Hierarchical models

What are Bayesian Hierarchical Models?

Bayesian hierarchical models are also known as multilevel models

- Particularly useful in cases where the data are nested within groups
- Examples: patients in different hospitals, schools within a city, geographic data, etc.
- Hierarchical models borrow information across groups

Example: Cardiac surgery in babies

- Source: BUGS examples Volume 1
- We have data for mortality rates in 12 hospitals performing cardiac surgery in babies
- For each hospital, we have n_i performed surgeries and y_i deaths, $i \in \{1, 2, \ldots, 12\}$

How do we model this?

Given what you know... Two options

1. **No pooling:** Each hospital has its own independent probability of death θ_i

$$y_i \mid \theta_i \overset{\text{iid}}{\sim} \text{Binomial}(n_i, \theta_i), \quad \theta_i \overset{\text{iid}}{\sim} \text{Beta}(1,1)$$

2. **Complete pooling:** All hospitals share the same θ

$$y_i \mid \theta \overset{\text{iid}}{\sim} \text{Binomial}(n_i, \theta), \quad \theta \sim \text{Beta}(1,1)$$

Hierarchical models

- There is a third way: hierarchical models
- Hierarchical models allow for **partial pooling**
- How do we do that?

Building a hierarchical model

We start with a model with hospital-specific θ_i:

$$y_i \mid \theta_i \overset{\text{iid}}{\sim} \text{Binomial}(n_i, \theta_i)$$

- There is a population of θ_i: we took a sample of 12 schools
- Conceptually, you can think that the θ_i are drawn from a distribution

A distribution for θ_i

Since the θ_i are probabilities,

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) \mid \mu, \tau \overset{\text{iid}}{\sim} N(\mu, 1/\tau)$$

μ and τ are unknown (and we're Bayesians), so we need distributions for them

Priors for μ and τ

The distributions on μ and τ characterize the heterogeneity between hospitals

We put a.vague normal prior on μ and a vague prior on τ, expressing prior ignorance about the problem

Complete model

$$
\begin{aligned}
y_i \mid \theta_i &\overset{\text{iid}}{\sim} \text{Binomial}(n_i, \theta_i) \\
\log\left(\frac{\theta_i}{1-\theta_i}\right) \mid \mu, \tau &\overset{\text{iid}}{\sim} N(\mu, 1/\tau) \\
\mu &\sim N(\mu_0, \sigma_0^2) \\
\tau &\sim \text{Gamma}(a_0, b_0)
\end{aligned}
$$

- It allows for between hospital heterogeneity, but it doesn't assume that the θ_i are independent (as in no pooling)

Advantages of using a hierarchical model

- The model allows us to make inferences and predictions for the hospitals we have surveyed
- It also allows to make inferences and predictions for a future hospital
- We'll see prediction next lecture

Example: Pumps

- Source: BUGS examples Volume 1
- Consider 10 power plant pumps. The number of failures at the i-th pump is assumed to be

$$x_i \mid \theta_i \sim \text{Poisson}(\theta_i t_i), \ \ i \in \{1, 2, \ldots, 10\}$$

  where θ_i is the failure rate for pump i and t_i is the length of operation time of the pump (in hours)
- Are there differences between pumps?

How do we model this?

- We can model this hierarchically: we can imagine there is a population of pumps

$$\begin{aligned} x_i \mid \theta_i &\sim \text{Poisson}(\theta_i t_i), \quad i \in \{1, 2, \ldots, 10\} \\ \theta_i \mid \alpha, \beta &\sim \text{Gamma}(\alpha, \beta) \\ \alpha &\sim \text{Exp}(1) \\ \beta &\sim \text{Gamma}(0.1, 1) \end{aligned}$$

Example: Homework time

- Source: Peter Hoff, A first course in Bayesian Analysis
- Weekly time spent doing homework (in hours) by students from 8 schools:

$$y_{ij} \mid \theta_i, \tau \sim \mathrm{Normal}(\theta_i, 1/\tau)$$

i is the school index and j indexes students within school

Hierarchical model

$$\begin{aligned}
y_{ij} \mid \theta_i, \tau &\sim \mathrm{Normal}(\theta_i, 1/\tau) \\
\theta_i \mid \mu, \phi &\sim \mathrm{Normal}(\mu, 1/\phi) \\
\mu &\sim \mathrm{Normal}(0, 100) \\
\tau^{-1/2} &\sim \mathrm{Unif}(0, 1000) \\
\phi^{-1/2} &\sim \mathrm{Unif}(0, 1000)
\end{aligned}$$

$\tau^{-1/2}$ and $\phi^{-1/2}$ are standard deviations

Exercise: 8 schools

- Source: Bayesian Data Analysis, Gelman et al.
- Eight schools participated in a program to see if tutoring improves SAT scores. For each school, they consider a group of students that was tutored and a group of student that wasn't tutored.
- The data below shows the difference of mean SAT scores for "tutored" vs "untutored" students. It also includes standard errors for the average differences:

Model

The statistical model for the j-th school is

$$y_j \overset{\text{iid}}{\sim} \text{Normal}(\theta_j, \sigma^2)$$

where we assume that $\sigma^2 = \text{se}^2$ is known. This is simplification isn't very important if we're interested in comparing schools (i.e., the θ_j)

Exercise

- Compare complete pooling, no pooling, and hierarchical models for the data. For the hierarchical model, you can use similar distributions to the ones proposed for the homework data
- With the hierarchical model, do a ranking of schools (like we did for hospitals)