Package ‘bindata’

October 12, 2022

Version 0.9-20
Date 2021-01-29
Title Generation of Artificial Binary Data
Description Generation of correlated artificial binary data.
License GPL-2
Imports e1071, mvtnorm (>= 0.7-0)
NeedsCompilation no
Author Friedrich Leisch [aut, cre],
      Andreas Weingessel [aut],
      Kurt Hornik [aut]
Maintainer Friedrich Leisch <Friedrich.Leisch@R-project.org>
Repository CRAN
Date/Publication 2021-01-29 12:00:09 UTC

R topics documented:

  bincorr2commonprob .................................................. 2
  check.commonprob .................................................... 3
  commonprob2sigma .................................................... 4
  condprob ............................................................ 5
  ra2ba ............................................................... 5
  rmvbin .............................................................. 6
  simul.commonprob .................................................... 7
  SimulVals .......................................................... 8

Index 9
bincorr2commonprob  

Convert Binary Correlation Matrix to Matrix of Joint Probabilities

Description

Compute a matrix of common probabilities for a binary random vector from given marginal probabilities and correlations.

Usage

\[
bincorr2commonprob(margprob, bincorr)
\]

Arguments

- **margprob**: vector of marginal probabilities.
- **bincorr**: matrix of binary correlations.

Value

The matrix of common probabilities. This has the probabilities that variable \( i \) equals 1 in element \((i, i)\), and the joint probability that variables \( i \) and \( j \) both equal 1 in element \((i, j)\) (if \( i \neq j \)).

Author(s)

Friedrich Leisch

References


See Also

commonprob2sigma, simul.commonprob.
check.commonprob

Check Joint Binary Probabilities

Description

The main diagonal elements commonprob[i, i] are interpreted as probabilities \( p_{A_i} \) that a binary variable \( A_i \) equals 1. The off-diagonal elements commonprob[i, j] are the probabilities \( p_{A_i A_j} \) that both \( A_i \) and \( A_j \) are 1.

This program checks some necessary conditions on these probabilities which must be fulfilled in order that a joint distribution of the \( A_i \) with the given probabilities can exist. The conditions checked are

\[
0 \leq p_{A_i} \leq 1
\]

\[
\max(0, p_{A_i} + p_{A_j} - 1) \leq p_{A_i A_j} \leq \min(p_{A_i}, p_{A_j}), \quad i \neq j
\]

\[
p_{A_i} + p_{A_j} + p_{A_k} - p_{A_i A_j} - p_{A_i A_k} - p_{A_j A_k} \leq 1, \quad i \neq j, i \neq k, j \neq k
\]

Usage

check.commonprob(commonprob)

Arguments

commonprob Matrix of pairwise probabilities.

Value

check.commonprob returns TRUE, if all conditions are fulfilled. The attribute "message" of the return value contains some information on the errors that were found.

Author(s)

Andreas Weingessel

References


See Also

simul.commonprob.commonprob2sigma
Examples

check.commonprob(cbind(c(0.5, 0.4), c(0.4, 0.8)))
check.commonprob(cbind(c(0.5, 0.25), c(0.25, 0.8)))
check.commonprob(cbind(c(0.5, 0, 0), c(0, 0.5, 0), c(0, 0, 0.5)))

commonprob2sigma  Calculate a Covariance Matrix for the Normal Distribution from a
                   Matrix of Joint Probabilities

Description

Computes a covariance matrix for a normal distribution which corresponds to a binary distribution with marginal probabilities given by diag(commonprob) and pairwise probabilities given by commonprob.

For the simulations the values of simulvals are used.

If a non-valid covariance matrix is the result, the program stops with an error in the case of NA arguments and yields are warning message if the matrix is not positive definite.

Usage

commonprob2sigma(commonprob, simulvals)

Arguments

commonprob  matrix of pairwise probabilities.
simulvals   array received by simul.commonprob.

Value

A covariance matrix is returned with the same dimensions as commonprob.

Author(s)

Friedrich Leisch

References


See Also

simul.commonprob
### Examples

```r
m <- cbind(c(1/2,1/5,1/6),c(1/5,1/2,1/6),c(1/6,1/6,1/2))
sigma <- commonprob2sigma(m)
```

### Description

Returns a matrix containing the conditional probabilities \( P(x_i = 1|x_j = 1) \) where \( x_i \) corresponds to the \( i \)-th column of \( x \).

### Usage

```r
condprob(x)
```

#### Arguments

- **x**: matrix of binary data with rows corresponding to cases and columns corresponding to variables.

### Author(s)

Friedrich Leisch

---

### ra2ba

Convert Real Valued Array to Binary Array

### Description

Converts all values of the real valued array \( x \) to binary values by thresholding at 0.

### Usage

```r
ra2ba(x)
```

#### Arguments

- **x**: array of arbitrary dimension

### Author(s)

Friedrich Leisch

### Examples

```r
x <- array(rnorm(10), dim=c(2,5))
ra2ba(x)
```
**rmvbin**  
*Multivariate Binary Random Variates*

**Description**

Creates correlated multivariate binary random variables by thresholding a normal distribution. The correlations of the components can be specified either as common probabilities, correlation matrix of the binary distribution, or covariance matrix of the normal distribution.

**Usage**

```r
rmvbin(n, margprob, commonprob=diag(margprob),
       bincorr=diag(length(margprob)),
       sigma=diag(length(margprob)),
       colnames=NULL, simulvals=NULL)
```

**Arguments**

- `n`: number of observations.
- `margprob`: margin probabilities that the components are 1.
- `commonprob`: matrix of probabilities that components $i$ and $j$ are simultaneously 1.
- `bincorr`: matrix of binary correlations.
- `sigma`: covariance matrix for the normal distribution.
- `colnames`: vector of column names for the resulting observation matrix.
- `simulvals`: result from `simul.commonprob`, a default data array is automatically loaded if this argument is omitted.

**Details**

Only one of the arguments `commonprob`, `bincorr` and `sigma` may be specified. Default are uncorrelated components.

$n$ samples from a multivariate normal distribution with mean and variance chosen in order to get the desired margin and common probabilities are sampled. Negative values are converted to 0, positive values to 1.

**Author(s)**

Friedrich Leisch

**References**

**simul.commonprob**

**See Also**

commonprob2sigma, check.commonprob, simul.commonprob

**Examples**

```r
## uncorrelated columns:
rmvbin(10, margprob=c(0.3,0.9))

## correlated columns
m <- cbind(c(1/2,1/5,1/6),c(1/5,1/2,1/6),c(1/6,1/6,1/2))
rmvbin(10,commonprob=m)

## same as the second example, but faster if the same probabilities are
## used repeatedly (commonprob2sigma rather slow)
sigma <- commonprob2sigma(m)
rmvbin(10,margprob=diag(m),sigma=sigma)
```

---

**simul.commonprob**  
*Simulate Joint Binary Probabilities*

**Description**

Compute common probabilities of binary random variates generated by thresholding normal variates at 0.

**Usage**

```r
simul.commonprob(margprob, corr=0, method="integrate", n1=10^5, n2=10)
```

**Arguments**

- `margprob`: vector of marginal probabilities.
- `corr`: vector of correlation values for normal distribution.
- `method`: either "integrate" or "monte carlo".
- `n1`: number of normal variates if method is "monte carlo".
- `n2`: number of repetitions if method is "monte carlo".

**Details**

The output of this function is used by `rmvbin`. For all combinations of `margprob[i]`, `margprob[j]` and `corr[k]`, the probability that both components of a normal random variable with mean `qnorm(margprob[c(i,j)])` and correlation `corr[k]` are larger than zero is computed.

The probabilities are either computed by numerical integration of the multivariate normal density, or by Monte Carlo simulation.

For normal usage of `rmvbin` it is not necessary to use this function, one simulation result is provided as variable `SimulVals` in this package and loaded by default.
Value

`simul.commonprob` returns an array of dimension `c(length(margprob), length(margprob), length(corr))`.

Author(s)

Friedrich Leisch

References


See Also

`rmvbin`

Examples

```r
simul.commonprob(seq(0,1,0.5), seq(-1,1,0.5), meth="mo", n1=10^4)
data(SimulVals)
```

<table>
<thead>
<tr>
<th>SimulVals</th>
<th>Pre-simulated Joint Binary Probabilities</th>
</tr>
</thead>
</table>

Description

This variable provides a pre-fabricated result from `simul.commonprob` such that it is normally not necessary to use this (time consuming) function, and is used by `rmvbin`.

Usage

`SimulVals`

Author(s)

Friedrich Leisch

References


See Also

`simul.commonprob, rmvbin`
Index

* array
  ra2ba. 5
* distribution
  bincorr2commonprob. 2
  check.commonprob. 3
  commonprob2sigma. 4
  condprob. 5
  rmvbin. 6
  simul.commonprob. 7
* multivariate
  bincorr2commonprob. 2
  check.commonprob. 3
  commonprob2sigma. 4
  rmvbin. 6
  simul.commonprob. 7
* sysdata
  SimulVals. 8
  bincorr2commonprob. 2
  check.commonprob. 3, 7
  commonprob2sigma. 2, 3, 4, 7
  condprob. 5
  ra2ba. 5
  rmvbin. 6, 7, 8
  simul.commonprob. 2–4, 7, 7, 8
  SimulVals. 7, 8