Package ‘SimCorMultRes’

July 10, 2018

Type Package

Title Simulates Correlated Multinomial Responses

Description
  Simulates correlated multinomial responses conditional on a marginal model specification.

Version 1.6.0

Depends R(>= 2.15.0)

Imports evd, methods, stats

Suggests bookdown, gee, knitr, markdown, multgee(>= 1.2)

URL http://github.com/AnestisTouloumis/SimCorMultRes

BugReports http://github.com/AnestisTouloumis/SimCorMultRes/issues

License GPL-3

VignetteBuilder knitr

R topics documented:

  SimCorMultRes-package ........................................... 2
  rbin ................................................................. 3
  rmult.bcl .......................................................... 5
  rmult.clm ........................................................... 8
  rmult.crm ........................................................... 10
  rmorta ............................................................... 13
  rsmvnorm .............................................................. 15

Index 16
SimCorMultRes-package  Simulating Correlated Multinomial Responses

Description

Functions to simulate correlated multinomial responses (three or more nominal or ordinal response categories) and correlated binary responses subject to a marginal model specification.

Details

The simulated correlated binary or multinomial responses are drawn as realizations of a latent regression model for continuous random vectors with the correlation structure expressed in terms of the latent correlation.

For an ordinal response scale, the multinomial variables are simulated conditional on either a marginal cumulative link model (rmult.clm) or a marginal continuation-ratio model (rmult.crm).

For a nominal response scale, the multinomial responses are simulated conditional on a marginal baseline-category logit model (rmult.bcl).

Correlated binary responses are simulated using the function rbin.

The threshold approaches that give rise to the implemented marginal models are fully described in Touloumis (2016).

The formulae are easier to read from either the Vignette or the Reference Manual (both available here).

Author(s)

Anestis Touloumis

Maintainer: Anestis Touloumis <A.Touloumis@brighton.ac.uk>

References


Simulating Correlated Binary Responses Conditional on a Marginal Model Specification

Description

Simulates correlated binary responses assuming a regression model for the marginal probabilities.

Usage

```r
rbin(clsize = clsize, intercepts = intercepts, betas = betas,
     xformula = formula(xdata), xdata = parent.frame(), link = "logit",
     cor.matrix = cor.matrix, rlatent = NULL)
```

Arguments

- `clsize`: integer indicating the common cluster size.
- `intercepts`: numerical (or numeric vector of length `clsize`) containing the intercept(s) of the marginal model.
- `betas`: numerical vector or matrix containing the value of the marginal regression parameter vector associated with the covariates (i.e., excluding `intercepts`).
- `xformula`: formula expression as in other marginal regression models but without including a response variable.
- `xdata`: optional data frame containing the variables provided in `xformula`.
- `link`: character string indicating the link function in the marginal model. Options include 'probit', 'logit', 'cloglog' or 'cauchit'. Required when `rlatent = NULL`.
- `cor.matrix`: matrix indicating the correlation matrix of the multivariate normal distribution when the NORTA method is employed (`rlatent = NULL`).
- `rlatent`: matrix with `clsize` columns containing realizations of the latent random vectors when the NORTA method is not preferred. See details for more info.

Details

The formulae are easier to read from either the Vignette or the Reference Manual (both available here).

The assumed marginal model is

\[ Pr(Y_{it} = 1 | x_{it}) = F(\beta_{i0} + \beta'_i x_{it}) \]
where $F$ is the cumulative distribution function determined by link. For subject $i$, $Y_{it}$ is the $t$-th binary response and $x_{it}$ is the associated covariates vector. Finally, $\beta_{t0}$ and $\beta_t$ are the intercept and regression parameter vector at the $t$-th measurement occasion.

The binary response $Y_{it}$ is obtained by extending the approach of Emrich and Piedmonte (1991) as suggested in Touloumis (2016).

When $\beta_{t0} = \beta_0$ for all $t$, then intercepts should be provided as a single number. Otherwise, intercepts must be provided as a numeric vector such that the $t$-th element corresponds to the intercept at measurement occasion $t$.

betas should be provided as a numeric vector only when $\beta_t = \beta$ for all $t$. Otherwise, betas must be provided as a numeric matrix with clsize rows such that the $t$-th row contains the value of $\beta_t$. In either case, betas should reflect the order of the terms implied by xformula.

The appropriate use of xformula is xformula = ~ covariates, where covariates indicate the linear predictor as in other marginal regression models.

The optional argument xdata should be provided in “long” format.

The NORTA method is the default option for simulating the latent random vectors denoted by $e_{it}^B$ in Touloumis (2016). To import simulated values for the latent random vectors without utilizing the NORTA method, the user can employ the rlatent argument. In this case, element $(i, t)$ of rlatent represents the realization of $e_{it}^B$.

Value

Returns a list that has components:

- Ysim the simulated binary responses. Element $(i,t)$ represents the realization of $Y_{it}$.
- simdata a data frame that includes the simulated response variables (y), the covariates specified by xformula, subjects’ identities (id) and the corresponding measurement occasions (time).
- rlatent the latent random variables denoted by $e_{it}^B$ in Touloumis (2016).

Author(s)

Anestis Touloumis

References


See Also

`rmult.bcl` for simulating correlated nominal responses, `rmult.clm` and `rmult.crm` for simulating correlated ordinal responses.

Examples

```r
## See Example 3.4 in the Vignette.
set.seed(123)
N <- 5000
cls <- 4
intercepts <- 0
betas <- 0.2
cor.matrix <- toeplitz(c(1, 0.9, 0.9, 0.9))
x <- rep(rnorm(N), each = cls)
CorBinRes <- rbin(clsize = cls, intercepts = intercepts, betas = betas,
xformula = ~x, cor.matrix = cor.matrix, link = 'probit')
library(gee)
binGEEmod <- gee(y ~ x, family = binomial('probit'), id = id, data = CorBinRes$simdata)
summary(binGEEmod)$coefficients

## See Example 3.5 in the Vignette.
set.seed(8)
library(evd)
rlatent1 <- rmvevd(N, dep = sqrt(1 - 0.9), model = 'log', d = cls)
rlatent2 <- rmvevd(N, dep = sqrt(1 - 0.9), model = 'log', d = cls)
rlatent <- rlatent1 - rlatent2
CorBinRes <- rbin(clsize = cls, intercepts = intercepts, betas = betas,
xformula = ~x, latent = rlatent)
binGEEmod <- gee(y ~ x, family = binomial('logit'), id = id, data = CorBinRes$simdata)
summary(binGEEmod)$coefficients
```

---

**rmult.bcl**  
**Simulating Correlated Nominal Responses Conditional on a Marginal Baseline-Category Logit Model Specification**

**Description**

Simulates correlated nominal responses assuming a baseline-category logit model for the marginal probabilities.

**Usage**

```r
rmult.bcl(clsize = clsize, ncategories = ncategories, betas = betas,
xformula = formula(xdata), xdata = parent.frame(),
cor.matrix = cor.matrix, rlatent = NULL)
```
Arguments

clsize  integer indicating the common cluster size.
ncategories  integer indicating the number of nominal response categories.
betas  numerical vector or matrix containing the value of the marginal regression parameter vector.
xformula  formula expression as in other marginal regression models but without including a response variable.
xdata  optional data frame containing the variables provided in xformula.
cor matrix  matrix indicating the correlation matrix of the multivariate normal distribution when the NORTA method is employed (rlatent = null).
rlatent  matrix with (clsize * ncategories) columns containing realizations of the latent random vectors when the NORTA method is not preferred. See details for more info.

Details

The formulae are easier to read from either the Vignette or the Reference Manual (both available here).

The assumed marginal baseline category logit model is

\[
\log \frac{Pr(Y_{it} = j|x_{it})}{Pr(Y_{it} = J|x_{it})} = (\beta_{tj0} - \beta_{tJ0}) + (\beta^*_{tj} - \beta^*_{tJ})x_{it} = \beta^*_{tj0} + \beta^*_{tj}x_{it}
\]

For subject \(i\), \(Y_{it}\) is the \(t\)-th nominal response and \(x_{it}\) is the associated covariates vector. Also \(\beta_{tj0}\) is the \(j\)-th category-specific intercept at the \(t\)-th measurement occasion and \(\beta_{tj}\) is the \(j\)-th category-specific regression parameter vector at the \(t\)-th measurement occasion.

The nominal response \(Y_{it}\) is obtained by extending the principle of maximum random utility (McFadden, 1974) as suggested in Touloumis (2016).

betas should be provided as a numeric vector only when \(\beta_{tj0} = \beta_{j0}\) and \(\beta_{tj} = \beta_{j}\) for all \(t\). Otherwise, betas must be provided as a numeric matrix with clsize rows such that the \(t\)-th row contains the value of \((\beta_{t10}, \beta_{t1}, \beta_{t20}, \beta_{t2}, ..., \beta_{tJ0}, \beta_{tJ})\). In either case, betas should reflect the order of the terms implied by xformula.

The appropriate use of xformula is xformula = ~ covariates, where covariates indicate the linear predictor as in other marginal regression models.

The optional argument xdata should be provided in “long” format.

The NORTA method is the default option for simulating the latent random vectors denoted by \(e_{itj}^{NO}\) in Touloumis (2016). In this case, the algorithm forces cor matrix to respect the assumption of choice independence. To import simulated values for the latent random vectors without utilizing the NORTA method, the user can employ the rlatent argument. In this case, row \(i\) corresponds to subject \(i\) and columns \((t-1)*ncategories + 1,...,t*ncategories\) should contain the realization of \(e_{it1}^{NO},...,e_{itJ}^{NO}\), respectively, for \(t = 1,...,clsize\).
Value

Returns a list that has components:

- **Ysim**
  - the simulated nominal responses. Element \((i,t)\) represents the realization of \(Y_{it}\).

- **simdata**
  - a data frame that includes the simulated response variables (y), the covariates specified by `xformula`, subjects' identities (id) and the corresponding measurement occasions (time).

- **rlatent**
  - the latent random variables denoted by \(e^{NO}_{it}\) in Touloumis (2016).

Author(s)

Anestis Touloumis

References


See Also

- `rbin` for simulating correlated binary responses, `rmult.clm` and `rmult.crm` for simulating correlated ordinal responses.

Examples

```r
## See Example 3.1 in the Vignette.
betas <- c(1, 3, 2, 1.25, 3.25, 1.75, 0.75, 2.75, 2.25, 2, 0, 0)
N <- 500
ncategories <- 4
clsize <- 3
set.seed(1)
x1 <- rep(rnorm(N), each = clsize)
x2 <- rnorm(N * clsize)
data <- data.frame(x1, x2)
cor.matrix <- kronecker(toeplitz(c(1, rep(0.95, clsize - 1)))), diag(ncategories))
CorNorRes <- rmult.bcl(clsize = clsize, ncategories = ncategories, betas = betas,
                        xformula = ~x1 + x2, xdata = xdata, cor.matrix = cor.matrix)
suppressPackageStartupMessages(library('multgee'))
```
rmult.clm

Simulating Correlated Ordinal Responses Conditional on a Marginal Cumulative Link Model Specification

Description

Simulates correlated ordinal responses assuming a cumulative link model for the marginal probabilities.

Usage

```r
rmult.clm(clsize = clsize, intercepts = intercepts, betas = betas, 
xformula = formula(xdata), xdata = parent.frame(), link = "logit", 
cor.matrix = cor.matrix, rlatent = NULL)
```

Arguments

- **clsize**: integer indicating the common cluster size.
- **intercepts**: numerical vector or matrix containing the intercepts of the marginal cumulative link model.
- **betas**: numerical vector or matrix containing the value of the marginal regression parameter vector associated with the covariates (i.e., excluding intercepts).
- **xformula**: formula expression as in other marginal regression models but without including a response variable.
- **xdata**: optional data frame containing the variables provided in xformula.
- **link**: character string indicating the link function in the marginal cumulative link model. Options include 'probit', 'logit', 'cloglog' or 'cauchit'. Required when rlatent = NULL.
- **cor.matrix**: matrix indicating the correlation matrix of the multivariate normal distribution when the NORTA method is employed (rlatent = NULL).
- **rlatent**: matrix with clsize columns containing realizations of the latent random vectors when the NORTA method is not preferred. See details for more info.

Details

The formulae are easier to read from either the Vignette or the Reference Manual (both available here).

The assumed marginal cumulative link model is

\[
Pr(Y_{it} \leq j|x_{it}) = F(\beta_{j0} + \beta'_1 x_{it})
\]
where $F$ is the cumulative distribution function determined by link. For subject $i$, $Y_{it}$ is the $t$-th ordinal response and $x_{it}$ is the associated covariates vector. Finally, $\beta_{tj0}$ is the $j$-th category-specific intercept at the $t$-th measurement occasion and $\beta_{tj}$ is the $j$-th category-specific regression parameter vector at the $t$-th measurement occasion.

The ordinal response $Y_{it}$ is obtained by extending the approach of McCullagh (1980) as suggested in Touloumis (2016).

When $\beta_{tj0} = \beta_{j0}$ for all $t$, then intercepts should be provided as a numerical vector. Otherwise, intercepts must be a numerical matrix such that row $t$ contains the category-specific intercepts at the $t$-th measurement occasion.

bets should be provided as a numeric vector only when $\beta_t = \beta$ for all $t$. Otherwise, betas must be provided as a numeric matrix with clsize rows such that the $t$-th row contains the value of $\beta_t$.

In either case, betas should reflect the order of the terms implied by xformula.

The appropriate use of xformula is xformula = ~ covariates, where covariates indicate the linear predictor as in other marginal regression models.

The optional argument xdata should be provided in “long” format.

The NORTA method is the default option for simulating the latent random vectors denoted by $e_{i1}^{O}$ in Touloumis (2016). To import simulated values for the latent random vectors without utilizing the NORTA method, the user can employ the rlatent argument. In this case, element $(i, t)$ of rlatent represents the realization of $e_{i1}^{O}$.

Value

Returns a list that has components:

- Ysim: the simulated ordinal responses. Element $(i, t)$ represents the realization of $Y_{it}$.
- simdata: a data frame that includes the simulated response variables (y), the covariates specified by xformula, subjects’ identities (id) and the corresponding measurement occasions (time).
- rlatent: the latent random variables denoted by $e_{i1}^{O}$ in Touloumis (2016).

Author(s)

Anestis Touloumis

References


See Also

`rmult.bcl` for simulating correlated nominal responses, `rmult.crm` for simulating correlated ordinal responses and `rbin` for simulating correlated binary responses.

Examples

```r
## See Example 3.2 in the Vignette.
set.seed(12345)
N <- 500
classize <- 4
intercepts <- c(-1.5, -0.5, 0.5, 1.5)
betas <- matrix(c(1, 2, 3, 4, 4, 1), 2, 3)
x <- rep(rnorm(N), each = classize)
cor.matrix <- toeplitz(c(1, 0.85, 0.5, 0.15))
CorOrdRes <- rmult.crm(classize = classize, intercepts = intercepts, betas = betas, xformula = ~x, cor.matrix = cor.matrix, link = 'probit')
head(CorOrdRes$simdata, n = 8)

## Same sampling scheme except that the parameter vector is now time-stationary.
set.seed(12345)
x <- rep(rnorm(N), each = classize)
CorOrdRes <- rmult.crm(classize = classize, betas = 1, xformula = ~x, cor.matrix = toeplitz(c(1, 0.85, 0.5, 0.15)), intercepts = c(-1.5, -0.5, 0.5, 1.5), link = 'probit')
## Fit a GEE model (Touloumis et al., 2013) to estimate the regression
## coefficients.
library(multgee)
fitmod <- ordLORgee(y ~ x, id = id, repeated = time, link = 'probit', data = CorOrdRes$simdata)
coef(fitmod)
```

---

**Description**

Simulates correlated ordinal responses assuming a continuation-ratio model for the marginal probabilities.

**Usage**

```r
rmult.crm(classize = classize, intercepts = intercepts, betas = betas,
          xformula = formula(xdata), xdata = parent.frame(), link = "logit",
          cor.matrix = cor.matrix, rlatent = NULL)
```
Arguments

- `clsize` integer indicating the common cluster size.
- `intercepts` numerical vector or matrix containing the intercepts of the marginal continuation-ratio model.
- `betas` numerical vector or matrix containing the value of the marginal regression parameter vector associated with the covariates (i.e., excluding intercepts).
- `xformula` formula expression as in other marginal regression models but without including a response variable.
- `xdata` optional data frame containing the variables provided in `xformula`.
- `link` character string indicating the link function of the marginal continuation-ratio model. Options include 'probit', 'logit', 'cloglog' or 'cauchit'. Required when `rlatent = NULL`.
- `cor.matrix` matrix indicating the correlation matrix of the multivariate normal distribution when the NORTA method is employed (`rlatent = NULL`).
- `rlatent` matrix with `clsize` rows and `ncategories` columns containing realizations of the latent random vectors when the NORTA method is not employed. See details for more info.

Details

The formulae are easier to read from either the Vignette or the Reference Manual (both available here).

The assumed marginal continuation-ratio model is

$$ Pr(Y_{it} = j | Y_{it} \geq j, x_{it}) = F(\beta_{tj0} + \beta_j'x_{it}) $$

where $F$ is the cumulative distribution function determined by `link`. For subject $i$, $Y_{it}$ is the $t$-th multinomial response and $x_{it}$ is the associated covariates vector. Finally, $\beta_{tj0}$ is the $j$-th category-specific intercept at the $t$-th measurement occasion and $\beta_j$ is the $j$-th category-specific regression parameter vector at the $t$-th measurement occasion.

The ordinal response $Y_{it}$ is determined by extending the latent variable threshold approach of Tutz (1991) as suggested in Touloumis (2016).

When $\beta_{tj0} = \beta_j0$ for all $t$, then intercepts should be provided as a numerical vector. Otherwise, intercepts must be a numerical matrix such that row $t$ contains the category-specific intercepts at the $t$-th measurement occasion.

`betas` should be provided as a numeric vector only when $\beta_t = \beta$ for all $t$. Otherwise, `betas` must be provided as a numeric matrix with `clsize` rows such that the $t$-th row contains the value of $\beta_t$. In either case, `betas` should reflect the order of the terms implied by `xformula`.

The appropriate use of `xformula` is `xformula = ~ covariates`, where `covariates` indicate the linear predictor as in other marginal regression models.

The optional argument `xdata` should be provided in “long” format.

The NORTA method is the default option for simulating the latent random vectors denoted by $e^{O_{itj}}$ in Touloumis (2016). In this case, the algorithm forces `cor.matrix` to respect the local independence assumption. To import simulated values for the latent random vectors without utilizing the NORTA
method, the user can employ the \texttt{rlatent} argument. In this case, row \(i\) corresponds to subject \(i\) and columns \((t-1) \ast \text{ncategories} + 1, \ldots, t \ast \text{ncategories}\) should contain the realization of \(e_{it1}, \ldots, e_{itJ}\), respectively, for \(t = 1, \ldots, \text{clsize}\).

**Value**

Returns a list that has components:

- \texttt{Ysim} the simulated ordinal responses. Element \((i,t)\) represents the realization of \(Y_{it}\).
- \texttt{simdata} a data frame that includes the simulated response variables \((y)\), the covariates specified by \texttt{xformula}, subjects’ identities \((id)\) and the corresponding measurement occasions \((time)\).
- \texttt{rlatent} the latent random variables denoted by \(e_{it}^O\) in \cite{Touloumis(2016)}.

**Author(s)**

Anestis Touloumis

**References**


**See Also**

\texttt{rmult.bcl} for simulating correlated nominal responses, \texttt{rmult.clm} for simulating correlated ordinal responses and \texttt{rbin} for simulating correlated binary responses.

**Examples**

```r
## See Example 3.3 in the Vignette.
set.seed(1)
N <- 500
csize <- 4
intercepts <- c(-1.5, -0.5, 0.5, 1.5)
betas <- 1
x <- rnorm(N \* csize)
ncategories <- 5
cor.matrix <- diag(1, (nccategories - 1) \* csize) + kronecker(toeplitz(c(0, rep(0.24, ncategories - 2))), matrix(1, csize, csize))
CorOrdRes <- rmult.crm(clsize = csize, intercepts = intercepts, betas = betas,

```

rnorta

Simulating Random Vectors using the NORTA Method

Description

Utility function to simulate random vectors with predefined marginal distributions via the NORTA method.

Usage

rnorta(R = R, cor.matrix = cor.matrix, distr = distr, qparameters = NULL)

Arguments

R integer indicating the sample size.

cor.matrix matrix indicating the correlation matrix of the multivariate normal distribution employed in the NORTA method.

distr character string vector of length ncol(cor.matrix) naming the quantile functions of the desired marginal distributions.

qparameters list of ncol(cor.matrix) lists indicating the parameter values of the quantile functions specified by distr.

Details

Checks are made to ensure that cor.matrix is a positive definite correlation matrix. The positive definiteness of cor.matrix is assessed via eigenvalues.

The $t$-th character string in distr indicates the quantile function of the $t$-th marginal distribution. See Distributions for the most common distributions. Quantile functions supported by other R packages are allowed provided that these packages have been uploaded first. However, note that no checks are made to ensure that the character strings in distr correspond to valid names of quantile functions.

If qparameters = NULL then the default parameter values for the quantile functions specified by distr are used. Otherwise, qparameters should be provided as a list of ncol(cor.matrix) lists such that the $t$-th list contains the desired parameter values of the $t$-th quantile function.

Value

Returns R random vectors of size ncol(cor.matrix) with marginal distributions specified by distr (and qparameters).
Author(s)

Anestis Touloumis

References


Examples

```r
## An example with standard logistic as marginal distribution.
set.seed(1)
R <- 1000
LatentCorrelation <- toeplitz(c(1, rep(0.8, 2)))
LatentCorrelation
CommonMarginals <- rep('qlogis', 3)
SimLogistic <- rnorta(r = R, cor.matrix = LatentCorrelation, distr = CommonMarginals)

## The following lines exemplify the NORTA method.
set.seed(1)
SimNormal <- rsrnorm(R = R, cor.matrix = LatentCorrelation)
all(SimLogistic == qlogis(pnorm(SimNormal)))

## Change the marginal distributions to standard normal, standard logistic and standard extreme value distribution.
set.seed(1)
DiffMarginals <- c('qnorm', 'qlogis', 'qgumbel')
SimDiffMars <- rnorta(R = R, cor.matrix = LatentCorrelation, distr = DiffMarginals)
cor(SimDiffMars)
colMeans(SimDiffMars)
apply(SimDiffMars, 2, sd)

## Same as above but using parameter values other than the default ones.
set.seed(1)
qpars <- list(c(mean = 1, sd = 9), c(location = 2, scale = 1), c(loc = 3, scale = 1))
SimDiffMars2 <- rnorta(R = R, cor.matrix = LatentCorrelation, distr = DiffMarginals, qparameters = qpars)
cor(SimDiffMars2)
colMeans(SimDiffMars2)
apply(SimDiffMars2, 2, sd)
```
**rsmvnorm**

**Simulating Continuous Random Vectors from a Multivariate Normal Distribution**

**Description**

Utility function to simulate continuous random vectors from a multivariate normal distribution such that all marginal distributions are univariate standard normal.

**Usage**

```r
rsmvnorm(r = rL corNmatrix = corNmatrix)
```

**Arguments**

- `r` integer indicating the sample size.
- `corNmatrix` matrix indicating the correlation matrix of the multivariate normal distribution.

**Details**

Checks are made to ensure that `corNmatrix` is a positive definite correlation matrix. The positive definiteness of `corNmatrix` is assessed via eigenvalues.

**Value**

Returns `r` random vectors of size `ncol(corNmatrix)`.

**Author(s)**

Anestis Touloumis

**Examples**

```r
## Simulating 10000 bivariate random vectors with correlation parameter
## equal to 0.4.
set.seed(1)
R <- 10000
cor.matrix <- toeplitz(c(1, 0.4))
SimBivariateNormal <- rsmvnorm(R = R, cor.matrix = cor.matrix)
colMeans(SimBivariateNormal)
apply(SimBivariateNormal, 2, sd)
cor(SimBivariateNormal)
```
Index

*Topic **package**
  SimCorMultRes-package, 2

Distributions, 13

rbin, 2, 3, 7, 10, 12
rmult.bcl, 2, 5, 5, 10, 12
rmult.clm, 2, 5, 7, 8, 12
rmult.crm, 2, 5, 7, 10, 10
rnorta, 13
rsmvnorm, 15

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