Exercises 5: Hierarchy and data augmentation

Hierarchical models and shrinkage

The data set in "mathtest.csv" shows the scores on a standardized math test from a sample of 10th-grade students at 100 different U.S. urban high schools, all having enrollment of at least 400 10th-grade students. (A lot of educational research involves "survey tests" of this sort, with tests administered to all students being the rare exception.)

Let θ_i be the underlying mean test score for school i, and let y_{ij} be the score for the jth student in school i. Starting with the "mathtest.R" script, you'll notice that the extreme school-level averages $\bar{y}_i$ (both high and low) tend to be at schools where fewer students were sampled.

1. Explain briefly why this would be.

2. Fit a normal hierarchical model to these data via Gibbs sampling:

$$
\begin{aligned}
y_{ij} &\sim \mathrm{N}(\theta_i, \sigma^2) \\
\theta_i &\sim \mathrm{N}(\mu, \tau^2)
\end{aligned}
$$

   Decide upon sensible priors for the unknown model parameters (μ, σ^2, τ^2).

3. Suppose you use the posterior mean $\hat{\theta}_i$ from the above model to estimate each school-level mean θ_i. Define the shrinkage coefficient κ_i as

$$
\kappa_i = \frac{\bar{y}_i - \hat{\theta}_i}{\bar{y}_i},
$$

   which tells you how much the posterior mean shrinks the observed sample mean. Plot this shrinkage coefficient for each school as a function of that school's sample size, and comment.

Data augmentation

Read the following paper:

> "Bayesian Analysis of Binary and Polychotomous Response Data."
> James H. Albert and Siddhartha Chib. *Journal of the American Statistical Association*, Vol. 88, No. 422 (Jun., 1993), pp. 669-679

The surefire way to get this paper is via access to JStor through the UT Library website. Let me know if this is an issue for you.

The paper describes a Bayesian treatment of probit regression (similar to logistic regression) using the trick of *data augmentation*—that is, introducing "latent variables" that turn a hard problem into a much easier one. Briefly summarize your understanding of the key trick proposed by this paper. Then see you if you can apply the trick in the following context, which is more complex than ordinary probit regression.

In "polls.csv" you will find the results of several political polls from the 1988 U.S. presidential election. The outcome of interest is whether someone plans to vote for George Bush (senior, not junior). There are several potentially relevant demographic predictors here, including the respondent's state of residence. The goal is to understand how these relate to the probability that someone will support Bush in the election. You can imagine this information would help a great deal in poll re-weighting and aggregation (ala Nate Silver).

Use Gibbs sampling, together with the Albert and Chib trick, to fit a hierarchical probit model of the following form:

$$\Pr(y_{ij} = 1) = \Phi(z_{ij})$$
$$z_{ij} = \mu_i + x_{ij}^T \beta.$$

Here y_{ij} is the response (Bush=1, other=0) for respondent j in state i; $\Phi(\cdot)$ is the probit link function, i.e. the CDF of the standard normal distribution; μ_i is a state-level intercept term; x_{ij} is a vector of respondent-level demographic predictors; and β is a vector of state-invariant regression coefficients.

Note: there are severe imbalances among the states in terms of numbers of survey respondents! Following the last problem, the key is to impose a hierarchical prior on the state-level intercepts.