**STAT 431 — Applied Bayesian Analysis — Spring 2019**

# Homework 5

Please submit your assignment *on paper*. Make sure your answers are completely justified and clear enough to read! Any computer code and output should be included.

1. Suppose

$$\left.\begin{array}{rcl} Z & \sim & \mathrm{N}(0,1) \\ \varepsilon_1 & \sim & \mathrm{N}(0,\sigma_1^2) \\ \varepsilon_2 & \sim & \mathrm{N}(0,\sigma_2^2) \end{array}\right\} \text{ independent}$$

and let

$$Y_1 = Z + \varepsilon_1 \qquad Y_2 = Z + \varepsilon_2$$

(a) [2 pts] Under what conditions (if any) are $Y_1$ and $Y_2$ *exchangeable*? Justify your answer.

(b) [2 pts] Determine the covariance of $Y_1$ and $Y_2$. Under what conditions (if any) are they (marginally) independent?

2. Twelve separate case-control studies were run to investigate the potential link between presence of a certain genetic trait (the $\mathrm{Pl}^{\mathrm{A2}}$ polymorphism of the glycoprotein IIIa subunit of the fibrinogen receptor) and risk of heart attack.[1] For the $i^{\text{th}}$ study, an estimated log-odds ratio, $\hat{\psi}_i$, and its (estimated) standard error, $\sigma_i$, were computed:

| $i$ | $\hat{\psi}_i$ | $\sigma_i$ | $i$ | $\hat{\psi}_i$ | $\sigma_i$ | $i$ | $\hat{\psi}_i$ | $\sigma_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1.055 | 0.373 | 5 | 1.068 | 0.471 | 9 | 0.507 | 0.186 |
| 2 | -0.097 | 0.116 | 6 | -0.025 | 0.120 | 10 | 0.000 | 0.328 |
| 3 | 0.626 | 0.229 | 7 | -0.117 | 0.220 | 11 | 0.385 | 0.206 |
| 4 | 0.017 | 0.117 | 8 | -0.381 | 0.239 | 12 | 0.405 | 0.254 |

Consider this Bayesian hierarchical model:

$$\begin{aligned} \hat{\psi}_i \mid \psi_i &\sim \text{indep. } \mathrm{N}(\psi_i, \sigma_i^2) \qquad i = 1, ..., 12 \\ \psi_i \mid \psi_0, \sigma_0^2 &\sim \text{indep. } \mathrm{N}(\psi_0, \sigma_0^2) \qquad i = 1, ..., 12 \\ \psi_0 &\sim \mathrm{N}(0, 1000) \\ \tau_0^2 = 1/\sigma_0^2 &\sim \text{gamma}(0.001, 0.001) \end{aligned}$$

with $\psi_0$ and $\tau_0^2$ independent, and the values $\sigma_i^2$, $i = 1, \ldots, 12$, regarded as fixed and known.

(a) [2 pts] Draw a directed acyclic graph (DAG) appropriate for this model. (Use the notation introduced in lecture, including "plates".)

(b) [3 pts] Using the template `prob2template.bug` provided on the Compass course web site, form a JAGS model statement (corresponding to your graph). [ Remember: JAGS "`dnorm`" uses precisions, not variances! ]

[1] From Burr, et al. (2003), *Statistics in Medicine*, **22**: 1741–1760.

(c) [3 pts] Perform a preliminary run of your model using `rjags`, to check for convergence of your sampler. Use three chains, separately initialized. Include the lists of the initial values you used. Show the plots you used to monitor convergence. Explicitly determine how many iterations to burn (i.e. exclude from inference), and justify the number you chose.

(d) [3 pts] Perform at least 100000 iterations (after burn-in), and produce a summary of your inference results, for both $\psi_0$ and $\sigma_0^2$. Be sure to include estimates of the posterior expected value and *standard deviation*, the Monte Carlo error, and a 95% posterior credible interval. Also include graphical estimates of the densities.

3. GRADUATE SECTION ONLY

Consider the situation and data of Problem 2. Suppose a new case-control study is to be performed, and assume that its log-odds standard error ($\sigma_{\text{new}}$) is anticipated to be 0.4.

This new study will produce a new *estimated* log-odds ratio $\hat{\psi}_{\text{new}}$, and it will also have a $\psi_{\text{new}}$, comparable to the $\psi_i$s of the previous studies. Model the $\psi_{\text{new}}$ for the new study as exchangeable with the $\psi_i$s of the previous studies.

Modify your JAGS model (*not* the data) and analysis to answer the following:

(a) [3 pts] Estimate the posterior mean and *standard deviation*, and form a 95% posterior predictive interval for the *estimated* log-odds ratio that the new study will obtain.

(b) [2 pts] Estimate the posterior predictive probability that the new *estimated* log-odds ratio will be at least twice its standard error ($2\sigma_{\text{new}}$), i.e., at least two standard errors greater than zero. (This is roughly the probability that the new study will find a statistically significant result, and in the positive direction.)

Show the modified JAGS code you used and also a summary of the approximate inference results from `rjags/coda` that you used to answer parts (a) and (b).