Review on Simrel Paper

Raju Rimal

TODO Corrections in Paper

Introduction

Page 2:

- Add references (Johnson 2013) along with (Ripley 2009) and (Gamerman and Lopes 2006).
- In particular, methods based on covariance structure has been discussed by (Arteaga and Ferrer 2010, Arteaga and Ferrer (2013), Camacho (2017)), which follows an algorithmic approach to find simulated data satisfying the desired correlation structure.

Page3:

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

Relevant Components

Page 6

- Equation number to the expression of generating lambda from gamma. Subscript changed from j to i.
- ~~Here we assume that $\mathbf{w}$'s are independent and has mutinormal distribution wnith variances 1, hence $\mathbf{\Sigma}_{ww} = \mathbf{I}_m$.~~ In similar manner, a parametric representation of eigenvalues corresponding to Σ_{ww} is adopted as,

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

Here, the decline of eigenvalues becomes steep as η gets far from zero. At $\eta = 0$, all $\boldsymbol{w}$ will have equal variance.

Page 8

- $\cancel{\Sigma_{ww} = I_m}$

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

- $\cancel{\rho_w^2 = \Sigma_{zw}^t A^{-1} \Sigma_{zw} \Sigma_{ww}^{-1}}$

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

- $\cancel{\rho_{w_j}^2 = \sum_{i=1}^p \frac{\sigma_{ij}^2}{\lambda_i}}$ is replaced by,

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

- $\cancel{\rho_{w_j}^2 = \sum_{i \in \mathcal{P}_j} \frac{\sigma_{ij}^2}{\lambda_i}}$

$$\rho_{w_j}^2 = \sum_{i \in \mathcal{P}_j} \frac{\sigma_{ij}^2}{\lambda_i \kappa_j}$$

- $\cancel{\sigma_{ij} = \text{Sign}(\mathcal{S}_i)\sqrt{\frac{\rho_{w_j}^2 |\mathcal{S}_i|}{\sum_{k \in \mathcal{P}_j} |\mathcal{S}_k|} \lambda_i}}$

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

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

Implementation

Page 13

- `gamma` $\to \gamma$ (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 for comparison, we have chosen $n > p$ situation in this example.

- **Figure 3 Caption:** Here (a) is the covariance structure of 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

- ~~(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)),
  # Position of predictor components
  R2     = c(0.8, 0.7),
  # Coefficient of determination for each response components
  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"
)
```

- Figure-5 moved to this page
- **Figure-5 Caption:** Web interface of shiny application of 'simrel': (**a**) Buttons to trigger simulation, (**b**) Parameters for simulation – here with these settings 200 training sets (**n**) and 50 test sets (`ntest`) will be simulated with 15 predictor variables (**p**) and 4

response variables (`m`). The 4 response variables will have 2 true latent dimension which is referred as *response components.* The first response component is rotated together with third (uninformative) response component and second response component is rotated together with fourth (uninformative) response components (`ypos`). Out of 15 predictors, 5 will be relevant for first response component and 4 will be relevant for second response component (`q`). The relevant principal components of the predictor variables are *predictor components.* The 5 predictor variables which are relevant for first response components span the same space as the predictor components at position 1 and 2. Similarly, the 4 predictor variables which are relevant for second response components span the same space as the predictor components at position 3, 4 and 6 (`relpos`). The coefficient of determination for first response component is 0.8 and second response component is 0.7 (`R2`). The eigenvalues of predictor components decay exponentially by the factor of 0.6 (`gamma`) but the eigenvalues of response components are constant (but can be set to exponential decay) (`eta`). (**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.

Conclusion

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

TODO R package:: `simrel` [1/7]

- [ ] Convert `simulatr` → `simrel`
- [ ] Univariate simrel (`simrel`) → `unisimrel` function
- [ ] Include `mbrd` function for creating fractional design from old simrel package
- [ ] Update documentation with many examples and cases
- [ ] Introduce about shiny application and rstudio gadget on the readme file
- [ ] Link to package documentation site
- [X] Move the repository `therimalaya/simulatr -> simulatr/simrel`

TODO Shiny application for `simrel` [0/2]

- [ ] Fix covariance plot (bigger text, aspect ratio etc)
- [ ] Model fitting tab:: Coefficient comparison plot, estimation and prediction error plot

References

Arteaga, Francisco, and Alberto Ferrer. 2010. "How to Simulate Normal Data Sets with the Desired Correlation Structure." *Chemometrics and Intelligent Laboratory Systems* 101 (1). Elsevier: 38–42.

———. 2013. "Building Covariance Matrices with the Desired Structure." *Chemometrics and Intelligent Laboratory Systems* 127. Elsevier: 80–88.

Camacho, José. 2017. "On the Generation of Random Multivariate Data." *Chemometrics and Intelligent Laboratory Systems* 160. Elsevier: 40–51.

Gamerman, Dani, and Hedibert F Lopes. 2006. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference.* CRC Press.

Johnson, Mark E. 2013. *Multivariate Statistical Simulation: A Guide to Selecting and Generating Continuous Multivariate Distributions.* John Wiley & Sons.

Ripley, Brian D. 2009. *Stochastic Simulation.* Vol. 316. John Wiley & Sons.

<script type="text/x-mathjax-config"> MathJax.Hub.Register.StartupHook("TeX Jax Ready",function () MathJax.Hub.Insert(MathJax.InputJax.TeX.Definitions.macros, cancel : ["Extension","cancel"], bcancel : ["Extension","cancel"], xcancel : ["Extension","cancel"], cancelto : ["Extension","cancel"]);); </script>