Exercises 4: Hierarchical Models

Hierarchical models: data-analysis problems

Math tests

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 this normal hierarchical model to these data via Gibbs sampling:

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

   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 (in absolute value) for each school as a function of that school's sample size, and comment.

Price elasticity of demand

The data in "cheese.csv" are about sales volume, price, and advertisting display activity for packages of Borden sliced "cheese." The data are taken from Rossi, Allenby, and McCulloch's textbook on *Bayesian Statistics and Marketing.* For each of 88 stores (store) in different US cities, we have repeated observations of the weekly sales volume (vol, in terms of packages sold), unit price (price), and whether the product was advertised with an in-store display during that week (disp = 1 for display).

Your goal is to estimate, on a store-by-store basis, the effect of display ads on the demand curve for cheese. A standard form of a demand curve in economics is of the form $Q = \alpha P^{\beta}$, where Q is quantity demanded (i.e. sales volume), P is price, and α and β are parameters to be estimated. You'll notice that this is linear on a log-log scale,

$$\log P = \log \alpha + \beta \log Q$$

which you should feel free to assume here. Economists would refer to β as the price elasticity of demand (PED). Notice that on a log-log scale, the errors enter multiplicatively.

There are several things for you to consider in analyzing this data set.

1. The demand curve might shift (different α) and also change shape (different β) depending on whether there is a display ad or not in the store.
2. Different stores will have very different typical volumes, and your model should account for this.
3. Do different stores have different PEDs? If so, do you really want to estimate a separate, unrelated β for each store?
4. If there is an effect on the demand curve due to showing a display ad, does this effect differ store by store, or does it look relatively stable across stores?
5. Once you build the best model you can using the log-log specification, do see you any evidence of major model mis-fit?

Propose an appropriate hierarchical model that allows you to address these issues, and use Gibbs sampling to fit your model.

A hierarchical probit model via 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:

$$\begin{aligned} \Pr(y_{ij} = 1) &= \Phi(z_{ij}) \\ z_{ij} &= \mu_i + x_{ij}^T \beta_i \, . \end{aligned}$$

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_i is a vector of regression coefficients for state i.

Notes:

1. 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 parameters.

2. The data-augmentation trick from the Albert and Chib paper above is explained in many standard references on Bayesian analysis. If you want to get a quick introduction to the idea, you can consult one of these. A good presentation is in Section 8.1.1 of "Bayesian analysis for the social sciences" by Simon Jackman, available as an ebook through lib.utexas.edu.

3. You are welcome to use the logit model instead of the probit model. If you do this, you'll need to read the following paper, rather than Albert and Chib: Polson, N.G., Scott, J.G. and Windle, J. (2013). Bayesian inference for logistic models using Polya-Gamma latent variables. J. Amer. Statist. Assoc. 108 1339–1349.