation_specs validates the specified quantities to avoid obvious specification errors. The functions corr.nn4bb, corr.nn4bn, corr.nn4on, corr.nn4pbo, corr.nn4pn, and corr.nn4pp each computes the intermediate correlation coefficient for binary-binary combinations, binary-normal combinations, ordinal-normal combinations, count-binary/ordinal combinations, count-normal and count-count combinations, respectively. The function intermat assembles the intermediate correlation matrix for the multivariate data based on input from functions corr.nn4bb, corr.nn4bn, corr.nn4on, corr.nn4pbo, corr.nn4pn and corr.nn4pp. The engine function genPBONdata computes the final correlation matrix and generates mixed data in accordance with the specified marginal and correlational quantities.

Author(s)

Hakan Demirtas, Yiran Hu, Rawan Allozi

Maintainer: Rawan Allozi <ralloz2@uic.edu>

References


Finds the tetrachoric correlation based on user-specified correlation between binary variables.

Description

This function computes the tetrachoric correlation given the correlation for a pair of binary variables (phi coefficient).

Usage

\texttt{corr.nn4bb(p1, p2, BB.cor)}

Arguments

- \texttt{p1} Probability parameter for the first binary variable.
- \texttt{p2} Probability parameter for the second binary variable.
- \texttt{BB.cor} Pre-specified correlation for a pair of binary variables.

Value

A tetrachoric correlation coefficient.

References


Examples

```r
## Not run:
corr.nn4bb(0.43, 0.7, 0.129)
## End(Not run)
```
**corr.nn4bn**  
Finds the biserial correlation given the correlation for a binary-normal pair.

### Description

This function computes the biserial correlation given the specified correlation for a pair of binary and normal variables (point-biserial correlation).

### Usage

```r
corr.nn4bn(p, BN.cor)
```

### Arguments

- `p`  
  Probability parameter for the binary variable.
- `BN.cor`  
  Pre-specified correlation for a pair of binary and normal variables.

### Value

A biserial correlation coefficient.

### Examples

```r
## Not run:
corr.nn4bn(0.43, 0.12)
## End(Not run)
```

---

**corr.nn4on**  
Finds polyserial correlation for given the correlation for an ordinal-normal pair.

### Description

This function computes the polyserial correlation given the specified correlation for a pair of ordinal and normal variables (point-polyserial correlation).

### Usage

```r
corr.nn4on(p, ON.cor)
```
**corr.nn4pbo**

**Arguments**

- **p**  
  A vector of probabilities for an ordinal variable. The i-th element of the pvec is the cumulative probability defining the marginal distribution of the ordinal variable. If the variable has k categories, the i-th element of p will contain k-1 probabilities. The k-th element is implicitly 1.

- **ON.cor**  
  Pre-specified correlation for a pair of ordinal-normal variables.

**Value**

A tetrachoric correlation coefficient.

**Examples**

```r
## Not run:
corr.nn4on(c(0.33, 0.66), 0.22)
## End(Not run)
```

---

**corr.nn4pbo**  
Finds the underlying bivariate normal correlation given the correlation for a count-binary or count-ordinal pair.

---

**Description**

This function computes the underlying bivariate normal correlation given the correlation for a pair of count and binary variables or a pair of count and ordinal variables.

**Usage**

```r
corr.nn4pbo(lam, p, P0.cor)
```

**Arguments**

- **lam**  
  Rate parameter for the count variable.

- **p**  
  A vector of probabilities for an ordinal variable. The i-th element of the pvec is the cumulative probability defining the marginal distribution of the ordinal variable. If the variable has k categories, the i-th element of p will contain k-1 probabilities. The k-th element is implicitly 1.

- **P0.cor**  
  Pre-specified correlation for a pair of count and binary, or count and ordinal, variables.

**Value**

A tetrachoric correlation coefficient.
References

Examples
## not run:
corr.nn4pn(c(PNPUL PNP7L c(PN2L PNU)L PN23U)
## end(not run)

corr.nn4pn
Finds the underlying bivariate normal correlation given the correlation for a count-normal pair.

Description
This function computes the underlying bivariate normal correlation given the specified correlation for a pair of count and normal variables.

Usage
corr.nn4pn(lam, PN.cor)

Arguments
lam        Rate parameter for the count variable.
PN.cor     Pre-specified correlation for a pair of count and normal variables.

Value
Correlation of underlying bivariate normal data.

Examples
## not run:
corr.nn4pn(0.05, 0.32)
## End(Not run)
corr.nn4pp

Finds the underlying bivariate normal correlation given the correlation for a pair of count variables.

Description

This function computes the underlying bivariate normal correlation given the specified correlation for a pair of count variables.

Usage

corr.nn4pp(lambda1, lambda2, PP.cor)

Arguments

lambda1       Rate parameter for the first count variable.
lambda2       Rate parameter for the second count variable.
PP.cor        Pre-specified correlation for a pair of count variables.

Value

Correlation of underlying bivariate normal data.

References


Examples

## Not run:
corr.nn4pp(0.05, 0.02, 0.34)

## End(Not run)

genPBONdata

Generates correlated data with multiple count, binary, ordinal and normal variables

Description

This function simulates a multivariate data set that is composed of count, binary, ordinal and normal variables with specified marginals and a correlation matrix.
Usage

\texttt{genPBONdata(n, no\_pois, no\_bin, no\_ord, no\_norm, inter\_mat, lamvec, prop\_vec\_bin, prop\_vec\_ord, nor\_mean, nor\_var)}

Arguments

\begin{itemize}
\item \texttt{n} \hspace{1cm} \text{Number of rows}
\item \texttt{no\_pois} \hspace{1cm} \text{Number of count variables}
\item \texttt{no\_bin} \hspace{1cm} \text{Number of binary variables}
\item \texttt{no\_ord} \hspace{1cm} \text{Number of ordinal variables}
\item \texttt{no\_norm} \hspace{1cm} \text{Number of normal variables}
\item \texttt{inter\_mat} \hspace{1cm} \text{The intermediate correlation matrix obtained from function intermat}
\item \texttt{lamvec} \hspace{1cm} \text{A vector of marginal rates for the count variables}
\item \texttt{prop\_vec\_bin} \hspace{1cm} \text{A vector of probabilities for the binary variables}
\item \texttt{prop\_vec\_ord} \hspace{1cm} \text{A vector of probabilities for the ordinal variables. For each of the variable, the i-th element of the pvec is the cumulative probability defining the marginal distribution of the ordinal variable. If the variable has k categories, the i-th element of p will contain k-1 probabilities. The k-th element is implicitly 1.}
\item \texttt{nor\_mean} \hspace{1cm} \text{A vector of means for the normal variables}
\item \texttt{nor\_var} \hspace{1cm} \text{A vector of variances for the normal variables}
\end{itemize}

Value

\begin{itemize}
\item \texttt{data} \hspace{1cm} \text{A simulated data matrix of size nx(no\_pois + no\_bin + no\_ord + no\_norm), of which the first no\_pois are count variables, followed by no\_bin binary variables, no\_ord ordinal variables, and lastly no\_norm normal variables.}
\item \texttt{n\_rows} \hspace{1cm} \text{Number of rows in the simulated data}
\item \texttt{prob\_bin} \hspace{1cm} \text{A vector of probabilities for the binary variables}
\item \texttt{prob\_ord} \hspace{1cm} \text{A vector of probabilities for the ordinal variables}
\item \texttt{nor\_mean} \hspace{1cm} \text{A vector of means for the normal variables}
\item \texttt{nor\_var} \hspace{1cm} \text{A vector of variances for the normal variables}
\item \texttt{lamvec} \hspace{1cm} \text{A vector of rate parameters for the count variables}
\item \texttt{n\_pois} \hspace{1cm} \text{Number of count variables}
\item \texttt{n\_bin} \hspace{1cm} \text{Number of binary variables}
\item \texttt{n\_ord} \hspace{1cm} \text{Number of ordinal variables}
\item \texttt{n\_norm} \hspace{1cm} \text{Number of normal variables}
\item \texttt{final\_corr} \hspace{1cm} \text{The final correlation matrix for the simulated data}
\end{itemize}
**Examples**

```r
## Not run:
ss=10000
num_pois<-2
num_bin<-1
num_ord<-2
num_norm<-1

lamvec=sample(10,2)
pbin=runif(1)
pord=list(c(0.1, 0.9), c(0.2, 0.3, 0.5))
nor_mean=3.1
nor.var=0.85
M=c(-0.05, 0.26, 0.14, 0.09, 0.14, 0.12, 0.13, -0.02, 0.17, 0.29,
    -0.04, 0.19, 0.10, 0.35, 0.39)
N=diag(6)
N[lower.tri(N)]=M
TV=N+t(N)
diag(TV)<-1
intmat<-intermat(num_pois,num_bin,num_ord,num_norm,corr_mat=TV,pbin,pord,lamvec,
nor.mean,nor.var)

genPBOndata(ss,num_pois,num_bin,num_ord,num_norm,intmat,lamvec,pbin,pord,nor.mean,nor.var)

## End(Not run)
```

---

**intermat**

*Calculates and assembles the intermediate correlation matrix entries for the multivariate normal data.*

**Description**

This function computes and assembles the correlation entries for the intermediate multivariate normal data.

**Usage**

```r
intermat(no_pois, no_bin, no_ord, no_norm, corr_mat, prop_vec_bin, prop_vec_ord,
lam_vec, nor_mean, nor_var)
```

**Arguments**

- `no_pois` Number of the count variables.
- `no_bin` Number of the binary variables.
- `no_ord` Number of the ordinal variables.
- `no_norm` Number of the normal variables.
- `corr_mat` Pre-specified correlation matrix for the multivariate data.
intermat

prop_vec_bin  Vector of probabilities for the binary variables.
prop_vec_ord  Vector of probabilities for the ordinal variables.
lam_vec       Vector of rate parameters for the count variables.
nor_mean      Vector of means for the normal variables.
nor_var       Vector of variances for the normal variables.

Value

The intermediate correlation matrix that will be used later for multivariate normal data simulation.

References


See Also

Examples

```r
## Not run:
num_pois<M2
num_bin<M1
num_ord<M2
num_norm<M1
lamvec]sample(1PL2)
pbin]runif(1)
pord]list(c(PN3L PN7L c(PN2L PN3L PN9U))
norNmean]3N1
norNvar]PN8U
m]c(MPNPUL PN2VL PN1TL PNP9L PN1TL PN12L PN13L MPNP2L PN17L PN29L MPNPTL PN19L PN1PL PN3UL PN39)

n]diag(V)
n[lowerNtri(n)]m
tv]n+t(n)
diag(tv)<M1

intmat<M

## end(not run)
```

## End(Not run)
validation_specs

Validates user-specified parameters

Description

This function checks the validity of user specified parameters including rate parameters for count variables, proportion parameters for binary and ordinary variables, mean and variance parameters for normal data, as well as the validity of entries in the correlation matrix. This function also computes the lower and upper limits for each pairwise correlation based on the marginal probabilities for range violation checks.

Usage

validation_specs(no.pois, no.bin, no.ord, no.norm, corr.mat, prop.vec.bin, prop.vec.ord, lamvec, nor.mean, nor.var)

Arguments

- **no.pois**: Number of count variables.
- **no.bin**: Number of binary variables.
- **no.ord**: Number of ordinal variables.
- **no.norm**: Number of normal variables.
- **corr.mat**: User specified correlation matrix for the multivariate data.
- **prop.vec.bin**: Vector of probabilities corresponding to each of the binary variables.
- **prop.vec.ord**: Vector of probabilities corresponding to each of the ordinal variables. For each of the ordinal variable, the i-th element of the probability vector is the cumulative probability defining the marginal distribution of the ordinal variable. If the variable has k categories, the i-th element of p will contain k-1 probabilities. The k-th element is implicitly 1.
- **lamvec**: Vector of rate parameters for the count variables.
- **nor.mean**: Vector of means for the normal variables.
- **nor.var**: Vector of variances for the normal variables.

Details

This function computes the lower and upper bounds for all possible pairs that involve count, binary, ordinal and normal variables.

Value

The function returns TRUE if no specification problem is encountered. Otherwise, it returns an error message.
References


Examples

```r
## not run:
num_pois<-1
num_bin<-1
num_ord<-1
num_norm<-1
lambda<-c(1)
pbin<-c(0.3)
pord<-list(c(0.3,0.6))
normean<-15
norvar<-7
corr.mat=matrix(c(1,2,0.1,0.3,0.2,1,0.5,0.4,0.1,0.5,1,0.7,0.3,0.4,0.7,1),4,4)
validation_specs(num_pois, num_bin, num_ord, num_norm, corr.mat, pbin, pord, lambda, normean,norvar)

num_pois<-2
num_bin<-2
num_ord<-2
num_norm<-0
lambda<-c(1,2)
pbin<-c(0.3,0.5)
pord<-list(c(0.3,0.6),c(0.5,0.6))
corr.mat=matrix(0.64,6,6)
diag(corr.mat)=1
validation_specs(num_pois, num_bin, num_ord, num_norm, corr.mat, pbin, pord, lambda, nor.mean=NULL, nor.var=NULL)

# An example with an invalid target correlation matrix (bound violation).
num_pois<-1
num_bin<-2
num_ord<-2
num_norm<-1
lamvec=c(1)
pbin=c(0.3, 0.7)
pord=list(c(0.2, 0.5), c(0.4, 0.7, 0.8))
nor.mean=2.1
nor.var=0.75
M=c(-0.35, 0.26, 0.34, 0.09, 0.14, 0.12, 0.30, -0.02, 0.17, 0.29, -0.04, 0.19, 0.10, 0.35, 0.39)
N=diag(6)
N[lower.tri(N)]=M
TV=N+t(N)
```
diag(TV)<-1
validation_specs(num_pois, num_bin, num_ord, num_norm, corr.mat=TV, pbin, pord, lamvec, normean, norvar)

# An example with a non-positive definite correlation matrix.
pbin=c(0.3, 0.7)
TV1=TV
TV1[3,2]=TV[2,3]=5
validation_specs(num_pois, num_bin, num_ord, num_norm, corr.mat=TV1, pbin, pord, lamvec, normean, norvar)

## End(Not run)
Index

* Topic concurrent generation of count, binary, ordinal and normal variables
  PoisBinOrdNor-package, 2

* Topic generating multivariate binary variables
  PoisBinOrdNor-package, 2

* Topic generating multivariate count variables
  PoisBinOrdNor-package, 2

* Topic generating multivariate normal variables
  PoisBinOrdNor-package, 2

* Topic generating multivariate ordinal variables
  PoisBinOrdNor-package, 2

  corr.nn4bb, 2, 3, 10
  corr.nn4bn, 2, 4, 10
  corr.nn4on, 2, 4, 10
  corr.nn4pbo, 2, 5, 10
  corr.nn4pn, 2, 6, 10
  corr.nn4pp, 2, 7, 10

  genPBOndata, 2, 7

  intermat, 2, 9

  PoisBinOrdNor (PoisBinOrdNor-package), 2
  PoisBinOrdNor-package, 2

  validation_specs, 2, 10, 11