Bayesian Case Studies
Practical 5

Hierarchical models. Posterior predictive checks. Adaptive MCMC.

Robin Ryder

Aim: Sampling from the posterior in a hierarchical model using an adaptive method. Computing posterior predictive probabilities.
Reference: Bayesian Data Analysis (Gelman, Carlin, Stern, Dunson, Vehtari & Rubin), chapter 6.

We wish to estimate the influence of a coaching program on the SAT scores of Boston high school students. In each of 8 schools, students are separated into two groups: some receive coaching, some do not. We record the SAT scores (which take values between 200 and 800) for each student. Education experts have obtained an estimate of the effect of the coaching program in each school, using a regression technique which we shall take as granted. In school i, the effect of the coaching program is denoted by θ_i and we observe a single realization $y_i \sim \mathcal{N}(\theta_i, \sigma_i^2)$ with σ_i known. We wish to infer on the (θ_i).
The data (from Gelman et al. 2004 and Rubin 1981) are summarized in this table:

School	y_i	σ_i
A	28	15
B	8	10
C	-3	16
D	7	11
E	-1	9
F	1	11
G	18	10
H	12	18

1. Think about these data: which aspects are striking and will need to be particularly kept in mind when checking the model fit?

2. Suppose the 8 experiments are independent. What would your interpretation be? Does this assumption seem reasonable?

3. Fit a model where all the θ_i are equal: $\exists \mu, \forall i, \theta_i = \mu$. Give the posterior mean and variance on μ using a non-informative constant prior. Interpret your results. Is this model reasonable given the data?

4. We propose a hierarchical model: $\theta_i \sim \mathcal{N}(\mu, \tau^2)$ with μ and τ unknown. Check that the two previous approaches can be viewed as special cases of this model.

5. Using a non-informative constant prior for μ and τ, write a Metropolis-Hastings algorithm to sample from the posterior distribution. Aim for an effective sample size around 200.

6. Use your posterior sample to estimate the treatment effect in each school. Give approximate distributions for $\max(\theta_i)$ and estimate the posterior probability of the event $\{\theta_A > \theta_C\}$.

7. Check the influence of the prior distribution.

8. Simulate 200 replicates of the data and perform the relevant posterior predictive checks.

9. Replace the distribution of the (θ_i) by a heavier-tailed distribution and perform a sensitivity analysis.

10. Modify the code for the Metropolis-Hastings algorithm to make it adaptive and evaluate the gain in efficiency.

11. Write a Gibbs' sampler to solve this problem, and compare the output to the Metropolis-Hastings approach.

12. Implement this model in STAN and use the `rstan` package to infer the parameters and check convergence. Compare with the previous approaches.

Adaptive MH (question 10) and STAN (question 12) will not be on the final examination.