STATS 551 Homework 3

MCMC & STAN

Due date: 6:00 pm (EST) Mar. 14, 2018

Estimating the risk of tumor in a group of rats. In the evaluation of drugs for possible clinical application, studies are routinely performed on rodents. For a particular study drawn from the statistical literature, suppose the immediate aim is to estimate θ, the probability of tumor in a population of female laboratory rats of type F344 that receive a zero dose of the drug (a control group). The data show that 4 out of 14 rats developed endometrial stromal polyps (a kind of tumor). It is natural to assume a binomial model for the number of tumors, given θ_{68}. For convenience, we select a prior distribution for θ_{68} from the conjugate family, $\theta \sim Beta(\alpha, \beta)$.

Typically, the mean and standard deviation of underlying tumor risks are not available. Rather, historical data are available on previous experiments on similar groups of rats. In the rat tumor example, the historical data were in fact a set of observations of tumor incidence in 67 groups of rats. In the j-th historical experiment, let the number of rats with tumors be y_j and the total number of rats be n_j. We model the y_j's as independent binomial data, given sample sizes n_j and study-specific means θ_j. Assuming that the beta prior distribution with parameters α, β is a good description of the population distribution of the θ_j's in the historical experiments, we can display the hierarchical model schematically as in Figure 1, with θ_{68} and y_{68} corresponding to the current experiment.

1. Write down the joint posterior distribution of $(\theta_1, \ldots, \theta_{68})$ given (α, β). Visualize the posterior distributions of $(\theta_1, \ldots, \theta_{68})$ when $(\alpha, \beta) = (1.4, 8.6)$.
2. What is the expectation of each y_j given α, β? What is the expectation of each y_j^2 given α, β? Can you give a possible value of (α, β) based on summaries of these moments? Plug in your guess for (α, β), calculate the posterior distribution for $(\theta_1, \ldots, \theta_{68})$ and compare with the results in 1.

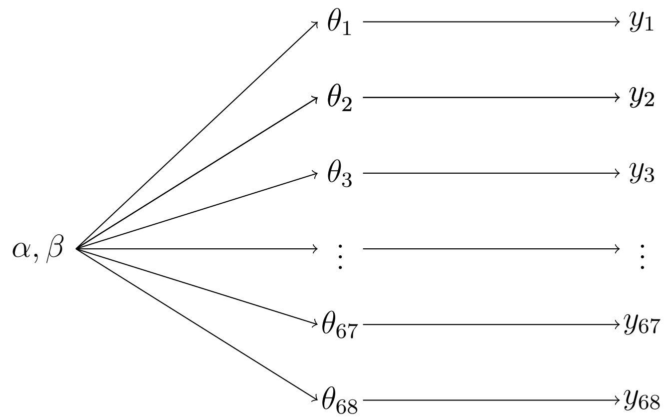

Figure 1: Structure of hierarchical model.

3. Show that a uniform prior on $(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2})$ yields a prior on the original scale $p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}$ and on natural transformed scale

$$p\left(\log \frac{\alpha}{\beta}, \log(\alpha+\beta)\right) \propto \alpha\beta(\alpha+\beta)^{-5/2}.$$

4. Write down the joint posterior distribution of $(\theta_1, \dots, \theta_{68}, \alpha, \beta)$. Write down the marginal posterior distribution of (α, β) on the original scale $p(\alpha, \beta|y)$ and on the log transformed scale $p\left(\log \frac{\alpha}{\beta}, \log(\alpha+\beta)|y\right)$.

5. Work with either $p(\alpha, \beta|y)$ or $p\left(\log \frac{\alpha}{\beta}, \log(\alpha+\beta)|y\right)$, design a Metropolis-Hastings algorithm to obtain posterior samples for α, β. Visualize your results.

6. Given the posterior samples of α, β, can you obtain posterior samples for $(\theta_1, \dots, \theta_{68})$? Visualize your results.

7. Work with either $p(\alpha, \beta; \theta_1, \dots, \theta_{68}|y)$ or $p\left(\log \frac{\alpha}{\beta}, \log(\alpha+\beta); \theta_1, \dots, \theta_{68}|y\right)$, design a Metropolis-Hastings (maybe combined with Gibbs sampling) algorithm to obtain posterior samples for $(\alpha, \beta; \theta_1, \dots, \theta_{68})$. Visualize your results.

8. Implement 7 in STAN.

9. Calculate $\mathbb{E}(\alpha|y)$ and $Pr(\frac{\alpha}{\alpha+\beta} < 0.2|y)$ based on posterior samples in 7.

10. Calculate $\mathbb{E}(\alpha|y)$ and $Pr(\frac{\alpha}{\alpha+\beta} < 0.2|y)$ based on posterior samples in 8.

11. (Optional) If you assign flat priors on (α, β), is the posterior proper? If your answer is yes, test out the results in STAN; otherwise, show that the posterior is improper theoretically or numerically.

The data is as follows, you can directly copy-and-paste to your R script.

```
y <- c(0,0,0,0,0,0,0,0,0,0,0,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,
      2,1,5,2,5,3,2,7,7,3,3,2,9,10,4,4,4,4,4,4,4,10,4,4,4,5,
      11,12, 5,5,6,5,6,6,6,6,16,15,15,9,4)
n <- c(20,20,20,20,20,19,19,19,19,18,17,20,20,20,20,19,19,18,18,25,24,
       23,20,20,20,20,20,20,10,49,19,46,27,17,49,47,20,20,13,48,50,20,
       20,20,20,20,20,20,48,19,19,19,22,46,49,20,20,23,19,22,20,20,20,
       52,46,47,24,14)
```

Remark. This example is modified from BDA, chapter 5, section 5.1. Please work out your own solutions before referring to resources on the Internet.

Guideline for Submission: Submit R markdown (or jupyter notebook) with annotated code followed by results. Discussions about the results should follow the results.

Optional Reading. Read one of the following papers and post your summary and thoughts on Canvas. Bonus points up to 5 will be rewarded.

1. Use of Bayesian statistics in drug development: Advantages and challenges, Sandeep K Gupta, Int J Appl Basic Med Res. 2012 Jan-Jun; 2(1): 3?6.

2. Sequential imputation and Bayesian missing data problems, Kong, A., Liu, J. S., and Wong, W. H. (1994), Journal of the American Statistical Association 89, 278-288.

3. Cowles, Mary Kathryn, and Bradley P. Carlin. “Markov chain Monte Carlo convergence diagnostics: a comparative review.” Journal of the American Statistical Association 91.434 (1996): 883-904.

4. Neal, Radford M. “Probabilistic inference using Markov chain Monte Carlo methods.” (1993).

5. Kass, Robert E., et al. “Markov chain Monte Carlo in practice: a roundtable discussion.” The American Statistician 52.2 (1998): 93-100.

For students interested in theoretical foundations:

1. Betancourt, Michael. "The Convergence of Markov chain Monte Carlo Methods: From the Metropolis method to Hamiltonian Monte Carlo." arXiv preprint arXiv:1706.01520 (2017).

2. Salimans, Tim, Diederik Kingma, and Max Welling. "Markov chain monte carlo and variational inference: Bridging the gap." Proceedings of the 32nd International Conference on Machine Learning (ICML-15). 2015.

3. Rosenthal, Jeffrey S. "Minorization conditions and convergence rates for Markov chain Monte Carlo." Journal of the American Statistical Association 90.430 (1995): 558-566.

4. Rosenthal, Jeffrey S. "Asymptotic variance and convergence rates of nearly-periodic Markov chain Monte Carlo algorithms." Journal of the American Statistical Association 98.461 (2003): 169-177.

5. Walker, Stephen, and Nils Lid Hjort. "On Bayesian consistency." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63.4 (2001): 811-821.

6. Walker, Stephen G. "Modern Bayesian asymptotics." Statistical Science (2004): 111-117.

7. Walker, Stephen. "New approaches to Bayesian consistency." Annals of Statistics (2004): 2028-2043.

8. De Blasi, Pierpaolo, and Stephen G. Walker. "Bayesian asymptotics with misspecified models." Statistica Sinica (2013): 169-187.

9. Gelfand, Alan E., and Dipak K. Dey. "Bayesian model choice: asymptotics and exact calculations." Journal of the Royal Statistical Society. Series B (Methodological) (1994): 501-514.

10. Carlin, Bradley P., and Siddhartha Chib. "Bayesian model choice via Markov chain Monte Carlo methods." Journal of the Royal Statistical Society. Series B (Methodological) (1995): 473-484.