LMM Covariance Structure Shiny App

Eduardo Coronado

11/18/2019

Concept

Sample covariance is important concept to understand when fitting linear mixed effects models since it changes as we include random effects (add heirarchies) into our models. We normally start with a simple fixed effects linear model with zero-centered Gaussian noise

$$y_{ij} = \beta_0 + \beta_{\cdot,j}; x_{ij} + \epsilon \qquad \epsilon \sim \text{ i.i.d. } N(0, \sigma^2)$$

where y is our response variable, x is our dependent variable and are indexed as i^{th} observation in the j^{th} group. Under this model, the main assumption is there isn't any correlation among observations at all (neither within group or between groups).

$$Cov(y_i) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} \quad \Rightarrow \quad Corr(y_i) = 0$$

This is problematic since in reality observations within-group have dependencies. Thus, we model using random effects we change this covariance structure since we allow for additional within-group variation. For example, for a mixed effects model with a group-specific random intercept $b_{0,j}$

$$y_{ij} = \beta_0 + \beta_{\cdot,j}; x_{ij} + + b_{0,j} + \epsilon$$
$$\epsilon \sim \text{ i.i.d. } N(0, \sigma^2) \perp b_{0,j} \sim \text{ i.i.d. } N(0, \tau^2)$$

we see the following changes in the covariance structure and intraclass correlation

$$Cov(y_i) = \begin{bmatrix} \sigma^2 + \tau^2 & \tau^2 & \cdots & \tau^2 \\ \tau^2 & \sigma^2 + \tau^2 & \cdots & \tau^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tau^2 & \tau^2 & \cdots & \sigma^2 \end{bmatrix} \quad \Rightarrow \quad Corr(y_i) = \frac{\tau^2}{\sigma^2 + \tau^2}$$

Using the same model formulation above but now adding an additional slope random effects $b_{1,j}$ as follows

$$\begin{bmatrix} b_{0,j} \\ b_{1,j} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \tau_{11} & \tau_{12} \\ \tau_{12} & \tau_{22} \end{bmatrix} \right)$$

we see further changes to the covariance and correlation structure

$$Cov(y_i) = \begin{bmatrix} \sigma^2 + \tau_{11} + \tau_{12} + 2\tau_{12} & \tau_{11} + \tau_{12} + 2\tau_{12} & \cdots & \tau_{11} + \tau_{12} + 2\tau_{12} \\ \tau_{11} + \tau_{12} + 2\tau_{12} & \sigma^2 + \tau_{11} + \tau_{12} + 2\tau_{12} & \cdots & \tau_{11} + \tau_{12} + 2\tau_{12} \\ \vdots & \vdots & \ddots & \vdots \\ \tau_{11} + \tau_{12} + 2\tau_{12} & \tau_{11} + \tau_{12} + 2\tau_{12} & \cdots & \sigma^2 + \tau_{11} + \tau_{12} + 2\tau_{12} \end{bmatrix}$$

$$\Rightarrow \quad Corr(y_i) = \frac{\tau_{11} + \tau_{12} + 2\tau_{12}}{\sigma^2 + \tau_{11} + \tau_{12} + 2\tau_{12}}$$

$

As seen above, deriving these covariance/correlation structures can be somewhat tedious and abstract for students at times. Therefore, the main goal of this shiny app `shiny_varcov_lmm.R` is to demonstrate how these covariance structures change in mixed effects models via interactive grid plot that visually constructs sample-wide group-specific variational structures.

Shiny App Functionality

The shiny was built with an instructor in mind, allowing for easy modifications to expand its capabilities to additional datasets as necessary (see Appendix).

Currently the app runs on two preset datasets used in STA610 class - the National Education Longitudinal Study (`nels.rds`) and Household Radon data (`srrs2.rds`). The user interface allows the instructor to select the dataset to be used to build a model and subsequent covariance/correlation visualization. The user interface has the following options once a dataset has been selected

- **Response variable**: Loads available variables based on the dataset, allows for a unique selection
- **Fixed effects**: Loads available variables (excluding the response selection), allows for multiple selection
- **Grouping variable (i.e. Random Intercept)**: Loads available variables (excluding the response selection), allows for a unique selection
- **Random Slope**: Loads available variables (excluding the response and grouping selections), allows for a unique selection
- **State selection** (*available for* `srrs2` *only*): Allows user to filter dataset for a specific state to reduce sample size to be rendered

As the user selects these, the app will show the formula it will use to fit the model. This is helpful to ensure the user is fitting the model he/she intends before running any computations or fits.

Once the user defines the model to be fit, the **Fit Model** button fits the model and generates the visualization. Once these are done, the app will show the formula it fit, the intraclass correlation ρ, as well as an interactive grid chart where the user can zoom-in into specific groups using the cursor and use the hover functionality to get the specific variance/covariance values of each sample.

Appendix

The shiny app can easily be updated to visualize new datasets as longs as these files are `.rds`. However, it is important that for datasets with many observations or groups additionals modifications are necessary when rendering the plots (i.e. samples $n > 1,000$ will require these additional modifications).

The visualization capabilities can be updated via the following: * Add `.rds` file to same folder as the shiny app script (`shiny_varcov_lmm.R`) * In the `ui` component of the shiny script, add the name of the `rds` file without the `rds` extension * Save and run

Again, as noted above, if the number of samples to be visualized is $n > 1000$ we need to add code to the `server` component of the shiny script where the data set is loaded to build the variance covariance structure, NOT to where it the model is fit.