Predictions across groups in multi-level models

N. Thompson Hobbs and Mevin B. Hooten

March 30, 2019

1 Overview

Multi-level models facilitate statistical inference on designs where there are multiple groups that contain observations of responses and covariates within them. Predicting new observations and the mean of new observations at the group level is straightforward. However, prediction across groups is a problem that is not well treated in existing texts, including widely used ones (e.g. Gelman and Hill, 2009; Gelman et al., 2013). Here, we develop a general approach for making predictions across groups in multi-level models.

Consider the hierarchical model

$$
\begin{aligned}
g(\beta_{0_j}, \boldsymbol{\beta}, \mathbf{x}_{ij}) &= \beta_{0_j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} \\
y_{ij} &\sim [y_{ij} \mid g(\beta_{0j}, \boldsymbol{\beta}, \mathbf{x}_{ij}), \sigma_y^2] \\
\beta_{0j} &\sim [\beta_{0j} \mid \mu_{\beta_0}, \sigma^2_{\beta_0}],
\end{aligned}
$$

which can be modified by appropriate transformation and moment matching to accommodate many types of data in the general linear modeling framework. The quantity y_{ij} is the i^{th} observation in group j group j, $i = 1, ..., n_j$.

2 Scales of inference

Inference across groups consists of three distinct types of predictions: predictions that are limited to the groups studied, predictions for groups that are a finite sample of a relatively small population, and predictions to all possible groups where the sample is a small part of the total population.

We treat the first and last type of inference in this brief document to allow you to proceed with the laboratory exercise. We will provide a more thorough treatment of all three types of inference later in the semester.

3 Prediction limited to a specific set of groups

3.1 Algorithm

We first consider the case where we desire inference across groups using the observed values of covariates, that is, predictions using in-sample values. You might seek such predictions when you simply want to portray the model fit to the data conditional on the sites studies and the covariate values observed. For example, presume you are interested in temporal trends in cover of native species in a desert grassland as it is modified by annual rainfall, which we presume can be measure at a set of sites (= groups). Covariates are sequential years $\boldsymbol{x}_1$ and rainfall $\boldsymbol{x_2}$

We compute the following quantities at each iteration (k) of the MCMC algorithm:

$$\begin{aligned}
j^k &\sim \text{discrete uniform}(1, J) \\
i^k &\sim \text{discrete uniform}(1, n_{j^k}) \\
\tilde{y}^k &\sim [\tilde{y}^k \mid g(\beta_{0j^k}^k, \boldsymbol{\beta}^k, \mathbf{x}_{i^k j^k}), \sigma_y^{2k}] \\
\tilde{\mu}^k &= g(\beta_{0j^k}^k, \boldsymbol{\beta}^k, \mathbf{x}_{i^k j^k}).
\end{aligned}$$

The quantity $\tilde{y}^k$ represents a prediction of a new observation and $\tilde{\mu}^k$ is the mean of new observations at observed values of covariates for group j^k. We calculate statistics of interest on $\tilde{y}$ or $\tilde{\mu}$ (e.g., medians, credible intervals) using the K converged iterations. Note that there is no i or j subscript on $\tilde{y}^k$ and $\tilde{\mu}^k$ because we compute one value of $\tilde{y}^k$ and $\tilde{\mu}^k$ at each iteration.[1]

This approach can be easily modified to accommodate intercepts and slopes that vary by group simply by adding a subscript j^k to the vector of coefficients $\boldsymbol{\beta}_{j^k}$ where the distribution of slopes and intercepts across groups is $\boldsymbol{\beta}_j \sim$ multivariate normal$(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma})$.

A key idea here is that we do not seek inference to any population from which the sites were drawn, but rather focus on predicting new observations and the mean of new observations at the sties studied and the covariates observed.

[1] We could compute values for each replicate without altering the results, but it is not necessary to do so because different replicates will be drawn at different iterations.

3.2 Plotting predictions

It is important to note that one of the values of the covariate, i.e. the first or second element of the covariate vector $\mathbf{x}_{ij}$, can be specified at points chosen by the analyst to facilitate plotting of means, medians, and credible intervals as a function of the specified values of that covariate. If there is a single covariate, specifying its value obviates the need to make a draw of i^k to produce plots.

4 Predictions across all possible groups

4.1 Algorithm

Inference on all possible groups including those that were *not* observed can be achieved for the random intercept model[2] as follows. We draw a new vector of covariates $\tilde{\mathbf{x}}^k$ from the distribution of covariate values with parameters mean and variance of each covariate computed across all replications and groups. We then draw a new intercept from the distribution of intercepts $\tilde{\beta}_0^k \sim [\tilde{\beta}_0^k \mid \mu_{\beta_0}^k, \sigma_{\beta_0}^{2k}]$ and obtain

$$\tilde{y}^k \sim [\tilde{y}^k \mid g(\tilde{\beta}_0^k, \boldsymbol{\beta}^k, \tilde{\mathbf{x}}^k), \sigma_y^{2k}]$$
$$\tilde{\mu}^k = g(\tilde{\beta}_{0_j}^k, \boldsymbol{\beta}^k, \tilde{\mathbf{x}}^k)$$

at each MCMC iteration. The quantity $\tilde{y}^k$ represents a prediction of a new observation from a new group and $\tilde{\mu}^k$ is the mean of new observations from a new group.

This approach assumes that sites are chosen completely at random from the population of possible sites, which is to say all sites are equally likely to be chosen. This means that the way the data were collected does not need to be specifically included in the analysis— in technical terms, the design is *ignorable* (Gelman et al. 1992). Information on the data collection process must be included in analysis of non-ignorable sampling or experimental designs.

As a brief example, consider a stratified random sample where groups were chosen randomly within $k = 1, ..., K$ strata. We find separate random intercepts and fixed slopes for each strata. In

[2] There is no simple way to make inference across all possible groups for models including both random intercepts and slopes. The reason for this limitation is that there is no simple way to make a draw of covariates that are properly aligned with a draw from the distribution of slopes. However, it *is* possible to make a draw of intercepts and slopes, specify values for a single covariate and hold the others at their means to facilitate plotting.

this case the draws from the distributions of sites and intercepts would need to be weighted by the proportion of the number of sites in strata k relative to the total number of potential sites. Thus the probability of drawing from strata k is $\frac{N_k}{N}$ where N_k is the total number of potential sites within strata k and $N = sum_{k=1}^{K}$ (Gelman et al. 2013 : page 207).

4.2 Plotting predictions

Again, values of one of the covariates can be specified to facilitate plotting predictions against the values of that covariate.

5 Acknowledgements

Conversations with Luke Zachmann and Chris Che-Castaldo helped formulate these ideas.

Literature Cited

Gelman, A., J. B. Carlin, H. S. Stern, D. Dunson, A. Vehhtari, and D. B. Rubin, 2013. Bayesian data analysis. Chapman and Hall / CRC, London, UK.

Gelman, A. and J. Hill, 2009. Data analysis using regression and multilievel / hierarchical modeling. Cambridge University Press, Cambridge, UK.