36-724 Spring 2006: Metropolis-Hastings Example

Brian Junker

February 22, 2006

- The Hierarchical Beta-Binomial
- An MCMC solution
- Example: Rat Tumors

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 θ_i, the probability that examinee i gets a question right — a measure of "proficiency" for examinee i.

Full model is

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

Using our "key observation" for each θ_i, we see that

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

If we fix α, β, we know how to analyze $Beta(\alpha + y_i, \beta + n_i - y_i)$!

In this model, the "population" of θ_i's has a $Beta(\alpha, \beta)$ distribution.

What if we want to estimate α, β (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

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

Again using the "key observation" for α, β, we see

$$\begin{aligned} p(\theta_i | y_i, \alpha, \beta) &= Beta(\theta_i | \alpha + y_i, \beta + n_i - y_i) \\ p(\alpha, \beta | y) &\propto p(\alpha, \beta) \prod_{i=1}^{N} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + y_i)\Gamma(\beta + n_i - y_i)}{\Gamma(\alpha + \beta + n_i)} \end{aligned}$$

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.

An MCMC solution

From the full model

$$p(y,\theta,\alpha,\beta) \;=\; \prod_{i=1}^{N} \binom{n_i}{y_i} \theta_i^{y_i}(1-\theta_i)^{n_i-y_i} \prod_{i=1}^{N} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_i^{\alpha-1}(1-\theta_i)^{\beta-1} p(\alpha,\beta)$$

the “key observation” gives the complete conditionals

$$\begin{aligned}
p(\theta_i|\text{ rest }) &= Beta(\theta_i|\alpha+y_i,\beta+n_i-y_i) \\
p(\alpha|\text{ rest }) &\propto \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)}\right]^N \prod_{i=1}^{N} \theta_i^{\alpha} p(\alpha,\beta) \\
p(\beta|\text{ rest }) &\propto \left[\frac{\Gamma(\alpha+\beta)}{\Gamma(\beta)}\right]^N \prod_{i=1}^{N} (1-\theta_i)^{\beta} p(\alpha,\beta)
\end{aligned}$$

This suggests:

- Gibbs step for θ_i’s: sample $\theta_i \sim Beta(\cdots)$ directly
- Metropolis steps for α and β using Normal proposal draws (“random walk M-H”). Normal variances are “tuning parameters”.

Choosing $p(\alpha,\beta)$

We note that for $\theta \sim B(\alpha,\beta)$, the "prior mean" is $\mu = E[\theta] = \alpha/\beta$ and $\tau^2 = \mathrm{Var}(\theta) = \mu(1-\mu)/(\alpha+\beta+1)$. It is generally easier to think about priors for μ and τ^2 than it is

Imitating what we did in the hierarchical normal case we might take

$$p(\mu) \propto 1\,, \quad \Rightarrow \quad p(\alpha/\beta) \propto 1$$
$$p(\tau) \equiv p(\tau|\mu) \propto 1\,, \quad \Rightarrow \quad p(1/\sqrt{\alpha+\beta}) \propto 1$$

Transforming back to (α,β) we get

$$p(\alpha,\beta) \;\propto\; (\alpha+\beta)^{-5/2}$$

One could imagine other schemes also, e.g.:

- Taking $(\alpha+\beta)^{-k}$ for higher values of k.
- Inventing some reasonable proper priors for α and β, e.g.

  $$p(\alpha) = Exp(\alpha|k) \qquad p(\beta) = Exp(\beta|\ell)$$

  (and further levels for k and $\ell\ldots$?)

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`