Package 'simAutoReg'

August 5, 2023

Title Simulate Data from Autoregressive Models
Version 0.9.1
Description Simulate data from autoregressive models.
<pre>URL https://github.com/ijapesigan/simAutoReg,</pre>
https://ijapesigan.github.io/simAutoReg/
BugReports https://github.com/ijapesigan/simAutoReg/issues
License GPL (>= 3)
Encoding UTF-8
Roxygen list(markdown = TRUE)
VignetteBuilder knitr
Depends R (>= $3.0.0$)
LinkingTo Rcpp, RcppArmadillo
Imports Rcpp
Suggests knitr, rmarkdown, testthat, microbenchmark
RoxygenNote 7.2.3
NeedsCompilation yes
Author Ivan Jacob Agaloos Pesigan [aut, cre, cph] (https://orcid.org/0000-0003-4818-8420)
Maintainer Ivan Jacob Agaloos Pesigan <ijapesigan@gmail.com></ijapesigan@gmail.com>
R topics documented:
SimAR 2 SimMVN 3 SimMVNFixed 4 SimVAR 6 SimVARExo 5 SimVariance 6 SimVARZIP 10

2 SimAR

SimAR	Simulate Data from an Autoregressive Model with Constant Term	
Index		16
	YX YXExo	
	SimVARZIPExo	

Description

This function generates synthetic time series data from an autoregressive (AR) model.

Usage

```
SimAR(time, burn_in, constant, coef, sd)
```

Arguments

time	Integer. Number of time points to simulate.
burn_in	Integer. Number of burn-in periods before recording data.
constant	Numeric. The constant term of the AR model.
coef	Numeric vector. Autoregressive coefficients.
sd	Numeric. The standard deviation of the random process noise.

Details

The SimAR() function generates synthetic time series data from an autoregressive (AR) model. The generated data follows the AR(p) model, where p is the number of coefficients specified in coef. The generated time series data includes a constant term and autoregressive terms based on the provided coefficients. Random noise, sampled from a normal distribution with mean 0 and standard deviation sd, is added to the time series. A burn-in period is specified to exclude initial data points from the output.

The steps in generating the autoregressive time series with burn-in are as follows:

- Set the order of the AR model to p based on the length of coef.
- Create a vector data of size time + burn_in to store the generated AR time series data.
- Create a vector data of size time + burn_in of random process noise from a normal distribution with mean 0 and standard deviation sd.
- Generate the autoregressive time series with burn-in using the formula:

$$Y_{t} = constant + \sum_{i=1}^{p} (coef[i] * Y_{t-i}) + noise_{t}$$

where Y_t is the time series data at time t, constant is the constant term, coef[i] are the autoregressive coefficients, Y_{t-i} are the lagged data points up to order p, and $noise_t$ is the random noise at time t.

• Remove the burn-in period from the generated time series data.

SimMVN 3

Value

Numeric vector (column matrix) containing the simulated time series data.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

```
Other Simulation of Autoregressive Data Functions: SimMVNFixed(), SimMVN(), SimVARExo(), SimVARZIPExo(), SimVARZIP(), SimVAR(), SimVariance(), YXExo(), YX()
```

Examples

```
set.seed(42) SimAR(time = 10, burn_in = 5, constant = 2, coef = c(0.5, -0.3), sd = 0.1)
```

SimMVN

Simulate Multivariate Normal Random Numbers

Description

This function generates multivariate normal random numbers.

Usage

```
SimMVN(n, location, chol_scale)
```

Arguments

n Integer. Number of samples to generate.

location Numeric vector. Mean vector of length k, where k is the number of variables.

chol_scale Numeric matrix. Cholesky decomposition of the covariance matrix of dimen-

sions k by k.

Details

The SimMVN() function generates multivariate normal random numbers using the Cholesky decomposition method. Given the number of samples n, the mean vector location of length k (where k is the number of variables), and the Cholesky decomposition chol_scale of the covariance matrix of dimensions k by k, the function produces a matrix of multivariate normal random numbers.

The steps involved in generating multivariate normal random numbers are as follows:

- Determine the number of variables k from the length of the mean vector.
- Generate random data from a standard multivariate normal distribution, resulting in an n by k matrix of random numbers.

4 SimMVNFixed

 Transform the standard normal random data into multivariate normal random data using the Cholesky decomposition chol_scale.

- Add the mean vector location to the transformed data to obtain the final simulated multivariate normal random numbers.
- The function returns a matrix of simulated multivariate normal random numbers with dimensions n by k, where n is the number of samples and k is the number of variables. This matrix can be used for various statistical analyses and simulations.

Value

Matrix containing the simulated multivariate normal random numbers, with dimensions n by k, where n is the number of samples and k is the number of variables.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

The chol() function in R to obtain the Cholesky decomposition of a covariance matrix.

```
Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimVARExo(), SimVARZIPExo(), SimVARZIP(), SimVAR(), SimVariance(), YXExo(), YX()
```

Examples

```
set.seed(42)
n <- 1000L
location <- c(0.5, -0.2, 0.1)
scale <- matrix(
   data = c(1.0, 0.3, 0.3, 0.3, 1.0, 0.2, 0.3, 0.2, 1.0),
   nrow = 3,
   byrow = TRUE
)
chol_scale <- chol(scale)
y <- SimMVN(n = n, location = location, chol_scale = chol_scale)
colMeans(y)
var(y)</pre>
```

SimMVNFixed

Simulate Multivariate Normal Random Numbers with Optional Fixed Values

Description

This function generates multivariate normal random numbers with optinal fixed values where the variance is zero.

SimMVNFixed 5

Usage

```
SimMVNFixed(n, location, scale)
```

Arguments

n Integer. Number of samples to generate.

location Numeric vector. Mean vector of length k, where k is the number of variables.

scale Numeric matrix. Covariance matrix of dimensions k by k. Values for variables

with variance of 0 will be fixed to the corresponding value in location.

Details

The SimMVNFixed() function first identifies the indices of non-constant variables (i.e., variables with variance not equal to 0) in the covariance matrix. It then extracts the non-constant elements from the mean vector and the covariance matrix. A Cholesky decomposition is performed on the covariance matrix of non-constant variables. Random samples are generated for the non-constant variables using the Cholesky factor. The generated data matrix is constructed by setting the non-constant variables and constant variables to their corresponding values.

Value

Matrix containing the simulated multivariate normal random numbers, with dimensions n by k, where n is the number of samples and k is the number of variables.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

```
Other Simulation of Autoregressive Data Functions: SimAR(), SimMVN(), SimVARExo(), SimVARZIPExo(), SimVARZIP(), SimVAR(), SimVariance(), YXExo(), YX()
```

Examples

```
set.seed(42)
n <- 1000L
location <- c(0.5, -0.2, 0.1)
scale <- matrix(
   data = c(0, 0.3, 0.3, 0.3, 1.0, 0.2, 0.3, 0.2, 1.0),
   nrow = 3,
   byrow = TRUE
)
y <- SimMVNFixed(n = n, location = location, scale = scale)
colMeans(y)
var(y)</pre>
```

6 SimVAR

SimVAR	Simulate Data from a Vector Autoregressive (VAR) Model

Description

This function generates synthetic time series data from a Vector Autoregressive (VAR) model.

Usage

```
SimVAR(time, burn_in, constant, coef, chol_cov)
```

Arguments

time	Integer. Number of time points to simulate.
burn_in	Integer. Number of burn-in observations to exclude before returning the results.
constant	Numeric vector. The constant term vector of length k, where k is the number of variables.
coef	Numeric matrix. Coefficient matrix with dimensions k by $(k * p)$. Each k by k block corresponds to the coefficient matrix for a particular lag.
chol_cov	Numeric matrix. The Cholesky decomposition of the covariance matrix of the multivariate normal noise. It should have dimensions k by k.

Details

The SimVAR() function generates synthetic time series data from a Vector Autoregressive (VAR) model. The VAR model is defined by the constant term constant, the coefficient matrix coef, and the Cholesky decomposition of the covariance matrix of the multivariate normal process noise chol_cov. The generated time series data follows a VAR(p) process, where p is the number of lags specified by the size of coef. The generated data includes a burn-in period, which is excluded before returning the results.

The steps involved in generating the VAR time series data are as follows:

- Extract the number of variables k and the number of lags p from the input.
- Create a matrix data of size k by (time + burn_in) to store the generated VAR time series data.
- Set the initial values of the matrix data using the constant term constant.
- For each time point starting from the p-th time point to time + burn_in 1:
 - Generate a vector of random noise from a multivariate normal distribution with mean 0 and covariance matrix chol_cov.
 - Generate the VAR time series values for each variable j at time t using the formula:

$$Y_{tj} = constant_j + \sum_{l=1}^{p} \sum_{m=1}^{k} (coef_{jm} * Y_{im}) + noise_j$$

SimVAR 7

where Y_{tj} is the value of variable j at time t, $constant_j$ is the constant term for variable j, $coef_{jm}$ are the coefficients for variable j from lagged variables up to order p, Y_{tm} are the lagged values of variable m up to order p at time t, and $noise_j$ is the element j from the generated vector of random process noise.

• Transpose the matrix data and return only the required time period after the burn-in period, which is from column burn_in to column time + burn_in - 1.

Value

Numeric matrix containing the simulated time series data with dimensions k by time, where k is the number of variables and time is the number of observations.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimMVN(), SimVARExo(), SimVARZIPExo(), SimVARZIP(), SimVariance(), YXExo(), YX()

Examples

```
set.seed(42)
time <- 50L
burn_in <- 10L
k <- 3
p <- 2
constant <- c(1, 1, 1)
coef <- matrix(</pre>
  data = c(
    0.4, 0.0, 0.0, 0.1, 0.0, 0.0,
    0.0, 0.5, 0.0, 0.0, 0.2, 0.0,
    0.0, 0.0, 0.6, 0.0, 0.0, 0.3
  ),
  nrow = k,
  byrow = TRUE
)
chol_cov <- chol(diag(3))</pre>
y <- SimVAR(
  time = time,
  burn_in = burn_in,
  constant = constant,
  coef = coef,
  chol_cov = chol_cov
head(y)
```

8 SimVARExo

SimVARExo	This function generates synthetic time series data from a Vector Autoregressive (VAR) model with exogenous variables.

Description

This function generates synthetic time series data from a Vector Autoregressive (VAR) model with exogenous variables.

Usage

```
SimVARExo(time, burn_in, constant, coef, chol_cov, exo_mat, exo_coef)
```

Arguments

time	Integer. Number of time points to simulate.
burn_in	Integer. Number of burn-in observations to exclude before returning the results.
constant	Numeric vector. The constant term vector of length k, where k is the number of variables.
coef	Numeric matrix. Coefficient matrix with dimensions k by $(k * p)$. Each k by k block corresponds to the coefficient matrix for a particular lag.
chol_cov	Numeric matrix. The Cholesky decomposition of the covariance matrix of the multivariate normal noise. It should have dimensions k by k.
exo_mat	Numeric matrix. Matrix of exogenous covariates with dimensions time + burn_in by x. Each column corresponds to a different exogenous variable.
exo_coef	Numeric vector. Coefficient matrix with dimensions k by x associated with the exogenous covariates.

Value

Numeric matrix containing the simulated time series data with dimensions k by time, where k is the number of variables and time is the number of observations.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimMVN(), SimVARZIPExo(), SimVARZIP(), SimVAR(), SimVariance(), YXExo(), YX()

SimVariance 9

SimVariance	Generate Random Data for the Variance Vector

Description

This function generates random data for the variance vector given by

$$oldsymbol{\sigma^2} = \exp\left(oldsymbol{\mu} + oldsymbol{arepsilon}
ight) \quad ext{with} \quad oldsymbol{arepsilon} \sim \mathcal{N}\left(\mathbf{0}, oldsymbol{\Sigma}
ight)$$

.

Usage

SimVariance(n, location, chol_scale)

Arguments

n Integer. Number of samples to generate.

location Numeric vector. The constant term μ .

chol_scale Numeric matrix. Cholesky decomposition of the covariance matrix Σ for the

multivariate normal random error ε .

Details

The SimVariance() function generates random data for the variance vector based on the exponential of a multivariate normal distribution. Given the number of samples n, the constant term μ represented by the location vector, and the Cholesky decomposition matrix Σ for the multivariate normal random error ε , the function simulates n independent samples of the variance vector σ^2 . Each sample of the variance vector σ^2 is obtained by calculating the exponential of random variations to the mean vector μ . The random variations are generated using the Cholesky decomposition of the covariance matrix Σ . Finally, the function returns a matrix with each column containing the simulated variance vector for each sample.

Value

Matrix with each row containing the simulated variance vector for each sample.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

```
Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimMVN(), SimVARExo(), SimVARZIPExo(), SimVARZIP(), SimVAR(), YXExo(), YX()
```

10 SimVARZIP

Examples

```
set.seed(42)
n <- 100
location <- c(0.5, -0.2, 0.1)
chol_scale <- chol(
    matrix(
        data = c(1.0, 0.3, 0.3, 0.3, 1.0, 0.2, 0.3, 0.2, 1.0),
        nrow = 3,
        byrow = TRUE
    )
)
SimVariance(n = n, location = location, chol_scale = chol_scale)</pre>
```

SimVARZIP

Simulate Data from a Vector Autoregressive Zero-Inflated Poisson (VARZIP) Model

Description

This function generates synthetic time series data from a Vector Autoregressive Zero-Inflated Poisson (VARZIP) model.

Usage

```
SimVARZIP(time, burn_in, constant, coef, chol_cov)
```

Arguments

time	Integer. Number of time points to simulate.
burn_in	Integer. Number of burn-in observations to exclude before returning the results.
constant	Numeric vector. The constant term vector of length k, where k is the number of variables.
coef	Numeric matrix. Coefficient matrix with dimensions k by $(k * p)$. Each k by k block corresponds to the coefficient matrix for a particular lag.
chol_cov	Numeric matrix. The Cholesky decomposition of the covariance matrix of the multivariate normal noise. It should have dimensions k by k.

Details

The SimVARZIP() function generates synthetic time series data from a Vector Autoregressive (VAR) with Zero-Inflated Poisson (ZIP) model for the first observed variable. See SimVAR() for more details on generating data for VAR(p). The SimVARZIP function goes further by using the generated values for the first variable to generate data from the ZIP model. The exponential of the values from the first variable from the original VAR(p) model are used as the intensity parameter in the Poisson distribution in the ZIP model. Data from the ZIP model are used to replace the original

SimVARZIP 11

values for the first variable. Values for the rest of the variables are unchanged. The generated data includes a burn-in period, which is excluded before returning the results.

The steps involved in generating the time series data are as follows:

- Extract the number of variables k and the number of lags p from the input.
- Create a matrix data of size k x (time + burn_in) to store the generated data.
- Set the initial values of the matrix data using the constant term constant.
- For each time point starting from the p-th time point to time + burn_in 1:
 - Generate a vector of random process noise from a multivariate normal distribution with mean 0 and covariance matrix chol_cov.
 - Generate the VAR time series values for each variable j at time t by applying the autoregressive terms for each lag lag and each variable 1.
 - * Add the generated noise to the VAR time series values.
 - * For the first variable, apply the Zero-Inflated Poisson (ZIP) model:
 - · Compute the intensity intensity as the exponential of the first variable's value at time t.
 - · Sample a random value u from a uniform distribution on [0, 1].
 - · If u is less than intensity / (1 + intensity), set the first variable's value to zero (inflation).
 - · Otherwise, sample the first variable's value from a Poisson distribution with mean intensity (count process).
- Transpose the data matrix data and return only the required time period after burn-in as a numeric matrix.

Value

Numeric matrix containing the simulated time series data with dimensions k by time, where k is the number of variables and time is the number of observations.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

```
Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimMVN(), SimVARExo(), SimVARZIPExo(), SimVAR(), SimVariance(), YXExo(), YX()
```

Examples

```
set.seed(42)
time <- 50L
burn_in <- 10L
k <- 3
p <- 2
constant <- c(1, 1, 1)
coef <- matrix(</pre>
```

12 SimVARZIPExo

```
data = c(
    0.4, 0.0, 0.0, 0.1, 0.0, 0.0,
    0.0, 0.5, 0.0, 0.0, 0.2, 0.0,
    0.0, 0.0, 0.6, 0.0, 0.0, 0.3
),
    nrow = k,
    byrow = TRUE
)
chol_cov <- chol(diag(3))
y <- SimVARZIP(
    time = time,
    burn_in = burn_in,
    constant = constant,
    coef = coef,
    chol_cov = chol_cov
)
head(y)</pre>
```

SimVARZIPExo

Simulate Data from a Vector Autoregressive Zero-Inflated Poisson (VARZIP) Model with Exogenous Variables

Description

This function generates synthetic time series data from a Vector Autoregressive Zero-Inflated Poisson (VARZIP) model with exogenous variables.

Usage

```
SimVARZIPExo(time, burn_in, constant, coef, chol_cov, exo_mat, exo_coef)
```

Arguments

time	Integer. Number of time points to simulate.
burn_in	Integer. Number of burn-in observations to exclude before returning the results.
constant	Numeric vector. The constant term vector of length k, where k is the number of variables.
coef	Numeric matrix. Coefficient matrix with dimensions k by $(k * p)$. Each k by k block corresponds to the coefficient matrix for a particular lag.
chol_cov	Numeric matrix. The Cholesky decomposition of the covariance matrix of the multivariate normal noise. It should have dimensions k by k.
exo_mat	Numeric matrix. Matrix of exogenous covariates with dimensions time + burn_in by x. Each column corresponds to a different exogenous variable.
exo_coef	Numeric vector. Coefficient matrix with dimensions k by x associated with the exogenous covariates.

YX 13

Value

Numeric matrix containing the simulated time series data with dimensions k by time, where k is the number of variables and time is the number of observations.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimMVN(), SimVARExo(), SimVARZIP(), SimVAR(), SimVariance(), YXExo(), YX()

ΥX

Create Y and X Matrices

Description

This function creates the dependent variable (Y) and predictor variable (X) matrices.

Usage

```
YX(data, p)
```

Arguments

data	Numeric matrix. The time series data with dimensions t by k, where t is the number of observations and k is the number of variables.
р	Integer. The order of the VAR model (number of lags).

Details

The YX() function creates the Y and X matrices required for fitting a Vector Autoregressive (VAR) model. Given the input data matrix with dimensions t by k, where t is the number of observations and k is the number of variables, and the order of the VAR model p (number of lags), the function constructs lagged predictor matrix X and the dependent variable matrix Y.

The steps involved in creating the Y and X matrices are as follows:

- Determine the number of observations t and the number of variables k from the input data matrix.
- Create matrices X and Y to store lagged variables and the dependent variable, respectively.
- Populate the matrices X and Y with the appropriate lagged data. The predictors matrix X contains a column of ones and the lagged values of the dependent variables, while the dependent variable matrix Y contains the original values of the dependent variables.
- The function returns a list containing the Y and X matrices, which can be used for further analysis and estimation of the VAR model parameters.

14 YXExo

Value

List containing the dependent variable (Y) and predictor variable (X) matrices. Note that the resulting matrices will have t - p rows.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

The SimVAR() function for simulating time series data from a VAR model.

```
Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimMVN(), SimVARExo(), SimVARZIPExo(), SimVARZIP(), SimVAR(), SimVariance(), YXExo()
```

Examples

```
set.seed(42)
time <- 50L
burn_in <- 10L
k <- 3
p <- 2
constant \leftarrow c(1, 1, 1)
coef <- matrix(</pre>
  data = c(
    0.4, 0.0, 0.0, 0.1, 0.0, 0.0,
    0.0, 0.5, 0.0, 0.0, 0.2, 0.0,
    0.0, 0.0, 0.6, 0.0, 0.0, 0.3
  ),
  nrow = k,
  byrow = TRUE
)
chol_cov <- chol(diag(3))</pre>
y <- SimVAR(
  time = time,
  burn_in = burn_in,
  constant = constant,
  coef = coef,
  chol_cov = chol_cov
)
yx \leftarrow YX(data = y, p = 2)
str(yx)
```

YXExo

Create Y and X Matrices with Exogenous Variables

Description

This function creates the dependent variable (Y) and predictor variable (X) matrices.

YXExo 15

Usage

```
YXExo(data, p, exo_mat)
```

Arguments

data	Numeric matrix. The time series data with dimensions t by k, where t is the number of observations and k is the number of variables.
р	Integer. The order of the VAR model (number of lags).
exo_mat	Numeric matrix. Matrix of exogenous variables with dimensions t by m.

Details

The YX() function creates the Y and X matrices required for fitting a Vector Autoregressive (VAR) model. Given the input data matrix with dimensions t by k, where t is the number of observations and k is the number of variables, and the order of the VAR model p (number of lags), the function constructs lagged predictor matrix X and the dependent variable matrix Y.

The steps involved in creating the Y and X matrices are as follows:

- Determine the number of observations t and the number of variables k from the input data matrix.
- Create matrices X and Y to store lagged variables and the dependent variable, respectively.
- Populate the matrices X and Y with the appropriate lagged data. The predictors matrix X contains a column of ones and the lagged values of the dependent variables, while the dependent variable matrix Y contains the original values of the dependent variables.
- The function returns a list containing the Y and X matrices, which can be used for further analysis and estimation of the VAR model parameters.

Value

List containing the dependent variable (Y) and predictor variable (X) matrices. Note that the resulting matrices will have t - p rows.

Author(s)

Ivan Jacob Agaloos Pesigan

See Also

The SimVAR() function for simulating time series data from a VAR model.

```
Other Simulation of Autoregressive Data Functions: SimAR(), SimMVNFixed(), SimMVN(), SimVARExo(), SimVARZIPExo(), SimVARZIP(), SimVAR(), SimVariance(), YX()
```

Index

* Simulation of Autoregressive Data	SimMVN, 3, 3, 5, 7–9, 11, 13–15
Functions	SimMVN(), 3
SimAR, 2	SimMVNFixed, 3, 4, 4, 7–9, 11, 13–15
SimMVN, 3	SimMVNFixed(), 5
SimMVNFixed, 4	SimVAR, <i>3-5</i> , 6, 8, 9, <i>11</i> , <i>13-15</i>
SimVAR, 6	SimVAR(), 6, 10, 14, 15
SimVARExo, 8	SimVARExo, 3-5, 7, 8, 9, 11, 13-15
SimVariance, 9	SimVariance, <i>3</i> – <i>5</i> , <i>7</i> , <i>8</i> , 9, <i>11</i> , <i>13</i> – <i>15</i>
SimVARZIP, 10	SimVariance(), 9
SimVARZIPExo, 12	SimVARZIP, <i>3-5</i> , <i>7-9</i> , 10, <i>13-15</i>
YX, 13	SimVARZIP(), 10
YXExo, 14	SimVARZIPExo, 3-5, 7-9, 11, 12, 14, 15
* simAutoReg	
SimAR, 2	YX, 3–5, 7–9, 11, 13, 13, 15
SimMVN, 3	YX(), <i>13</i> , <i>15</i>
SimMVNFixed, 4	YXExo, 3–5, 7–9, 11, 13, 14, 14
SimVAR, 6	
SimVARExo, 8	
SimVariance, 9	
SimVARZIP, 10	
SimVARZIPExo, 12	
YX, 13	
YXExo, 14	
* sim	
SimAR, 2	
SimMVN, 3	
SimMVNFixed, 4	
SimVAR, 6	
SimVARExo, 8	
SimVariance, 9	
SimVARZIP, 10	
SimVARZIPExo, 12	
* utils	
YX, 13	
YXExo, 14	
chol(), 4	
SimAR, 2, 4, 5, 7–9, 11, 13–15 SimAR(), 2	