The Hierarchical Beta-Binomial

A simple model for multiple-choice testing: examinees \( i = 1, \ldots, N \), each getting \( y_i \) of \( n_i \) questions right.

1. Level 1: \( y_i | \theta_i \sim Bin(\theta_i, n_i) \);
2. Level 2: \( \theta_i | \alpha, \beta \sim Beta(\alpha, \beta) \)

We are interested in inference about \( \theta_i \), the probability that examinee \( i \) gets a question right — a measure of “proficiency” for examinee \( i \).

Full model is

\[
p(y, \theta | \alpha, \beta) = \prod_{i=1}^{N} \left( \binom{n_i}{y_i} \theta_i^{y_i} (1 - \theta_i)^{n_i-y_i} \right) \prod_{i=1}^{N} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1} (1 - \theta_i)^{\beta-1}
\]

Using our “key observation” for each \( \theta_i \), we see that

\[
p(\theta_i | y_i, \alpha, \beta) \propto \theta_i^{y_i+\gamma_i-1} (1 - \theta_i)^{\beta+n_i-y_i-1} \equiv \theta_i | y_i \sim Beta(\alpha + y_i, \beta + n_i - y_i)
\]

If we fix \( \alpha, \beta \), we know how to analyze \( Beta(\alpha + y_i, \beta + n_i - y_i) \)!
An MCMC solution

From the full model

\[ p(y, \theta, \alpha, \beta) = \prod_{i=1}^{N} \left( \frac{n_i!}{y_i!} \right) \theta_i^{y_i} (1 - \theta_i)^{n_i-y_i} \prod_{i=1}^{N} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1} (1 - \theta_i)^{\beta-1} p(\alpha, \beta) \]

the “key observation” gives the complete conditionals

\[ p(\theta|\text{rest}) = \text{Beta}(\theta|\alpha + y_i, \beta + n_i - y_i) \]

\[ p(\alpha|\text{rest}) \propto \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \right] \prod_{i=1}^{N} \theta_i^{\alpha} p(\alpha, \beta) \]

\[ p(\beta|\text{rest}) \propto \left[ \frac{\Gamma(\alpha + \beta)}{\Gamma(\beta)} \right] \prod_{i=1}^{N} (1 - \theta_i)^{\beta} p(\alpha, \beta) \]

This suggests:

- Gibbs step for \( \theta_i \)’s: sample \( \theta_i \sim \text{Beta}(\cdots) \) directly
- Metropolis steps for \( \alpha \) and \( \beta \) using Normal proposal draws (“random walk M-H”). Normal variances are “tuning parameters”.

In this model, the “population” of \( \theta_i \)’s has a \( \text{Beta}(\alpha, \beta) \) distribution.

What if we want to estimate \( \alpha, \beta \) (estimate the shape of the population / latent distribution)? I.e. what is the distribution of “proficiency” among students who took this test?

Add a third modeling assumption:

3. Level 3: \( \alpha, \beta \sim p(\alpha, \beta) \)

Now the full model is

\[ p(y, \theta, \alpha, \beta) = \prod_{i=1}^{N} \left( \frac{n_i!}{y_i!} \right) \theta_i^{y_i} (1 - \theta_i)^{n_i-y_i} \prod_{i=1}^{N} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1} (1 - \theta_i)^{\beta-1} p(\alpha, \beta) \]

Again using the “key observation” for \( \alpha, \beta \), we see

\[ p(\theta|y_i, \alpha, \beta) = \text{Beta}(\theta|\alpha + y_i, \beta + n_i - y_i) \]

\[ p(\alpha, \beta|y) \propto p(\alpha, \beta) \prod_{i=1}^{N} \frac{\Gamma(\alpha + \beta) \Gamma(\alpha + y_i) \Gamma(\beta + n_i - y_i)}{\Gamma(\alpha)\Gamma(\beta) \Gamma(\alpha + \beta + n_i)} \]

Gelman et al. (pp. 128ff.) suggest computing \( p(\alpha, \beta|y) \)—or actually \( p(\log(\alpha/\beta), \log(\alpha + \beta)|y) \)—on a grid, using trial and error to place the grid over the “interesting” part of the density.
Example: Rat Tumors

The data are from Gelman pp. 119 ff. In this case the binomial experiment is to observe the number $y_i$ of a group of $n_i$ rats that develop tumors when exposed to some risk factor. Each group of $n_i$ rats is from a different experiment and so the $\theta_i = P[\text{tumor in group } i]$ will vary from group to group.

We try the hierarchical beta-binomial model as above.

see R code for this lecture