Package ‘SimCorMultRes’

February 15, 2013

Type Package

Title Simulates Correlated Multinomial Responses

Version 1.0

Date 2012-11-12

Author Anestis Touloumis

Maintainer Anestis Touloumis <anestis@ebi.ac.uk>

Depends R(>= 2.15.0), mvtnorm, evd

Suggests multgee

Description This package simulates correlated multinomial responses utilizing threshold approaches and assuming a cumulative link model or a baseline category logit model for the univariate marginal probabilities.

License GPL-3

LazyLoad yes

Repository CRAN

Date/Publication 2012-11-12 18:37:14

NeedsCompilation no

R topics documented:

SimCorMultRes-package .......................................................... 2
rmult.bcl ............................................................................. 2
rmult.clm ............................................................................ 4
rnorta .................................................................................... 6

Index 8
SimCorMultRes-package  Simulating Correlated Multinomial Responses

Description

Functions to simulate correlated nominal or ordinal multinomial responses assuming a cumulative link model or a baseline category logit model for the univariate marginal probabilities.

Details

The simulated correlated multinomial responses are drawn as realizations of an underlying regression model for continuous random vectors. The correlation structure is expressed in terms of the latent random vectors. For ordinal response categories, the multinomial responses are produced by chopping off the corresponding latent variables (McCullagh, 1980), and for nominal response categories by utilizing the principle of maximum random utility (McFadden, 1973). The functions are suitable to generate correlated ordinal or nominal multinomial responses when a cumulative link model or a baseline category logit model, respectively, holds for the univariate marginal probabilities.

The package can ease the simulation of correlated multinomial responses when the interest lies on estimating the marginal regression coefficient parameters, e.g. using a GEE model. An example is given in Touloumis, Agresti and Kateri (2012).

Author(s)

Anestis Touloumis <anestis@ebi.ac.uk>
Maintainer: Anestis Touloumis <anestis@ebi.ac.uk>

References


rmult.bcl  Simulating Correlated Nominal Multinomial Responses

Description

This function generates correlated nominal multinomial responses assuming a baseline category logit model for the univariate marginal probabilities.
**Usage**

```r
rmult.bcl(clsize, ncategories, lin.pred, cor.matrix)
```

**Arguments**

- `clsize`: integer indicating the common cluster size.
- `ncategories`: integer indicating the number of response categories.
- `lin.pred`: matrix with `clsize` x `ncategories` columns. See details for more info.
- `cor.matrix`: matrix indicating the correlation matrix of the multivariate normal distribution in the NORTA method.

**Details**

1. Each multinomial response is obtained by utilizing the principle of maximum random utility.
2. The row dimension of `lin.pred` equals the sample size. Rows correspond to subjects. Columns $t*(\text{ncategories})+1, \ldots, 2*\text{ncategories}$ correspond to the $1-\ldots, \text{ncategories}$-th category specific linear predictor at occasion $t$ respectively.
3. The algorithm forces `cor.matrix` to respect the local independence assumption between the alternatives.

**Value**

- `Ysim`: the correlated nominal multinomial responses.
- `correlation`: the correlation matrix of the multivariate normal distribution in the NORTA method.
- `rlatent`: the latent random vectors after applying the NORTA method.

**Author(s)**

Anestis Touloumis

**References**


**See Also**

`rnorta`, `rmult.clm`
Examples

```r
## Simulating 500 correlated nominal multinomial responses assuming a baseline category logit model for the univariate marginal probabilities.
## In particular we assume that all the category-specific intercepts are equal to 1 and all the category-specific coefficients are equal to 2.
## Further we sample a cluster specific covariate from a standard normal distribution.

set.seed(1)
N <- 500
ncategories <- 4
clustersize <- 3
Xmat <- matrix(rnorm(N),N,ncategories)
betas <- c(1,2,1,2)
linpred <- matrix(c(betas[c(2,4,6)],0),N,4,byrow=TRUE)*Xmat + matrix(c(betas[c(1,3,5)],0),N,4,byrow=TRUE)
linpred <- matrix(linpred,N,ncategories*clustersize)
cormat <- toeplitz(c(1,rep(0,3),rep(c(0.9,0,0,0,0,0,0),2)))
```

```r
## We fit a GEE model to illustrate that the regression coefficients can be estimated
library(multgee)
Y <- rmult.clm(clsize=3,ncategories=4,lin.pred=linpred,cor.matrix=cormat)$Ysim
data <- cbind(c(t(Y)),c(t(Xmat[, -ncategories])))
data <- data.frame(data)
data$id <- rep(1:N,each=clustersize)
data$time <- rep(1:clustersize,N)
colnames(data) <- c("y","x","id","time")
fitmod <- nomLORgee(y~x,id="id",repeated="time",data=data,add=0.01)
```

```r
## The GEE estimates of the regression coefficients
coef(fitmod)
```

---

### rmult.clm

**Simulating Correlated Ordinal Multinomial Responses**

**Description**

This function generates correlated ordinal multinomial responses assuming a cumulative link model for the univariate marginal probabilities.

**Usage**

```r
rmult.clm(clsize, lin.pred, corr, cuts, link = "probit")
```

**Arguments**

- **clsize**: integer indicating the common cluster size.
- **lin.pred**: matrix with clsize columns. See details for more info.
**rmult.clm**

- **corr**: matrix or constant describing the latent correlation structure. See details for more info.
- **cuts**: vector containing the intercepts of the cumulative link model, and -Inf and Inf as first and last elements respectively.
- **link**: character string indicating the link function of the cumulative link model. Options include "probit", "logit", "clogclog" or "cauchit".

**Details**

1. Each multinomial response takes the value \( j \) if and only if the corresponding latent random variable is on \( (\text{cuts}_{j-1}, \text{cuts}_j] \).
2. The row dimension of \( \text{lin.pred} \) equals the sample size.
3. \( \text{lin.pred} \) contains the linear predictor of the cumulative link model after subtracting the intercepts.
4. If \( \text{corr} \) is a matrix, then the NORTA method is employed and the correlation matrix corresponds to that of the multivariate normal distribution. Otherwise, it must be a constant on \([0,1)\). For the cloglog link, it generates latent random vectors from a multivariate Gumbel distribution with correlation parameter \( \text{corr} \). For the logit link, it generates the latent random vectors as the difference of two independent random vectors that follow a multivariate Gumbel distribution with correlation parameter \( \text{corr} \).

**Value**

- \( \text{Ysim} \): the correlated ordinal multinomial responses.
- \( \text{correlation} \): the latent correlation matrix.
- \( \text{rlatent} \): the underlying latent random vectors.

**Author(s)**

Anestis Touloumis

**References**


**See Also**

rnorta, rmult.bcl
Examples

## Simulating 500 correlated ordinal multinomial responses assuming a probit cumulative
## link model for the univariate marginal probabilities. The model is described in Touloumis,
## Agresti and Kateri (2012) with a toeplitz underlying correlation structure and a
time-varying covariates design.

set.seed(12345)
N <- 500
collectorsize <- 4
intercepts <- c(-Inf,-1.5,-0.5,0.5,1.5,Inf)
cormat <- toeplitz(c(1,0.85,0.5,0.15))
linpred <- rmvnorm(N,sigma=toeplitz(c(1,rep(0.85,collectorsize-1))))

Y <- rmult.clm(clsize=collectorsize,lin.pred=linpred,corr=cormat,
cuts=intercepts,link="probit")$Ysim
data <- cbind(t(Y),t(linpred))
data <- data.frame(data)
data$id <- rep(1:N,each=collectorsize)
data$time <- rep(1:collectorsize,N)
colnames(data) <- c("y","x","id","time")

## We fit a GEE model to illustrate that the regression coefficients can be estimated
library(multgee)
fitmod <- ordLORgee(y~x,data=data,id="id",repeated="time",link="probit",IM="cholesky")

## The GEE estimates of the regression coefficients
coef(fitmod)

rnorta

Simulating Continuous Random Vectors Using The NORTA Method

Description

This function generates continuous random vectors with prescribed univariate marginal distributions
using the NORTA method.

Usage

rnorta(R = R, cor.matrix = cor.matrix, distr = "normal")

Arguments

R  integer indicating the sample size.
cor.matrix  matrix indicating the correlation matrix of the multivariate normal distribution
in the NORTA method.
distr  character string indicating the desired univariate marginal distributions. Options
include "normal", "logistic", "extreme" or "cauchy".
Details

Checks are made to ensure that `cor.matrix` is a valid semi-positive correlation matrix. The semi-positiveness of `cor.matrix` is determined by the eigenvalues.

Value

Returns R random vectors of size equal to the row dimension of `cor.matrix` and univariate marginal distributions `distr`.

Author(s)

Anestis Touloumis

References


Examples

```r
## An example with marginal standard logistic distributions.
sset.seed(1)
logit.sim <- rnorta(R = 1000, cor.matrix = toeplitz(c(1,rep(0.8,3))), distr = "logistic")

## The following illustrates how the NORTA method works.
sset.seed(1)
norm.sim <- rmvnorm(1000, sigma = toeplitz(c(1,rep(0.8,3))))
al(l(logit.sim==qlogis(pnorm(norm.sim))))
```
Index

*Topic** package
  SimCorMultRes-package, 2

rmult.bcl, 2, 5
rmult.clm, 3, 4
rnorta, 3, 5, 6

SimCorMultRes (SimCorMultRes-package), 2
SimCorMultRes-package, 2