

# Package: SimCorMultRes (via r-universe)

September 30, 2026

**Type** Package

**Title** Simulates Correlated Multinomial Responses

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

**Version** 1.9.1

**Depends** R(>= 2.15.0)

**Imports** evd, methods, stats

**Suggests** bookdown, covr, gee, knitr, multgee (>= 1.2), rmarkdown, R.rsp, testthat

**URL** <https://github.com/AnestisTouloumis/SimCorMultRes>

**BugReports** <https://github.com/AnestisTouloumis/SimCorMultRes/issues>

**License** GPL-3

**VignetteBuilder** knitr, R.rsp

**RoxygenNote** 7.3.1

**Roxygen** list(old\_usage = TRUE)

**Encoding** UTF-8

**LazyData** true

**Repository** <https://anestistouloumis.r-universe.dev>

**Date/Publication** 2024-03-18 21:08:10 UTC

**RemoteUrl** <https://github.com/anestistouloumis/simcormultres>

**RemoteRef** HEAD

**RemoteSha** a3843e9ae02eccc6772597ba174aff82ca550a98

## Contents

|                     |   |
|---------------------|---|
| rbin . . . . .      | 2 |
| rmult.acl . . . . . | 4 |
| rmult.bcl . . . . . | 7 |
| rmult.clm . . . . . | 9 |

|                      |    |
|----------------------|----|
| rmult.crm . . . . .  | 12 |
| rnorta . . . . .     | 14 |
| rsmvnorm . . . . .   | 16 |
| simulation . . . . . | 17 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>19</b> |
|--------------|-----------|

---

|      |                                                                                             |
|------|---------------------------------------------------------------------------------------------|
| rbin | <i>Simulating Correlated Binary Responses Conditional on a Marginal Model Specification</i> |
|------|---------------------------------------------------------------------------------------------|

---

## Description

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

## Usage

```
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', 'cauchit' or 'identity'. 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

$$Pr(Y_{it} = 1|x_{it}) = F(\beta_{t0} + \beta'_t x_{it})$$

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:

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

## Author(s)

Anestis Touloumis

## References

- Cario, M. C. and Nelson, B. L. (1997) *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.
- Emrich, L. J. and Piedmonte, M. R. (1991) A method for generating high-dimensional multivariate binary variates. *The American Statistician* **45**, 302–304.
- Li, S. T. and Hammond, J. L. (1975) Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man and Cybernetics* **5**, 557–561.
- Touloumis, A. (2016) Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The `SimCorMultRes` Package. *The R Journal* **8**, 79–91.

## See Also

[rmult.bcl](#) for simulating correlated nominal responses, [rmult.clm](#), [rmult.crm](#) and [rmult.acf](#) for simulating correlated ordinal responses.

## Examples

```
## See Example 3.5 in the Vignette.
set.seed(123)
sample_size <- 5000
cluster_size <- 4
beta_intercepts <- 0
beta_coefficients <- 0.2
latent_correlation_matrix <- toeplitz(c(1, 0.9, 0.9, 0.9))
x <- rep(rnorm(sample_size), each = cluster_size)
simulated_binary_dataset <- rbin(clsize = cluster_size,
  intercepts = beta_intercepts, betas = beta_coefficients,
  xformula = ~x, cor.matrix = latent_correlation_matrix, link = "probit")
library(gee)
binary_gee_model <- gee(y ~ x, family = binomial("probit"), id = id,
  data = simulated_binary_dataset$simdata)
summary(binary_gee_model)$coefficients

## See Example 3.6 in the Vignette.
set.seed(8)
library(evd)
simulated_latent_variables1 <- rmvevd(sample_size, dep = sqrt(1 - 0.9),
  model = "log", d = cluster_size)
simulated_latent_variables2 <- rmvevd(sample_size, dep = sqrt(1 - 0.9),
  model = "log", d = cluster_size)
simulated_latent_variables <- simulated_latent_variables1 -
  simulated_latent_variables2
simulated_binary_dataset <- rbin(clsize = cluster_size,
  intercepts = beta_intercepts, betas = beta_coefficients,
  xformula = ~x, rlatent = simulated_latent_variables)
binary_gee_model <- gee(y ~ x, family = binomial("logit"), id = id,
  data = simulated_binary_dataset$simdata)
summary(binary_gee_model)$coefficients
```

---

rmult.acf

*Simulating Correlated Ordinal Responses Conditional on a Marginal  
Adjacent-Category Logit Model Specification*

---

## Description

Simulates correlated ordinal responses assuming an adjacent-category logit model for the marginal probabilities.

**Usage**

```
rmult.acl(clsize = clsize, intercepts = intercepts, betas = betas,
  xformula = formula(xdata), xdata = parent.frame(),
  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 adjacent-category logit model.                                                                |
| 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 adjacent-category logit model is

$$\log \frac{Pr(Y_{it} = j | x_{it})}{Pr(Y_{it} = j + 1 | x_{it})} = \beta_{tj0} + \beta'_t x_{it}$$

For subject  $i$ ,  $Y_{it}$  is the  $t$ -th ordinal 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  $\beta_t$  is the regression parameter vector at the  $t$ -th measurement occasion.

The ordinal response  $Y_{it}$  is obtained by utilizing the threshold approach described in the Vignette. This approach is based on the connection between baseline-category and adjacent-category logit models.

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  $e_{itj}^{O3}$  in the Vignette. 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  $(t-1)*ncategories+1, \dots, t*ncategories$  should contain the realization of  $e_{it1}^{O3}, \dots, e_{itJ}^{O3}$ , respectively, for  $t = 1, \dots, clsize$ .

### Value

Returns a list that has components:

|                      |                                                                                                                                                                                                                                                 |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>Ysim</code>    | the simulated nominal responses. Element $(i,t)$ represents the realization of $Y_{it}$ .                                                                                                                                                       |
| <code>simdata</code> | a data frame that includes the simulated response variables ( <code>y</code> ), the covariates specified by <code>xformula</code> , subjects' identities ( <code>id</code> ) and the corresponding measurement occasions ( <code>time</code> ). |
| <code>rlatent</code> | the latent random variables denoted by $e_{itj}^{O3}$ in the Vignette.                                                                                                                                                                          |

### Author(s)

Anestis Touloumis

### References

- Cario, M. C. and Nelson, B. L. (1997) *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.
- Li, S. T. and Hammond, J. L. (1975) Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man and Cybernetics* **5**, 557–561.
- Touloumis, A. (2016) Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The `SimCorMultRes` Package. *The R Journal* **8**, 79–91.
- Touloumis, A., Agresti, A. and Kateri, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics* **69**, 633–640.

### See Also

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

### Examples

```
## See Example 3.4 in the Vignette.
beta_intercepts <- c(3, 1, 2)
beta_coefficients <- c(1, 1)
sample_size <- 500
cluster_size <- 3
set.seed(321)
x1 <- rep(rnorm(sample_size), each = cluster_size)
x2 <- rnorm(sample_size * cluster_size)
xdata <- data.frame(x1, x2)
```

```

identity_matrix <- diag(4)
equicorrelation_matrix <- toeplitz(c(1, rep(0.95, cluster_size - 1)))
latent_correlation_matrix <- kronecker(equicorrelation_matrix,
  identity_matrix)
simulated_ordinal_dataset <- rmult.acl(clsize = cluster_size,
  intercepts = beta_intercepts, betas = beta_coefficients,
  xformula = ~ x1 + x2, xdata = xdata,
  cor.matrix = latent_correlation_matrix)
suppressPackageStartupMessages(library("multgee"))
ordinal_gee_model <- ordLORgee(y ~ x1 + x2,
  data = simulated_ordinal_dataset$simdata, id = id, repeated = time,
  LORstr = "time.exch", link = "acl")
round(coef(ordinal_gee_model), 2)

```

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

```

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  $(\beta_{t10}, \beta_{t1}, \beta_{t20}, \beta_{t2}, \dots, \beta_{tJ0}, \beta_{tJ})$ . 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  $(t-1)*ncategories+1, \dots, t*ncategories$  should contain the realization of  $e_{it1}^{NO}, \dots, e_{itJ}^{NO}$ , respectively, for  $t = 1, \dots, clsize$ .

## Value

Returns a list that has components:

|                      |                                                                                                                                                                                                                                                 |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>Ysim</code>    | the simulated nominal responses. Element $(i,t)$ represents the realization of $Y_{it}$ .                                                                                                                                                       |
| <code>simdata</code> | a data frame that includes the simulated response variables ( <code>y</code> ), the covariates specified by <code>xformula</code> , subjects' identities ( <code>id</code> ) and the corresponding measurement occasions ( <code>time</code> ). |
| <code>rlatent</code> | the latent random variables denoted by $e_{it}^{NO}$ in Touloumis (2016).                                                                                                                                                                       |

## Author(s)

Anestis Touloumis

## References

Cario, M. C. and Nelson, B. L. (1997) *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Li, S. T. and Hammond, J. L. (1975) Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man and Cybernetics* **5**, 557–561.

McFadden, D. (1974) *Conditional logit analysis of qualitative choice behavior*. New York: Academic Press, 105–142.

Touloumis, A. (2016) Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The SimCorMultRes Package. *The R Journal* **8**, 79–91.

Touloumis, A., Agresti, A. and Kateri, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics* **69**, 633–640.

### See Also

[rbin](#) for simulating correlated binary responses, [rmult.clm](#), [rmult.crm](#) and [rmult.acl](#) for simulating correlated ordinal responses.

### Examples

```
## See Example 3.1 in the Vignette.
betas <- c(1, 3, 2, 1.25, 3.25, 1.75, 0.75, 2.75, 2.25, 0, 0, 0)
sample_size <- 500
categories_no <- 4
cluster_size <- 3
set.seed(1)
x1 <- rep(rnorm(sample_size), each = cluster_size)
x2 <- rnorm(sample_size * cluster_size)
xdata <- data.frame(x1, x2)
equicorrelation_matrix <- toeplitz(c(1, rep(0.95, cluster_size - 1)))
identity_matrix <- diag(categories_no)
latent_correlation_matrix <- kronecker(equicorrelation_matrix,
  identity_matrix)
simulated_nominal_dataset <- rmult.bcl(clsizes = cluster_size,
  ncategories = categories_no, betas = betas, xformula = ~ x1 + x2,
  xdata = xdata, cor.matrix = latent_correlation_matrix)
suppressPackageStartupMessages(library("multgee"))
nominal_gee_model <- nomLORgee(y ~ x1 + x2,
  data = simulated_nominal_dataset$simdata, id = id, repeated = time,
  LORstr = "time.exch")
round(coef(nominal_gee_model), 2)
```

---

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

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

## Arguments

|                         |                                                                                                                                                                                             |
|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>clsize</code>     | integer indicating the common cluster size.                                                                                                                                                 |
| <code>intercepts</code> | numerical vector or matrix containing the intercepts of the marginal cumulative link model.                                                                                                 |
| <code>betas</code>      | numerical vector or matrix containing the value of the marginal regression parameter vector associated with the covariates (i.e., excluding intercepts).                                    |
| <code>xformula</code>   | formula expression as in other marginal regression models but without including a response variable.                                                                                        |
| <code>xdata</code>      | optional data frame containing the variables provided in <code>xformula</code> .                                                                                                            |
| <code>link</code>       | character string indicating the link function in the marginal cumulative link model. Options include 'probit', 'logit', 'cloglog' or 'cauchit'. Required when <code>rlatent = NULL</code> . |
| <code>cor.matrix</code> | matrix indicating the correlation matrix of the multivariate normal distribution when the NORTA method is employed ( <code>rlatent = NULL</code> ).                                         |
| <code>rlatent</code>    | matrix with <code>clsize</code> 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

$$Pr(Y_{it} \leq j | x_{it}) = F(\beta_{tj0} + \beta'_t x_{it})$$

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.

`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  $e_{it}^{O1}$  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  $e_{it}^{O1}$ .

### Value

Returns a list that has components:

|                      |                                                                                                                                                                                                                                                 |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>ysim</code>    | the simulated ordinal responses. Element $(i, t)$ represents the realization of $Y_{it}$ .                                                                                                                                                      |
| <code>simdata</code> | a data frame that includes the simulated response variables ( <code>y</code> ), the covariates specified by <code>xformula</code> , subjects' identities ( <code>id</code> ) and the corresponding measurement occasions ( <code>time</code> ). |
| <code>rlatent</code> | the latent random variables denoted by $e_{it}^{O1}$ in Touloumis (2016).                                                                                                                                                                       |

### Author(s)

Anestis Touloumis

### References

- Cario, M. C. and Nelson, B. L. (1997) *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.
- Li, S. T. and Hammond, J. L. (1975) Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man and Cybernetics* **5**, 557–561.
- McCullagh, P. (1980) Regression models for ordinal data. *Journal of the Royal Statistical Society B* **42**, 109–142.
- Touloumis, A. (2016) Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The SimCorMultRes Package. *The R Journal* **8**, 79–91.
- Touloumis, A., Agresti, A. and Kateri, M. (2013) GEE for multinomial responses using a local odds ratios parameterization. *Biometrics* **69**, 633–640.

### See Also

[rmult.bcl](#) for simulating correlated nominal responses, [rmult.crm](#) and [rmult.acl](#) for simulating correlated ordinal responses and [rbin](#) for simulating correlated binary responses.

### Examples

```
## See Example 3.2 in the Vignette.
set.seed(12345)
sample_size <- 500
cluster_size <- 4
beta_intercepts <- c(-1.5, -0.5, 0.5, 1.5)
beta_coefficients <- matrix(c(1, 2, 3, 4), 4, 1)
```

```

x <- rep(rnorm(sample_size), each = cluster_size)
latent_correlation_matrix <- toeplitz(c(1, 0.85, 0.5, 0.15))
simulated_ordinal_dataset <- rmult.clm(clsize = cluster_size,
  intercepts = beta_intercepts, betas = beta_coefficients, xformula = ~x,
  cor.matrix = latent_correlation_matrix, link = "probit")
head(simulated_ordinal_dataset$simdata, n = 8)

## Same sampling scheme except that the parameter vector is time-stationary.
set.seed(12345)
simulated_ordinal_dataset <- rmult.clm(clsize = cluster_size, betas = 1,
  xformula = ~x, cor.matrix = latent_correlation_matrix,
  intercepts = beta_intercepts, link = "probit")
## Fit a GEE model (Touloumis et al., 2013) to estimate the regression
## coefficients.
library(multgee)
ordinal_gee_model <- ordLORgee(y ~ x, id = id, repeated = time,
  link = "probit", data = simulated_ordinal_dataset$simdata)
coef(ordinal_gee_model)

```

rmult.crm

*Simulating Correlated Ordinal Responses Conditional on a Marginal Continuation-Ratio Model Specification*

## Description

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

## Usage

```

rmult.crm(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 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. |

|                         |                                                                                                                                                                                                  |
|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>cor.matrix</code> | matrix indicating the correlation matrix of the multivariate normal distribution when the NORTA method is employed ( <code>rlatent = NULL</code> ).                                              |
| <code>rlatent</code>    | matrix with <code>clsize</code> rows and <code>ncategories</code> 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

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

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  $\beta_{tj}$  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  $e_{itj}^{O2}$  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 `rlatent` argument. In this case, row  $i$  corresponds to subject  $i$  and columns  $(t - 1) * \text{ncategories} + 1, \dots, t * \text{ncategories}$  should contain the realization of  $e_{it1}^{O2}, \dots, e_{itJ}^{O2}$ , respectively, for  $t = 1, \dots, \text{clsize}$ .

## Value

Returns a list that has components:

|                      |                                                                                                                                                                                                                                                 |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <code>ysim</code>    | the simulated ordinal responses. Element $(i, t)$ represents the realization of $Y_{it}$ .                                                                                                                                                      |
| <code>simdata</code> | a data frame that includes the simulated response variables ( <code>y</code> ), the covariates specified by <code>xformula</code> , subjects' identities ( <code>id</code> ) and the corresponding measurement occasions ( <code>time</code> ). |
| <code>rlatent</code> | the latent random variables denoted by $e_{it}^{O2}$ in <i>Touloumis (2016)</i> .                                                                                                                                                               |

**Author(s)**

Anestis Touloumis

**References**

Cario, M. C. and Nelson, B. L. (1997) *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Li, S. T. and Hammond, J. L. (1975) Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man and Cybernetics* **5**, 557–561.

Touloumis, A. (2016) Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The SimCorMultRes Package. *The R Journal* (forthcoming).

Tutz, G. (1991) Sequential models in categorical regression. *Computational Statistics & Data Analysis* **11**, 275–295.

**See Also**

[rmult.bcl](#) for simulating correlated nominal responses, [rmult.clm](#) and [rmult.acl](#) for simulating correlated ordinal responses and [rbin](#) for simulating correlated binary responses.

**Examples**

```
## See Example 3.3 in the Vignette.
set.seed(1)
sample_size <- 500
cluster_size <- 4
beta_intercepts <- c(-1.5, -0.5, 0.5, 1.5)
beta_coefficients <- 1
x <- rnorm(sample_size * cluster_size)
categories_no <- 5
identity_matrix <- diag(1, (categories_no - 1) * cluster_size)
equicorrelation_matrix <- toeplitz(c(0, rep(0.24, categories_no - 2)))
ones_matrix <- matrix(1, cluster_size, cluster_size)
latent_correlation_matrix <- identity_matrix +
  kronecker(equicorrelation_matrix, ones_matrix)
simulated_ordinal_dataset <- rmult.crm(clsize = cluster_size,
  intercepts = beta_intercepts, betas = beta_coefficients, xformula = ~x,
  cor.matrix = latent_correlation_matrix, link = "probit")
head(simulated_ordinal_dataset$Ysim)
```

**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

|                          |                                                                                                                                         |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|
| <code>R</code>           | integer indicating the sample size.                                                                                                     |
| <code>cor.matrix</code>  | matrix indicating the correlation matrix of the multivariate normal distribution employed in the NORTA method.                          |
| <code>distr</code>       | character string vector of length <code>ncol(cor.matrix)</code> naming the quantile functions of the desired marginal distributions.    |
| <code>qparameters</code> | list of <code>ncol(cor.matrix)</code> lists indicating the parameter values of the quantile functions specified by <code>distr</code> . |

## 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

- Cario, M. C. and Nelson, B. L. (1997) *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.
- Li, S. T. and Hammond, J. L. (1975) Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man and Cybernetics* **5**, 557–561.
- Touloumis, A. (2016) Simulating Correlated Binary and Multinomial Responses under Marginal Model Specification: The SimCorMultRes Package. *The R Journal* **8**, 79–91.

## Examples

```
## An example with standard logistic as marginal distribution.
set.seed(1)
sample_size <- 1000
latent_correlation_matrix <- toeplitz(c(1, rep(0.8, 2)))
latent_correlation_matrix
common_marginal_distribution <- rep("qlogis", 3)
simulated_logistic_responses <- rnorta(R = sample_size,
  cor.matrix = latent_correlation_matrix,
  distr = common_marginal_distribution)

## The following lines exemplify the NORTA method.
set.seed(1)
simulated_normal_responses <- rsmvnorm(R = sample_size,
  cor.matrix = latent_correlation_matrix)
norta_simulated <- qlogis(pnorm(simulated_normal_responses))
all(simulated_logistic_responses == norta_simulated)

## Change the marginal distributions to standard normal, standard
## logistic and standard extreme value distribution.
set.seed(1)
different_marginal_distributions <- c("qnorm", "qlogis", "qgumbel")
simulated_logistic_responses <- rnorta(R = sample_size,
  cor.matrix = latent_correlation_matrix,
  distr = different_marginal_distributions)
cor(simulated_logistic_responses)
colMeans(simulated_logistic_responses)
apply(simulated_logistic_responses, 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))
simulated_logistic_responses <- rnorta(R = sample_size,
  cor.matrix = latent_correlation_matrix,
  distr = different_marginal_distributions, qparameters = qpars)
cor(simulated_logistic_responses)
colMeans(simulated_logistic_responses)
apply(simulated_logistic_responses, 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**

```
rsmvnorm(R = R, cor.matrix = cor.matrix)
```

**Arguments**

`R` integer indicating the sample size.

`cor.matrix` matrix indicating the correlation matrix of the multivariate normal distribution.

**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.

**Value**

Returns `R` random vectors of size `ncol(cor.matrix)`.

**Author(s)**

Anestis Touloumis

**Examples**

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

---

simulation

---

*Simulated Correlation Parameters*


---

**Description**

A simulated dataset to explore the association between the correlation parameter of bivariate normally distributed random variables used in the intermediate step of the NORTA method and the correlation parameter of the resulting non-normal random responses generated by the NORTA method for all the threshold approached implemented in this package.

**Usage**

```
simulation
```

**Format**

A data frame with 100 rows and 4 columns:

**rho** numeric indicating the true value of the correlation parameter.

**normal** numeric indicating the simulated correlation parameter when the marginal distribution of each of the latent variables is normal.

**logistic** numeric indicating the simulated correlation parameter when the marginal distribution of each of the latent variables is logistic.

**gumbel** numeric indicating the simulated correlation parameter when the marginal distribution of each of the latent variables is Gumbel.

**Examples**

```
plot(rho ~ normal ~ rho, data = simulation, type = "l", col = "blue",  
     ylim = c(0, 0.016),  
     ylab = expression(rho ~ bar(rho)[sim]),  
     xlab = expression(rho))  
points(rho ~ logistic ~ rho, data = simulation, type = "l", col = "red")  
points(rho ~ gumbel ~ rho, data = simulation, type = "l", col = "grey")  
legend("topright", legend = c("Normal", "Logistic", "Gumbel"),  
       col = c("blue", "red", "grey"), pch = "l" )
```

# Index

## \* **datasets**

simulation, [17](#)

Distributions, [15](#)

rbin, [2](#), [6](#), [9](#), [11](#), [14](#)

rmult.acl, [4](#), [4](#), [9](#), [11](#), [14](#)

rmult.bcl, [4](#), [6](#), [7](#), [11](#), [14](#)

rmult.clm, [4](#), [6](#), [9](#), [9](#), [14](#)

rmult.crm, [4](#), [6](#), [9](#), [11](#), [12](#)

rnorta, [14](#)

rsmvnorm, [16](#)

simulation, [17](#)