Hierarchical Bayesian Modeling

What is it?

Hierarchical models (also known as multilevel models or mixed effects models) are an extension of linear modeling which account for a hierarchical structure in the data such as ecosubsections within ecosections or even provinces. This thesis implements *Bayesian* hierarchical models to accomplish small area estimation on forest attributes, meaning that these models do not only take into account the hierarchical structure of the data, but they also incorporate prior information about the covariates in the model. To best understand a Bayesian hierarchical model, I will first begin with the standard OLS regression form. We have

$$Y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon_i$$
$$= \vec{\beta} X_i + \epsilon_i$$

where Y_i is the response variable for the ith observation and all observations are assumed to be independent. In this case, β_0 is the intercept term, ϵ_i is the unexplained error, and $x_1, \ldots, x_p$ are predictors or "covariates" with coefficient estimates $\beta_1, \ldots, \beta_p$ fit with the methods of Least Squares. The method of Least Squares estimates the intercept and slopes of these covariates by fitting a hyperplane which minimizes sum of squared residuals, where each residual is defined as $e_i = Y_{observed} - Y_{expected}$.

Bayesian hierarchical modeling does not use the method of Least Squares to estimate parameter values, instead we specify priors on each covariate and then sample from the posterior distribution using Markov Chain Monte Carlo (MCMC) methods. The most common MCMC method for hierarchical modeling is the Gibbs Sampler. The Gibbs Sampler is useful because it uses the conditional distributions for each variable (which we will specify in our model) to estimate the joint distribution of our response variable.

Before we dig much deeper into how to use the Gibbs Sampler in this case, we should specify the model. I will first specify a very simple multilevel model and we will work up from there. With a multilevel model, we can choose to vary to slope(s), the intercept, or both. First, let's consider a varying intercept model with j levels:

$$Y_i \sim N(\alpha_j + \vec{\beta} \vec{X}_i, \ \sigma^2)$$
$$\alpha_j \sim N(\mu_\alpha, \ \sigma_\alpha^2)$$

In this model, we have the response variable, Y, which is modeled to have a Gaussian posterior distribution with mean $\alpha_j + \vec{\beta} \vec{X}_i$ which can change intercept based on the level that a given observation is in. Another type of Bayesian Multilevel Model is the varying slopes model, where the coefficient of each parameter estimate can change based on the level of the observation. We specify that model as follows:

$$Y_i \sim N(\alpha + \vec{\beta}_j \vec{X}_i, \ \sigma^2)$$
$$\vec{\beta}_j \sim N(\vec{\mu}_\beta, \ \sigma_\beta^2)$$

In this model, the intercept of each level is set to be the same, however the coefficients (β's) can change between groups. Neither of these models are too common, as it often makes sense to allow both the slope and intercept to vary across levels. From these first two models, it is easy to imagine what this varying-intercept-and-slopes model will look like:

$$Y_i \sim N(\alpha_j + \vec{\beta}_j \vec{X}_i, \ \sigma^2)$$
$$\alpha_j \sim N(\mu_\alpha, \ \sigma_\alpha^2)$$
$$\vec{\beta}_j \sim N(\vec{\mu}_\beta, \ \sigma_\beta^2)$$

This multilevel model allows for both the slopes and intercept to vary between groups by placing prior distributions on each which will update simultaneously between groups when we use the Gibbs sampler to sample from the posterior distribution. While these groups are specified to allow for different parameter estimates and intercept, they are also not unaware of each other. The model borrows strength from other levels and uses this information to help create the posterior distribution as well.

Why use it?

Now that we have explicitly written the model form out we can consider why one may want to use multilevel modeling in their research. I will focus on the main benefits of multilevel modeling to small area estimation. First, we have improved estimation for imbalance in sampling: multilevel models will automatically deal with issues of over- and under-sampling of clusters by assigning differing uncertainty across clusters. Also, when we are dealing with groups within the data, the multilevel models will model the variation between groups explicitly. Finally, with multilevel models we can avoid pre-averaging our data. This allows us to retain the variation in our data to be fully used in the model and we do not lose some valuable insight that this variation could give us. (Source: Statistical Rethinking, edition 2, page 414)

Example

As noted above, this thesis estimates forest attributes by implementing Bayesian Multilevel Models. To get a concrete sense of Bayesian Multilevel Models and the forestry data, consider a multilevel model estimating mean basal area per acre with one explanatory variable: forest biomass. Mean basal area is the average amount of area that tree stems occupy when we measure their area at breast height while forest biomass is the total biomass of the trees in the forest.

We could naively model this for the entire interior west as a linear function using OLS and get the following output:

$$\hat{Y}_i = 8.54 + 2.13x_i \quad (R^2 = 0.44)$$

This gives us a relatively poor fitting model that would not be any use for us in terms of prediction. Perhaps we want to know about basal area in some given small area, say an ecosubsection. Rather than using ecosubsection as an indicator variable, we can use Bayesian Multilevel Modeling which allows us to help update our priors for a given ecosubseciton with data from the surrounding area, say the ecosection that ecosubsection resides in. To understand this technique, let's specify the model form:

$$Y_i \sim N(\alpha_j + \beta_j X, \ \sigma^2)$$
$$\alpha_j \sim N(0, \ 10)$$
$$\beta_j \sim N(0, \ 10)$$
$$\sigma \sim Exp(1)$$

Here, Y_i is the mean basal area for a observation with a given ecosubsection, and we are allowing its slope and intercept to vary for each ecosubsection. The slope and intercept are given relatively uninformative priors, both Gaussian with mean 0 and variance 10. We must also put a prior on the variance of Y_i, and so we specify

```
library(hbsae)
```