

Multi-level regression models (MLM) combine multiple hierarchies of data into a single regression function by modeling the regression coefficients as random variables with prior distributions. This paper presents a MLM with the intercept coefficient as a random variable. Population regression lines representing $j=3$ groups are simulated using normal distributions for the purpose of comparing model estimates to specified parameters. The MLM is estimated in WinBUGS using a Gibbs Sequence. An analysis shows convergence for all variables. While the estimates all fall within 95% Bayesian uncertainty intervals, the group-level variance has a long tail. Group-level variation is compared to individual-level variation. I conclude with some recommendations for further research.

Blake Shurtz

Department of Statistics and Biostatistics

Cal State University, East Bay

Keywords: Hierarchical regression, multi-level modeling, WinBUGS

(1) Recent increases in computing power have made stochastic estimation of posterior distributions faster and cheaper. One application of the stochastic/Bayesian approach which is particular to linear regression is Multi-level modeling (MLM). MLM can be thought of as a Bayesian linear regression, using group-level data as a prior distribution for estimating the coefficients for the individual-level response. Common examples of hierarchical structures are “people within cities” or “players within teams”. In a MLM, and specifically for this paper, I have predictors on both levels and wish to combine them into a single regression function in order to measure the individual-level response.

(2) Although a MLM reduces to a single regression equation, the model can be conceptually divided into two separate equations, one for each level of data. As there are two levels of data, there are two regression equations.

At the individual-level, the MLM has a single, continuous response variable and a single, binary predictor. The intercept is a random variable and the slope is a fixed parameter. In Bayesian terms, the individual-level regression is the likelihood function as it represents the data we use to estimate our population parameters.

$$y_i = \alpha_{j[i]} + \beta x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma_y^2), \quad i = 1, \dots, n.$$

y_i is the response variable for observation i . $\alpha_{j[i]}$ is the random variable for the intercept coefficient. For each group j , there is a unique intercept that is a function of the group-level data (see below). β is the slope parameter which is assumed to be constant across all groups. x_i is the binary predictor. The error term ε_i is assumed to be normally distributed around zero with a constant variance σ_y^2 .

$\alpha_{j[i]}$ is modeled as a random variable with its own prior distribution, which is assumed to be normal with mean μ_α and group-level variance σ_α^2 . The group-level regression equation represents the prior distribution in the MLM.

$$\alpha_j = \gamma_0 + \gamma_1 \mu_j + \eta_j, \eta_j \sim N(0, \sigma_\alpha^2) \quad \alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$

γ_0 and γ_1 are unmodeled parameters with non-informative normal prior distributions. μ_j is the group level data. η_j is the error term with constant group-level variance σ_α^2 . Initially, a positive, non-uniform prior distribution was used. However, this led to a strongly right-skewed distribution that distorted the interpretation of the variances. A Gamma(1,2) prior was used to maintain some skewness in the pdf while providing a mean coefficient that was equivalent to a point estimate that was calculated through other methods in R.

(3) All data was simulated using R. The reason for doing this was for the purpose of comparison. If the goal is to test a MLM against a classical regression, then we need a standard with which to compare the models.

Three population regression lines are hypothesized, one for each of the three groups:

$$E(Y_1) = 2x_i + \varepsilon_i, \varepsilon \sim N(0, \sigma_y^2)$$

$$E(Y_2) = 6 + 2x_i + \varepsilon_i, \varepsilon \sim N(0, \sigma_y^2)$$

$$E(Y_3) = 12 + 2x_i + \varepsilon_i, \varepsilon \sim N(0, \sigma_y^2)$$

The intercept coefficients vary for each group. The slope coefficient is constant. The domain of the regression function is (0, 1) corresponding to the binary predictor. The three groups in the model meet the minimum requirement to prevent convergence to the classical regression model.

In simulating the distributions, normal distributions were chosen to ensure homoscedasticity of the residuals. The mean of each normal distribution corresponds to the $E(Y)$ for each particular “group/predictor” combination.

```
sample_size <- 15; SD <- 2
group0A <- rnorm(sample_size, 6, SD)
group1A <- rnorm(sample_size, 8, SD)
Code for Group 1:  $E(Y_A) = 6 + 2x_i$ 
```

The sample size for each group is 30 observations, which meets the simple assumptions for the t-statistic while preventing unrealistically high R^2 values in classical regression models. A standard deviation of $\sigma_y = 2$ allows for some overlap of \hat{y}_i values for each group.

Group-level data were given an ordinal ranked value of $\mu_j = 0, 1, 2$ for groups 1, 2 and 3 respectively. By assigning group predictors that are positively correlated with the intercepts, I would assume that the group-level data can have some explanatory power.

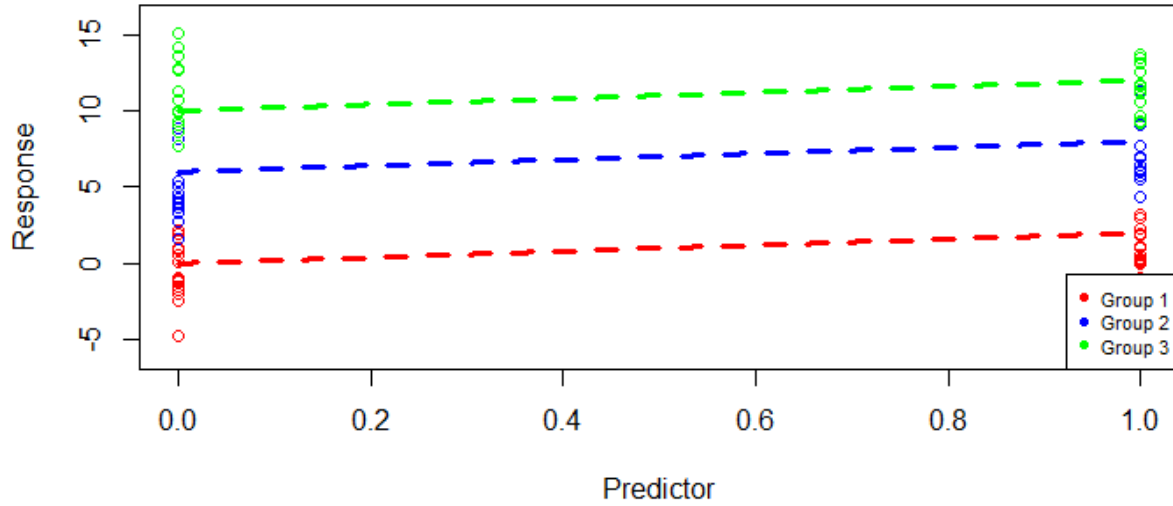


Figure 1: Population regression lines and simulated data.

(4) For each parameter, four Markov chains with 1000 iterations were simulated. Thus, the Markov-Chain Monte Carlo array contained 4 matrices of 9 variables by 1000 values.

For all parameters, \hat{R} values are equal to one, indicating that the model mixed well despite not including a “burn in” period.

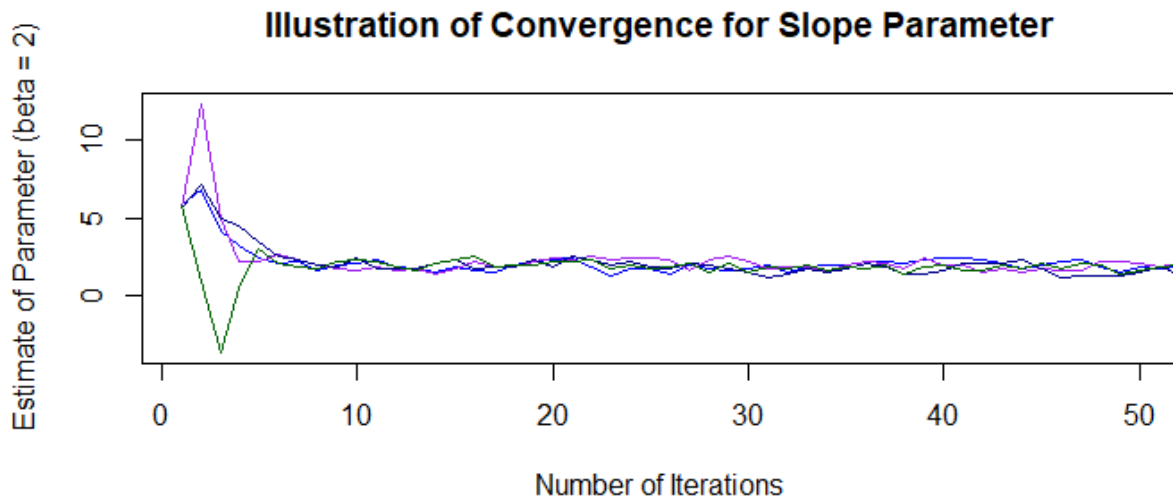


Figure 2: 4 chains showing convergence for the slope. The parameter converges to a median estimate of 1.6.

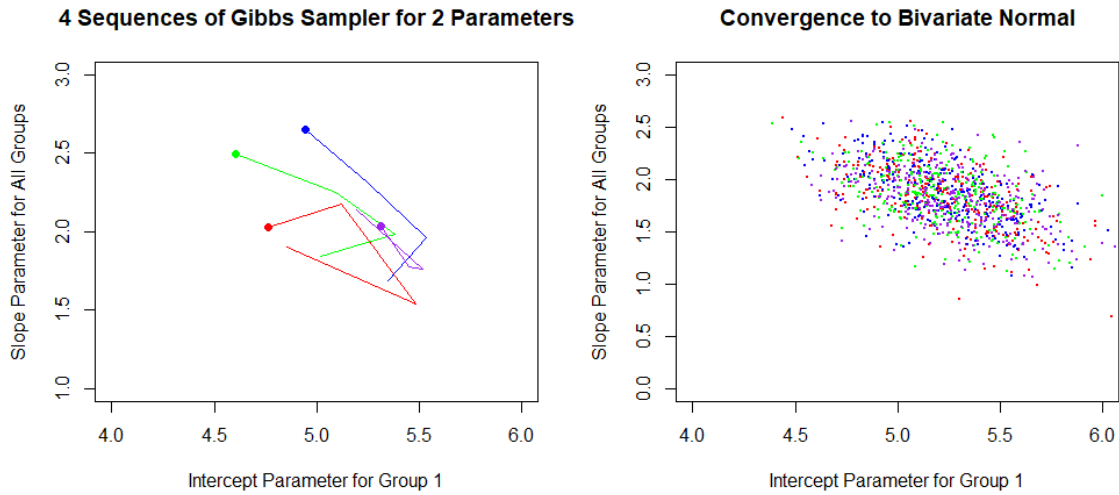


Figure 3: A few initial stages in Gibbs sampling of two coefficients. Right: Convergence to bivariate normal.

(5) Multi-level modeling was performed using WinBUGS software as run through the R package R2WinBUGS. Point-level estimates of coefficients we estimated using the lme4 package.

All of the specified parameters, including the individual-level variance, fall within the uncertainty intervals.

Predictor	Parameter	2.5%	50%	97.5%
α_1	2	-1.3	-0.5	0.3
α_2	4	4.5	5.4	6.1
α_3	5	11.6	12.3	13.2
β	2	0.8	1.6	2.4
γ_0	NA	-2.3	-0.7	-0.3
γ_1	NA	5.0	6.4	7.6
σ_y	2	1.7	1.9	2.3
σ_a	NA	0.0	0.5	1.8

Figure 4: Plot of Uncertainty Estimates as Inferred from BUGS model

The probability density functions for the parameters are normally distributed with the exception of the group-level variance, which is right skewed.

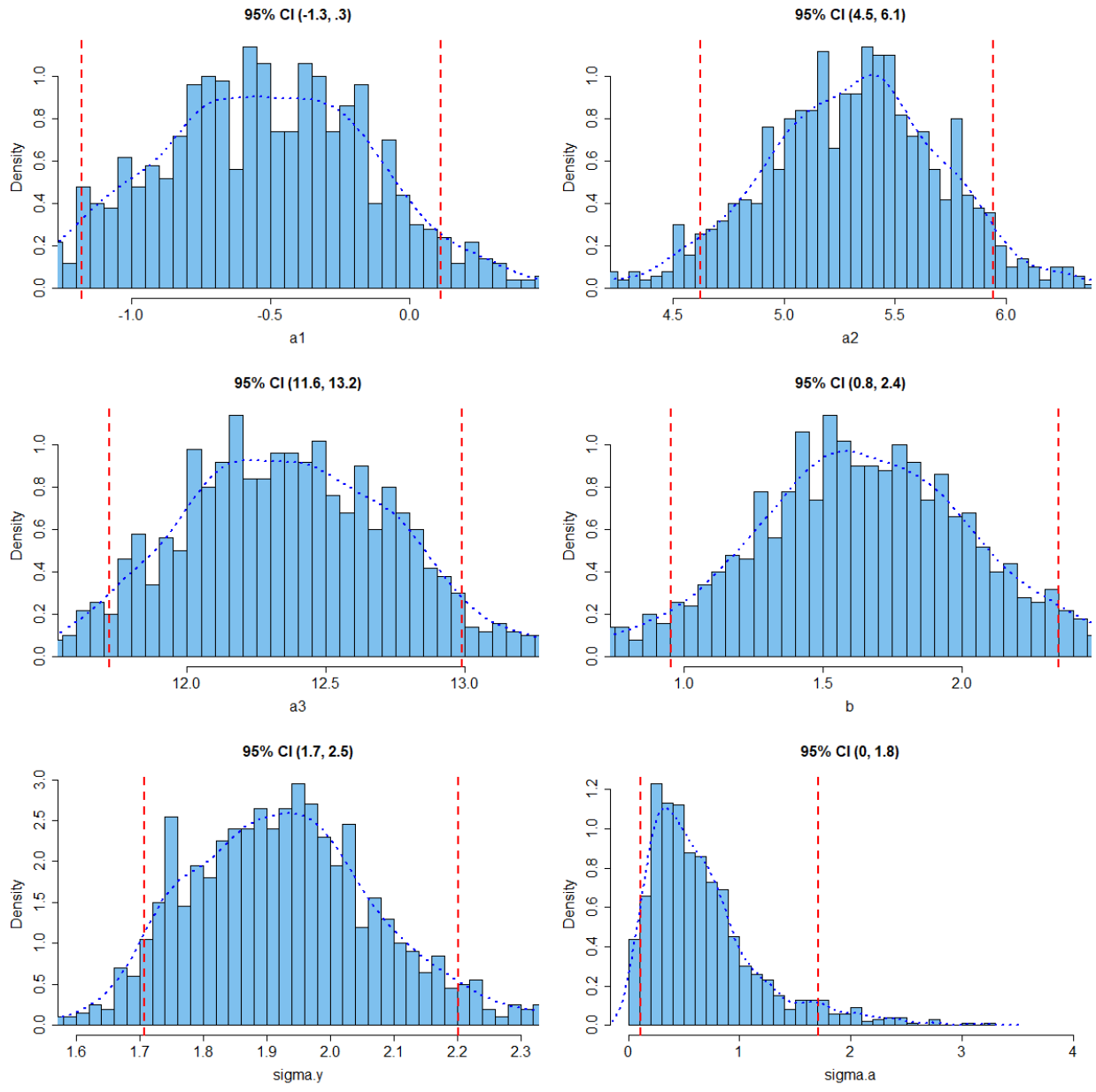


Figure 5: Plot of posterior distributions with 95% uncertainty intervals

The rightly-skewed distribution for the group-level variance was tampered by fitting a Gamma (1, 2) prior. At this value, the point estimate for the group-level variance is 0.7.

The intra-class correlation for the variances is about 12%, which suggests that about 12% of the variation can be explained by the group-level predictors. There is a sample size tradeoff of about 7.5, suggesting that you need about 7.5 observations before the individual-level variance explains more of the response than the group-level variance. A bivariate plot shows a single modality, but the right-skew of the group-level variance is evident and remains unexplained.

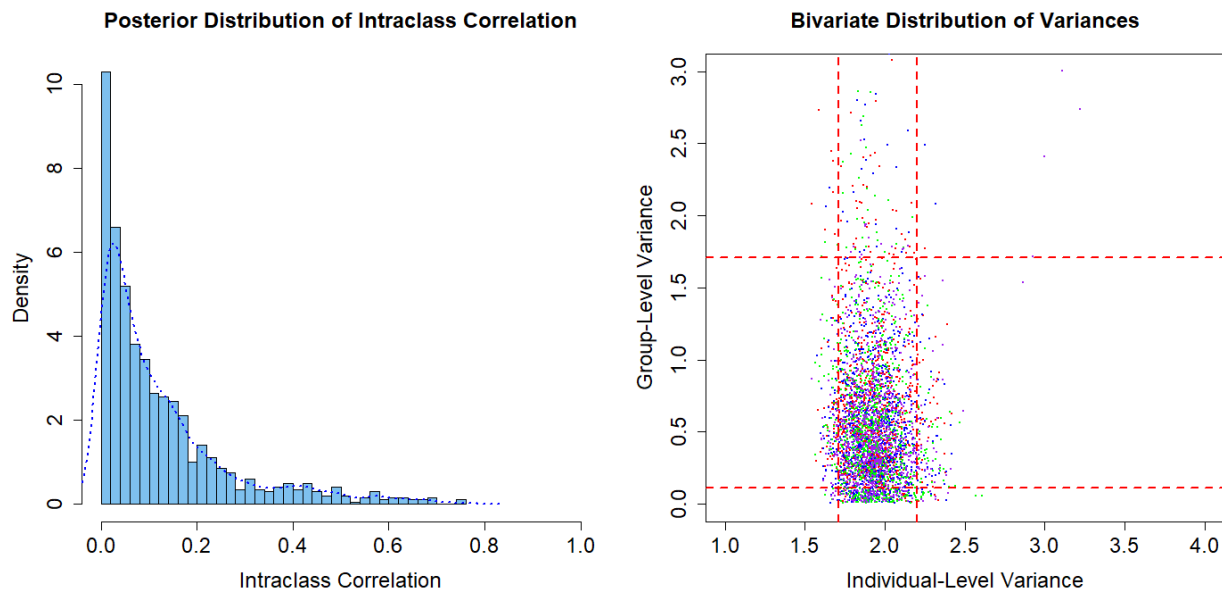


Figure 5: Plot of posterior estimate of intra-class correlation and distribution of variances.

(6) Future research topics can look for an explanation for the non-normal distribution of the group-level variance. Any or all of these variables may be having an effect and may therefore need to be tested: individual sample size, number of groups, individual variance, distances between intercept coefficients between groups.

Different seeds can be utilized also. Although randomness is an assumption and the seed should not be considered a variable per se, the number of outliers and skewness of the sample data may also be having an impact on the skewness of the group-level variance.

Bibliography

Gelman, Andrew and Jennifer Hill. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, 2007.

Suess, Eric and Bruce Trumbo. Introduction to Probability Simulation and Gibbs Sampling with R. Springer, 2010.