Introduction to Bayesian Data Analysis

Hierarchical Models

Alix I. Gitelman

Statistics Department
gitelman@science.oregonstate.edu

July 27, 2016

Hierarchical Models

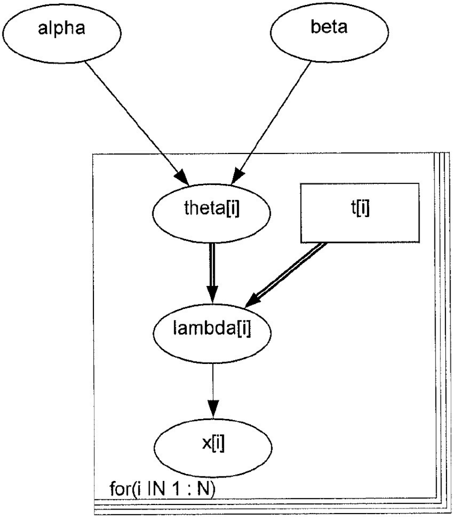
- General Ideas
- Example

Hierarchical/Multilevel Models

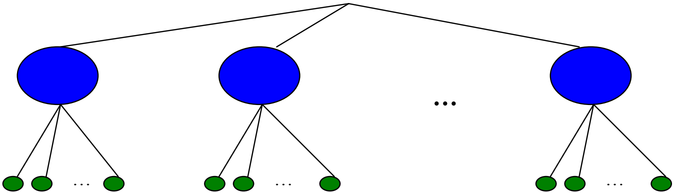
The idea of a hierarchical or multilevel model is that instead of having prior distributions where we specify values for the hyper-parameters, we actually model those hyper-parameters using priors and possibly additional covariate information.

- This may seem like a lot of extra modeling—and all in the prior distributions—but there are many situation in which it's actually quite natural to do this.
- We often have observations grouped together in some natural way—test scores of children in classrooms; repeated vital measurements on patients; measurements on trees within forest stands; measurements on ice cores taken in close proximity; etc.

The Pumps Example



Hierarchical/Multilevel Structure



Hierarchical/Multilevel Models

A hierarchical model allows us to account for the fact that observations that are grouped together—in time, space, classrooms, etc.—tend to be more correlated than observations that are not grouped together.

- When we do a better job of modeling the correlation among observations, we typically get more realistic estimates of the variability in the things we're really trying to make inference about.
- We might also be interested in studying the contributions to the overall variance in a system that come from the different levels of the system (e.g., how much of the variation in tree biomass has to do with covariate information, stand structure, larger-scale spatial structure, etc.).

The Lettuce Lichen Example

- We had $X \sim Bin(57, \pi)$, and I took $\pi \sim Unif(0, 1)$
- But suppose I now tell you that the researchers didn't just visit one forest stand, they visited 10 forest stands, and at each forest stand they looked at a sample of trees.
- We could just take the 10 different Binomial observations and use the same model as before:

$$X_1, \ldots, X_{10} \overset{ind}{\sim} Bin(n_j, \pi), \qquad \pi \sim U(0, 1)$$

(n_j is the number of trees at the j^{th} stand, $j = 1, \ldots, 10$).

  - But what if there are substantively different characteristics in the different stands of trees (e.g., slope, aspect, elevation)?

Hierarchical/Multilevel Models

A hierarchical model in this situation allows us to compromise between taking all of the presence probabilities to be identical and taking all of the presence probabilities to be unrelated (i.e., and essentially fit 10 separate models):

$$
\begin{aligned}
X_j &\sim Bin(n_j, \pi_j) \\
\pi_j &\sim Beta(\alpha, \beta) \\
\alpha &\sim [\alpha] \\
\beta &\sim [\beta]
\end{aligned}
$$

- This model gives each stand to have a distinct presence probability, π_j, but it connects those π_j's via a common distribution, the $Beta(\alpha, \beta)$ distribution.

Hierarchical Models, Random Effects, etc.

Connections between models:

- The pump example is a hierarchical model; we can also consider it to be a random effects model (pump failure rates are random).
- The lettuce lichen example is a hierarchical model; we can also think of it as a random effects model (stand-level probabilities are random).
- I'm about to show another example of a hierarchical model that we also call a random coefficients model (stand-level regression coefficients are random)

Example: Western Hemlock Establishment

- Western Hemlock (*Tsuga heterophylla*) is a key structural component of old-growth forests in the Pacific Northwest.
- It typically provides a multi-layer canopy and contributes to the diversity of tree ages (both features of old-growth forests).
- Forest managers would like to promote Western Hemlock establishment in the hope of accelerating development of old-growth characteristics.

Western Hemlock Establishment

Questions:

1. What is the relationship between hemlock establishment and the amount of coarse woody debris (CWD; large pieces of dead and decayed wood) *locally*, meaning at the micro-site level?

2. If there is a relationship between hemlock establishment and CWD locally, is that relationship different in forest stands with larger amounts of CWD at the stand level than it is in forest stands with less CWD at the stand level?

Western Hemlock Data

- 15 mature (approximately 80–120 years old), undisturbed forest stands in the northern Oregon Coast Range.
- Within each stand: between 5 and 18 hemlock saplings and between 4 and 11 points without a hemlock sapling, randomly selected.
- Each sample point—either with or without a sapling—represents the micro-site level.

Western Hemlock Data

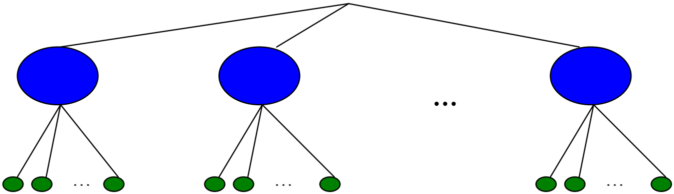
At each micro-site, a $2m$ diameter circular plot was used to measure the amount of local CWD.

- Locally, CWD was measured as a ground-cover area of all dead logs and snags. CWD at the micro-sites ranged from 0 to $0.78m^2\text{CWD}/m^2$.

At each stand, and independently of the local CWD measurements, CWD was sampled along a $500m$ transect.

- At the stand level, CWD was measured as a ground-cover volume of dead logs. CWD at the stands ranged from 140 to $190m^3/ha$.

Hierarchical/Multilevel Structure



Building a Statistical Model

Next, I'll construct a hierarchical generalized linear model for these data that is appropriate for answering the questions of interest:

1. What is the relationship between hemlock establishment and the amount of coarse woody debris (CWD; large pieces of dead and decayed wood) *locally*, meaning at the micro-site level?

2. If there is a relationship between hemlock establishment and CWD locally, is that relationship different in forest stands with larger amounts of CWD at the stand level than it is in forest stands with less CWD at the stand level?

Micro-site Level Model

Responses are 0/1 (presence/absence) of hemlock. For micro-site i at stand j let

$$Y_{ij} = \begin{cases} 1 & : \quad \text{hemlock is present} \\ 0 & : \quad \text{otherwise} \end{cases}$$

for $i = 1, \ldots, n_j$ and $j = 1, \ldots, 15$ with n_j denoting the number of micro-sites at stand j.

As is typical, we take $Y_{ij} \sim Bernoulli(\pi_{ij})$, where π_{ij} is the probability of hemlock presence at site i in stand j.

Micro-site Level Model

Now, let CWD_{ij} denote the micro-site level measurement of CWD.

Since the responses are binary, we use a **logistic regression model** to relate those responses to the explanatory variables, CWD_{ij}:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} + \beta_{1j} CWD_{ij}$$

Notice that there are 15 β_0's and 15 β_1's in this model, one of each for each of the stands.

Stand Level: Model 1

We could ignore the effects of stand-level CWD and just take:

$$\begin{aligned} (\beta_{0j}, \beta_{1j})' &\sim N_2((\mu_0, \mu_1)', \Omega) \text{ for all } j \\ \mu_0, \mu_1 &\sim N(0, 1000) \\ \Omega &\sim \mathit{Wishart}(R, 2) \end{aligned}$$

This is essentially a random coefficients model—the intercepts (β_{0j}) and slopes (β_{1j}) are random and different at each stand.

Stand Level: Model 2

Alternatively, we'll let the slope coefficients, β_{1j} depend on the measure of stand level CWD, denoted $CWDAVE_j$:

$$
\begin{aligned}
(\beta_{0j}, \beta_{1j})' &\sim N_2((\mu_0, \mu_{1j})', \Omega) \text{ for all } j \\
\mu_{1j} &= \alpha_0 + \alpha_1 CWDAVE_j \text{ for } j = 1, \ldots, 15 \\
\mu_0, \alpha_0, \alpha_1 &\sim N(0, 1000) \\
\Omega &\sim Wishart(R, 2)
\end{aligned}
$$

This is, in this model, we let the effect of CWD locally, namely β_{1j} depend on CWD at the stand level.

Some Results

- Under Model 2, the 95% posterior interval for α_1 is $(-0.21, 0.18)$, suggesting that the association between CWD and hemlock presence locally does not depend on the average amount of CWD available in the stand.
- From Model 1, we estimate that for every $0.1 m^2 CWD/m^2$ area increase in the amount of CWD at the micro-site, the odds of hemlock sapling presence to absence increases 2.45-fold (95% posterior interval: $(1.47, 3.96)$).
- In Model 1, there is substantial variability in the β_{1j}'s suggesting that other (unmeasured) stand-level factors may help to explain these differences.

Some Comments

1. Questions are clearly specified and answered by different parameters in the model.

2. Data that are collected at different **scales** are explicitly incorporated into different levels of the hierarchical model.

3. Results are easy to interpret—the 95% posterior interval for the odds of presence versus absence *is* a probability interval—the probability that the true odds is in the interval is 95% under this model.

Some Questions for You

1. How might we add a component or components to this model to account for the possibility that there is spatial dependence between sites within stands that's not entirely accounted for by the CWD covariate and the hierarchical model structure?

2. How might we add a component or components to this model to account for the possibility that there is spatial dependence **across** stands?

3. Can you do some model evaluation and/or exploratory analysis to see if spatial dependence is even an issue?