Review on Simrel Paper

Raju Rimal

Reference to the line or paragraph: Small Bold Text

Comments: *italics*

Actual changes: Normal text

DONE Corrections in Paper

Introduction

Page 2:

Reference added (Johnson 2013) along with (Ripley 2009) and (Gamerman and Lopes 2006).

Line 2 last paragraph

In particular, methods based on covariance structure has been discussed by (Arteaga and Ferrer 2010, Arteaga and Ferrer (2013), Camacho (2017)), which follow an algorithmic approach to find simulated data satisfying the desired correlation structure.

Page 3:

After first Paragraph

The github repository of the package at http://github.com/simulatr/simrel has rich documentation with many examples and cases along with detail description of simulation parameters. In the following two sections, the discussion encircle the mathematical framework behind. In addition, in section-4 and 5 we have also discussed the input parameters needed for simrel function in brief. In section 4, an implementation is presented as an example of use-cases and the final section introduces the shiny web application for this tool.

Last Paragraph (Coefficient of Determination)

~~**Coefficient of Determination** The diagonal elements of the coefficient-of-determination matrix $\rho_y^2(m \times m)$ give the amount of variation in each response variable that is explained~~

~~by **x**.~~

$$\boldsymbol{\rho}_y^2 = \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}$$

Coefficient of Determination Since, a matrix of coefficient-of-determination represents the proportion of variation explained by the predictors, we can write this matrix as,

$$\left(\rho_y^2\right)_{jj'} = \frac{\sigma_{xy_j}^t \Sigma_{xx}^{-1} \sigma_{xy_{j'}}}{\sqrt{\sigma_{y_j}^2 \sigma_{y_{j'}}^2}} \forall j, j' = 1\ldots m$$

where, σ_{xy_j}, $\sigma_{xy_{j'}}$ are covariance between **x** and y_j, $y_{j'}$ respectively. Also, $\sigma_{y_j}^2$ and $\sigma_{y_{j'}}^2$ are unconditional variances of y_j and $y_{j'}$.

Here the numerator gives the variation that **x** explained which is equivalent to the covariance of fitted **y** in sample space. The denominator gives the total variation present in **y**. The diagonal element of this matrix is the proportion of variation in a response $y_j, j = 1, \ldots m$ explained by the predictor.

Page 5:

~~**Coefficient of Determination** The coefficient-of-determination matrix for (4) is~~

$$\begin{aligned}
\boldsymbol{\rho}_w^2 &= \boldsymbol{\Sigma}_{wz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1} \\
&= \mathbf{Q}\boldsymbol{\Sigma}_{yx}\mathbf{R}^t \left(\mathbf{R}\boldsymbol{\Sigma}_{xx}\mathbf{R}^t\right)^{-1} \mathbf{R}\boldsymbol{\Sigma}_{xy}\mathbf{Q}^t \left(\mathbf{Q}\boldsymbol{\Sigma}_{yy}^{-1}\mathbf{Q}^t\right) \\
&= \mathbf{Q}\left[\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\right]\mathbf{Q} = \mathbf{Q}\boldsymbol{\rho}_y^2\mathbf{Q}^t
\end{aligned}$$

Coefficient of Determination The coefficient-of-determination matrix corresponding to **w** can be written as,

$$\begin{aligned}
\left(\rho_w^2\right)_{jj'} &= \boldsymbol{\Sigma}_{ww}^{-1/2}\boldsymbol{\Sigma}_{wz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zw}\boldsymbol{\Sigma}_{ww}^{-1/2} \\
&= \frac{\sigma_{zw_i}^t \Sigma_{zz}^{-1} \sigma_{zw_{j'}}}{\sqrt{\sigma_{w_j}^2 \sigma_{w_{j'}}^2}} \forall j, j' = 1\ldots m
\end{aligned}$$

where, σ_{zw_j} and $\sigma_{zw_{j'}}$ are covariances of **z** with w_j and $w_{j'}$ respectively. Also, $\sigma_{w_j}^2$ and $\sigma_{w_{j'}}^2$ are unconditional variances of w_j and $w_{j'}$. For simplicity, we will denote $\sigma_{z_i w_j}$ by σ_{ij}.

Since the rotation matrices gives direct connection between the covariance of (1) and (3), a straight forward relationship can be imagined between the termes in above matrix with their counterpart covariance matrices of **xy**-space.

Relevant Components

Page 6

Equation number has been added to the expression of generating lambda from gamma. Subscript changed from j to i.

Added to equation of γ

Hence, we can write $\Sigma_{zz} = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$.

Parameterization of Σ_{ww}

~~Here we assume that $\mathbf{w}$'s are independent and has mutinormal distribution wnith variances 1, hence $\Sigma_{ww} = \mathbf{I}_m$.~~ In similar manner, a parametric representation of eigenvalues corresponding to Σ_{ww} is adopted as,

$$\kappa_j = e^{-\eta(j-1), \eta>0 \text{ and } j=1...m}$$

Here, the decline of eigenvalues becomes steeper as η increases from zero. At $\eta = 0$, all w will have equal variance. Hence we can write $\Sigma_{ww} = \text{diag}(\kappa_1, \ldots, \kappa_m)$.

Page 8

Paragraph 2

$$\require{cancel}\cancel{\boldsymbol{\rho}_w^2 = \mathbf{\Sigma}_{zw}^t \mathbf{\Sigma}_{zz}^{-1} \mathbf{\Sigma}_{zw} \mathbf{\Sigma}_{ww}^{-1}}$$
$$\boldsymbol{\rho}_w^2 = \mathbf{\Sigma}_{ww}^{-1/2} \mathbf{\Sigma}_{zw}^t \mathbf{\Sigma}_{zz}^{-1} \mathbf{\Sigma}_{zw} \mathbf{\Sigma}_{ww}^{-1/2}$$
$$= \frac{\sigma_{ij}^t \Lambda^{-1} \sigma_{ij'}}{\sqrt{\sigma_j^2 \sigma_{j'}^2}} \forall j, j' = 1 \ldots m$$

Paragraph 3

h~~$\Sigma_{ww} = \mathbf{I}_m$~~

$$\Sigma_{ww} = \text{diag}(\kappa_1, \ldots, \kappa_m)$$

Paragraph 4 $\cancel{\rho_w^2 = \Sigma_{zw}^t \Lambda^{-1} \Sigma_{zw} \Sigma_{ww}^{-1}}$

$$\rho_w^2 = \Sigma_{ww}^{-1/2} \Sigma_{zw}^t \Lambda^{-1} \Sigma_{zw} \Sigma_{ww}^{-1/2} = \begin{bmatrix} \sum\limits_{i=1}^p \dfrac{\sigma_{i1}^2}{\lambda_i \kappa_1} & \cdots & \sum\limits_{i=1}^p \dfrac{\sigma_{i1} \sigma_{im}}{\lambda_i \sqrt{\kappa_1 \kappa_m}} \\ \vdots & \ddots & \vdots \\ \sum\limits_{i=1}^p \dfrac{\sigma_{i1} \sigma_{im}}{\lambda_i \sqrt{\kappa_1 \kappa_m}} & \cdots & \sum\limits_{i=1}^p \dfrac{\sigma_{im}^2}{\lambda_i \kappa_m} \end{bmatrix}$$

Paragraph 6

$\cancel{\rho_{w_j}^2 = \sum_{i=1}^p \frac{\sigma_{ij}^2}{\lambda_i}}$

$$\rho_{w_j}^2 = \sum_{i=1}^p \frac{\sigma_{ij}^2}{\lambda_i \kappa_j}$$

Paragraph 8

3

~~$\rho^2_{w_j} = \sum_{i \in \mathcal{P}_j} \frac{\sigma^2_{ij}}{\lambda_i}$~~

$$\rho^2_{w_j} = \sum_{i \in \mathcal{P}_j} \frac{\sigma^2_{ij}}{\lambda_i \kappa_j}$$

Paragraph 12 (last display equation)

~~$\sigma_{ij} = \text{Sign}(\mathcal{S}_i)\sqrt{\frac{\rho^2_{wj}|\mathcal{S}_i|}{\sum_{k \in \mathcal{P}_j}|\mathcal{S}_k|}\lambda_i}$~~

$$\sigma_{ij} = \text{Sign}(\mathcal{S}_i)\sqrt{\frac{\rho^2_{wj}|\mathcal{S}_i|}{\sum_{k \in \mathcal{P}_j}|\mathcal{S}_k|}\lambda_i \kappa_j}$$

for $i \in \mathcal{P}_j$ and $j = 1, \ldots m$

Implementation

Page 13

~~gamma~~ γ *(referred to corresponding equation)*

Line 8

Here, in this example we have assumed that all w's have equal variance, i.e. $\Sigma_{ww} = I_m$

Line 11

Although the simulation method is well equipped to simulate data with $p \gg n$, for incorporating envelope estimation methods which are based on maximization of likelihood, we have chosen a $n > p$ situation in the example.

Figure 3 Caption added

Here (a) is the covariance structure of the latent space, which is rotated by the block diagonal rotation matrix in (b) resulting the covariance structure of simulated data in (c).

Web Interface

Page 16

Last Paragraph

~~(uses `simrel` package in CRAN), (not yet available in CRAN), (`simrel-m`).~~

second last line

Users can also download simulated data in JSON and CSV format.

Page 17

After first paragraph

An R expression equivalent to the input parameters as shown in Figure - 5(b) can be written as,

```r
simrel(
  n      = 200, # Number of training observations
  ntest  = 50, # Number of test observations
  p      = 15, # Number of predictor variables
  q      = c(5, 4), # Number of relevant predictors
  relpos = list(c(1, 2), c(3, 4, 6)),
         # Relevant predictor components
  R2     = c(0.8, 0.7), # Rsq for each response component
  m      = 4, # Number of response variables
  gamma  = 0.6, # Decay factor of eigenvalues of predictors
  eta    = 0, # Decay factor of eigenvalues of responses
  ypos   = list(c(1, 3), c(2, 4)),
         # Combination of response components on rotation
  type   = "multivariate"
)
```

After Second paragraph until conclusion, major changes have been done. In order to make Figure-5 more clear. Figure-5 moved a page before

Figure-5 Caption:

Web interface of shiny application of `simrel`: (**a**) Buttons to trigger simulation, (**b**) Parameters for simulation, (**c**) Visualization of the true properties of simulated data (regression coefficients, true and estimated covariance between response and predictors components) (**d**) Additional analysis (**e**) Download option of simulated data.

Page 18

Figure 5 is updated.

Conclusion

Whether comparing methods or assessing and understanding the properties of any methods, tool or procedure; simulated data allows for controlled tests for researchers. However, researchers spend enormous amount of time creating such simulation tools so that they can obtain a particular nature of data. We believe that this tool along with the R-package and the easy-to-use shiny web interface will become an assistive tool for researchers in this respect.