Package 'corcounts'

November 13, 2009

Type Package

Title Generate correlated count random variables

Version 1.4

Date 2009-11-13

Author Vinzenz Erhardt

Maintainer Vinzenz Erhardt <erhardt@ma.tum.de>

Description Generate high-dimensional correlated count random variables with a prespecified Pearson correlation.

License GPL (>= 3)

LazyLoad yes

Repository CRAN

Date/Publication 2009-11-13 13:35:16

R topics documented:

corcounts-package *Generate correlated count random variables*

Description

Sample high-dimensional correlated count random variables with approximate prespecified Pearson correlation and exact margins.

Details

Package:	corcounts
Type:	Package
Version:	1.4
Date:	2009-11-13
License:	GPL (>= 3)
LazyLoad:	yes

Specify the marginal distributions and parameters and the desired correlation matrix and run 'rcounts()'.

In order to allow for regression, i.e. for individual parameters for each cluster, run 'rcounts.reg()'.

Author(s)

Maintainer: Vinzenz Erhardt <erhardt@ma.tum.de>

References

Erhardt, V., Czado, C. (2009) A method for approximately sampling high-dimensional count variables with prespecified Pearson correlation. Submitted for publication.

Czado, C., Erhardt, V., Min, A., Wagner, S. (2007) Zero-inflated generalized Poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates. Statistical Modelling 7 (2), 125-153.

c2pc *Calculate partial correlations from a correlation matrix*

Description

'c2pc' is used to calculate partial correlations from a correlation matrix.

Usage

```
c2pc(Cin)
```

Arguments

`Cin`	A symmetric positive definite correlation matrix.

Details

If you obtain values not in [-1,1], your correlation matrix is not positive definite.

This routine only calculates partial correlations conditional on 1, 12, 123, 1234, etc.. Partial correlations conditional on other margins can be obtained by a permutation of margins.

Value

The partial correlations calculated will be

Theta =

12.....13.....14.....15.....16

.....23|1.....24|1.....25|1.....26|1

.....34|12.....35|12.....36|12 ...

.....45|123.....46|123

.....56|1234

...

Author(s)

Vinzenz Erhardt

See Also

Package 'corpcor' calculates partial correlations conditional on ALL other margins.

Examples

```
# create 8 dimensional symmetric positive correlation matrix with random entries
Cin <- unstructured(8)
Cin

Theta <- c2pc(Cin)
Theta

# transform Theta back to obtain the correlation matrix
pc2c(Theta)

# identical with Cin
pc2c(Theta) - Cin
```

`modified.cvine.alg` *Auxiliary function*

Description

auxiliary function

`modified.cvine.alg.reg` *Auxiliary function*

Description

auxiliary function

`pc2c` *Calculate a corresponding correlation matrix to partial correlations*

Description

'c2pc' is used to calculate the corresponding correlation matrix of dimension T times T out of partial correlations.

Usage

```
pc2c(Theta)
```

Arguments

`Theta`	A T times T matrix with partial correlations. See details.

Details

The partial correlations in Theta have to be specified as

Theta =

12.....13.....14.....15.....16

.....23|1.....24|1.....25|1.....26|1

.....34|12.....35|12.....36|12 ...

.....45|123.....46|123

.....56|1234

...

and may be NA elsewhere. Theta has to be of dimension T times T.

This routine only calculates partial correlations conditional on 1, 12, 123, 1234, etc.. Partial correlations conditional on other margins can be obtained by a permutation of margins.

Value

A symmetric positive definite correlation matrix of dimension T times T.

Author(s)

Vinzenz Erhardt

Examples

```
# create random uniform(0,1) partial correlations in dimension 8
dimension <- 8
Theta <- matrix(NA,dimension,dimension)
for (i in 2:dimension) {
  for (j in 1:(i-1)) {
    Theta[j,i] <- runif(1,-1,1)
  }
}
Theta

# calculate corresponding correlation matrix
C <- pc2c(Theta)
C

# transform back to partial correlations
c2pc(C)

# equivalence with original Theta
Theta - c2pc(C)
```

---

pseudoinv.zigp *Auxiliary function*

Description

auxiliary function

R11.AR1 *Auxiliary function*

Description

auxiliary function

R11.exchangeable *Auxiliary function*

Description

auxiliary function

rcounts *Generate correlated count random variables*

Description

'rcounts' is used to sample high-dimensional correlated count random variables with approximate prespecified Pearson correlation and exact margins.

Usage

```
rcounts(N, margins, mu, phi=rep(NA,length(margins)),
        omega=rep(NA,length(margins)), psi=rep(NA,length(margins)),
        corstr, corpar, conv=0.01)
```

Arguments

N	number of observations to be generated per margin (should be at least 500).
margins	Vector of margin tokens. Its length T is the dimension. See details.
mu	Vector of length T of means for the Poisson, GP, ZIP, ZIGP and NB margins.
phi	Vector of length T of dispersion parameters for the GP, and ZIGP margins. For Poisson, ZIP and NB margins, an 'NA' can be provided.
omega	Vector of length T of zero-inflation parameters for the ZIP and ZIGP margins. For Poisson, GP and NB margins, an 'NA' can be provided.
psi	Vector of length T of size parameters for the NB margins. For Poisson, GP, ZIP and ZIGP margins, an 'NA' can be provided.
corstr	Correlation structure. Can be 'ex' for exchangeable, 'AR1' for AR(1) and 'unstr' for unstructured.
corpar	Correlation parameter. Scalar correlation for 'ex' and 'AR1' and matrix of dimension TxT for 'unstr'.
conv	Convergence criterion

Details

The entries in 'margins' can be specified as 'Poi' for Poisson, 'GP' for generalized Poisson, 'ZIP' for zero-inflated Poisson, 'ZIGP' for zero-inflated generalized Poisson and 'NB' for negative-binomial.

Value

The function will return a matrix of counts of dimension N x T.

Author(s)

Vinzenz Erhardt

Examples

```
N <- 5000

# high precision in dimension 2
margins <- c("ZIGP","GP")
mu <-    c(10,  15)
phi <-   c(1.5, 3.5)
omega <- c(.25, NA)
corstr <- "ex"
corpar <- .5
Y <- rcounts(N=N, margins=margins, mu=mu, phi=phi, omega=omega, corstr=corstr,
             corpar=corpar, conv=0.0001)
cor(Y)


# five-dimensional examples
margins <- c("ZIGP","GP","Poi","NB","ZIP")
mu <-    c(10,  25,  12, 20, 28)
phi <-   c(1.5, 2,   NA, NA, NA)
omega <- c(.25, NA,  NA, NA, .2)
psi <-   c(NA,  NA,  NA, 7,  NA)

# Exchangeable structure with correlation of 0.5
corstr <- "ex"
corpar <- .5
Y <- rcounts(N=N, margins=margins, mu=mu, phi=phi, omega=omega, psi=psi,
             corstr=corstr, corpar=corpar)
cor(Y)

# AR(1) structure with correlation of corr(Y(t1), Y(t2)) = .8 ^ |t1-t2|
corstr <- "AR1"
corpar <- .8
Y <- rcounts(N=N, margins=margins, mu=mu, phi=phi, omega=omega, psi=psi,
             corstr=corstr, corpar=corpar)
cor(Y)

# Unstructured correlation. Create random symmetric positive definite
# matrix using function 'unstructured'
corstr <- "unstr"
corpar <- unstructured(5)
corpar
Y <- rcounts(N=N, margins=margins, mu=mu, phi=phi, omega=omega, psi=psi,
             corstr=corstr, corpar=corpar)
cor(Y)
```

`rcounts.reg` *Generate correlated count random variables with individual parameters for each cluster*

Description

'rcounts.reg' is used to sample high-dimensional correlated count random variables with approximate prespecified Pearson correlation and exact margins.

Usage

```
rcounts.reg(N, margins, mu, phi=matrix(NA,N,length(margins)),
            omega=matrix(NA,N,length(margins)),
            psi=matrix(NA,N,length(margins)), corstr, corpar, conv=0.01)
```

Arguments

`N`	number of observations to be generated per margin (should be at least 500).
`margins`	Vector of margin tokens. Its length T is the dimension. See details.
`mu`	Matrix of dimension N x T of means for the Poisson, GP, ZIP, ZIGP and NB margins.
`phi`	Matrix of dimension N x T of dispersion parameters for the GP, and ZIGP margins. For Poisson, ZIP and NB margins, an 'NA' can be provided.
`omega`	Matrix of dimension N x T of zero-inflation parameters for the ZIP and ZIGP margins. For Poisson, GP and NB margins, an 'NA' can be provided.
`psi`	Matrix of dimension N x T of size parameters for the NB margins. For Poisson, GP, ZIP and ZIGP margins, an 'NA' can be provided.
`corstr`	Correlation structure. Can be 'ex' for exchangeable, 'AR1' for AR(1) and 'unstr' for unstructured.
`corpar`	Correlation parameter. Scalar correlation for 'ex' and 'AR1' and matrix of dimension TxT for 'unstr'.
`conv`	Convergence criterion

Details

The entries in 'margins' can be specified as 'Poi' for Poisson, 'GP' for generalized Poisson, 'ZIP' for zero-inflated Poisson, 'ZIGP' for zero-inflated generalized Poisson and 'NB' for negative-binomial.

NOTE: there is a tradeoff between too small N (decreasing accuracy of the resulting correlation) and too high N (dramatically increasing computation time).

Value

The function will return a matrix of counts of dimension N x T.

Author(s)

Vinzenz Erhardt

Examples

```
N <- 500

# bivariate example
margins <- c("ZIGP","GP")
mu <-    matrix(runif(N*2,10,20),N,2)
phi <-   matrix(runif(N*2,1,3),N,2)
omega <- matrix(c(runif(N,0,.3),rep(NA,N)),N,2)
corstr <- "ex"
corpar <- .5
Y <- rcounts.reg(N=N, margins=margins, mu=mu, phi=phi, omega=omega,
                 corstr=corstr, corpar=corpar)
cor(Y)
```

---

unstructured *Generate random correlation matrix*

Description

'unstructured' generates a random correlation matrix of dimension T with random entries. To ensure positive definiteness, a matrix of partial correlations with random entries uniform on [-0.9, 0.9] will be generated an the corresponding correlation matrix be calculated from it using a bijective recursive relation between them.

Usage

```
unstructured(dimension)
```

Arguments

dimension	Dimension T of the correlation matrix.

Value

A correlation matrix of dimension T x T.

Examples

```
unstructured(10)
```

Index