Homework 4: Using the Lkj Prior

Zeki Kazan

11/23/2021

I. Conceptual Basis

In this report, I create a Shiny application to demonstrate how, in a Bayesian hierarchical model, the choice of parameter for an Lkj prior on the correlation matrix of the random effects will effect posterior inference. I begin by defining the Lkj prior. Recall that a $K \times K$ correlation matrix, $\mathbf{\Omega}$, has the form

$$\mathbf{\Omega} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1K} \\ \rho_{12} & 1 & \rho_{23} & \cdots & \rho_{2K} \\ \rho_{13} & \rho_{23} & 1 & \cdots & \rho_{3K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1K} & \rho_{2K} & \rho_{3K} & \cdots & 1 \end{pmatrix}.$$

A Lewandowski-Kurowicka-Joe (Lkj) distribution with parameter η on $\mathbf{\Omega}$ has kernel $|\mathbf{\Omega}|^{\eta-1}$. The exact distribution is given by the pdf

$$p(\mathbf{\Omega}) = 2^{\sum_{k=1}^{K-1}(2(\eta-1)+K-k)(K-k)} \prod_{k=1}^{K-1} \left(B(\eta + (K-k-1)/2, \eta + (K-k-1)/2)\right)^{K-k} |\mathbf{\Omega}|^{\eta-1}.$$

For simplicity, I will focus on the case where $K = 2$. In this case, $\eta = 1$ corresponds to a uniform prior on the correlation, $\rho = \rho_{12}$. When $\eta > 1$, the prior is concentrated on values around 0 and when $\eta < 1$, the prior is concentrated around ± 1.

In order to best illustrate use of the Lkj prior, I construct the simplest possible model where this prior would be reasonable. Let $j \in \{1, \ldots, J\}$ index the groups and $i \in \{1, \ldots, n_j\}$ index observations within each group. Let y_{ij} be the response variable and x_{ij} be a predictor. I will generate data and model via a mixed effects model with a random intercept and slope.

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}, & \varepsilon_{ij} &\overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \\ \beta_{0j} = \beta_0 + b_{0j}, &\quad \beta_{1j} = \beta_1 + b_{1j}, & \begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} &\overset{iid}{\sim} \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}) \end{aligned}$$

See Section III for details on the data generation. I decompose the random effect covariance, $\mathbf{\Sigma}$, into

$$\mathbf{\Sigma} = \begin{pmatrix} \tau_0 & 0 \\ 0 & \tau_1 \end{pmatrix} \mathbf{\Omega} \begin{pmatrix} \tau_0 & 0 \\ 0 & \tau_1 \end{pmatrix},$$

for correlation matrix $\mathbf{\Omega} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The prior for $\mathbf{\Omega}$ is, of course, $\mathbf{\Omega} \sim \mathsf{LkjCorr}(\eta)$. The other priors are set to generic weakly informative choices (see Section III for details). The model is fit in `brms` with 4 chains of 4,000 post-warmup iterations (and 1,000 warmup iterations).

II. Using Shiny

The shiny application allows for the user to select two values. The first value is ρ, the true correlation between the random intercept and the random slope. ρ is selected using a slider input widget, which allows ρ to be set to any multiple of 0.05 between -0.95 and 0.95. The second value is η, the parameter for the Lkj prior on the correlation matrix. η is selected using a slider input widget which allows for $\eta \in \{0.1, 0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 10\}$, allowing for a variety of different prior shapes.

For whatever inputs the user selects, several output plots within the Shiny application will automatically adjust. The first two plots are a plot of the generated data (that is y_{ij} plotted against x_{ij}, colored by the group j) and a plot of the true (generated) intercept and slope for each group j. Only the choice of ρ will effect these plots. The next plot is of the prior distribution for $\boldsymbol{\Omega}$, which varies based on the choice of η, with a dashed line showing the selected value of ρ for comparison. Finally, the last plot overlays the prior and samples from the posterior, again with a dashed line showing the selected value of ρ for comparison.

The shiny application can be run via the file `hw4_shiny.R`, which is included on Gradescope or in the GitHub repository https://github.com/zekicankazan/STA-610-hw4. Cached posterior samples for every combination of ρ and η are saved as CSV files to the folder `hw4_cache`, which is compressed as `hw4_cache.zip`. Additional files include `hw4_cacheing.R`, which includes the code to produce the CSV files and `hw4.Rmd`, which includes the code to produce this document.

Note that because the posterior samples are cached, the user may need to adjust the file path in the `read_csv` command in `hw4_shiny.R`. This command is on line 139 of the code and looks like

```
read_csv(paste0("hw4_cache/",
```

The user should change `hw4_cache/` to be the path to whatever folder the CSV files are stored in.

III. Additional Details

- I set the number of groups to $J = 8$, the number of observations per group to $n_j = 25\ \forall j$, the fixed effects to $\beta_0 = 5$ and $\beta_1 = 2$, the random effect standard deviations to $\tau_0 = 2$ and $\tau_1 = 1$, and the residual standard deviation to $\sigma = 1$. Predictor values x_{ij} are drawn from a standard normal distribution. The code to generate the data is included below

```
J <- 8; n <- rep(25,J); beta0 <- 5; beta1 <- 2; tau0 <- 2; tau1 <- 1; sigma <- 1

tau_mat <- diag(c(tau0, tau1))
Omega <- matrix(c(1,rho,rho,1),ncol=2)
Sigma <- tau_mat %*% Omega %*% tau_mat

beta <- rmvnorm(J, mean = c(beta0, beta1),  sigma = Sigma)

x <- c(); y <- c()
for(j in 1:J){
  xj <- rnorm(n[j])
  yj <- rnorm(n[j], beta[j,1] + beta[j,2]*xj,sigma)
  x <- c(x,xj); y <- c(y,yj)
}
```

- Weakly informative prior choices for the other parameters: The priors for τ_0, τ_1, and σ are set to their `brms` defaults. That is, for MAD$(\{y_{ij}\})$ the mean-absolute deviation of the response,

$$\tau_0, \tau_1, \sigma \overset{iid}{\sim} t_3^+(0, \max\{2.5, \text{MAD}(\{y_{ij}\})\})$$

Rather than using the default flat priors, the priors for β_0 and β_1 are set to scaled dispersed normal distributions. That is, if s_x is the sample standard deviation of x_{ij}, s_y is the sample standard deviation of y_{ij}, and $\bar{y}$ is the grand mean, the priors are

$$\beta_0 \sim \mathcal{N}(\bar{y}, 10s_y), \qquad \beta_1 \sim \mathcal{N}(0, 2.5s_y/s_x)$$

- The Stan Functions Reference recommends decomposing $\mathbf{\Omega} = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is a lower-triangular Cholesky factor of $\mathbf{\Omega}$, and putting a prior on $\mathbf{L}$ that implies $\mathbf{\Omega} \sim \mathsf{Lkj}(\eta)$. This method is faster, is more numerically stable, and uses less memory than putting the prior directly on $\mathbf{\Omega}$. Thus I use this method to set the prior. The code to generate posterior samples in `brms` is presented below.

```
Int_prior <- paste0("normal(",as.character(mean(y)),", ", as.character(10*sd(y)),")")
b_prior <- paste0("normal(0, ", as.character(2.5*sd(y)/sd(x)),")")
L_prior <- paste0("lkj_corr_cholesky(", as.character(eta), ")")

prior <- c(set_prior(Int_prior, class = "Intercept"),
           set_prior(b_prior, class = "b"),
           set_prior(L_prior, class = "L"))

mod <- brm(y ~ x + (x | group), data = data, cores = 4, chains = 4,
           warmup = 1000, iter = 5000, prior = prior, seed = 1)
```

- For plotting the Lkj prior, I use the fact that the implied marginal distributions of the correlations are

$$\frac{\rho_{k_1 k_2} + 1}{2} \sim \mathsf{Beta}\left(\eta + 1 - \frac{K}{2}, \eta + 1 - \frac{K}{2}\right).$$

IV. References

- Original paper defining the Lkj distribution: *Daniel Lewandowski, Dorota Kurowicka, Harry Joe, Generating random correlation matrices based on vines and extended onion method, Journal of Multivariate Analysis, Volume 100, Issue 9*

- Useful reference for facts about the distribution: https://distribution-explorer.github.io/multivariate_continuous/lkj.html

- STAN function reference for Lkj distributions: https://mc-stan.org/docs/2_28/functions-reference/correlation-matrix-distributions.html