

Assignment 2

1. Consider the two different hyperprior formulations for the binomial hierarchical model of Lesson 3.2: Hierarchical Modeling Fundamentals. This exercise shows how different those priors are.

Note: Consult `help(distributions)` in R for the random number generators you will need. (You do not need JAGS.)

- (a) The first prior formulation was

$$\begin{aligned}\theta_j \mid \alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\ \alpha, \beta &\sim \text{iid Expon}(0.001)\end{aligned}$$

- (i) [2 pts] Independently simulate 1000 pairs (α, β) from their hyperprior, and produce a scatterplot of $\log(\beta)$ versus $\log(\alpha)$.
- (ii) [2 pts] Using the simulated pairs (α, β) , forward-simulate θ_j , and produce a histogram of the result (an approximation of its marginal prior).

- (b) The second prior formulation was

$$\begin{aligned}\theta_j \mid \alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\ \alpha &= \phi_1 / \phi_2^2 & \beta &= (1 - \phi_1) / \phi_2^2 \\ \phi_1 &\sim \text{U}(0, 1) & \phi_2 &\sim \text{U}(0, 1000)\end{aligned}$$

- (i) [2 pts] Independently simulate 1000 pairs (α, β) from their hyperprior, and produce a scatterplot of $\log(\beta)$ versus $\log(\alpha)$.
- (ii) [2 pts] Using the simulated pairs (α, β) , forward-simulate θ_j , and produce a histogram of the result (an approximation of its marginal prior).

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

j	$\hat{\psi}_j$	σ_j	j	$\hat{\psi}_j$	σ_j	j	$\hat{\psi}_j$	σ_j
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

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

Consider this Bayesian hierarchical model:

$$\begin{aligned}\hat{\psi}_j \mid \psi_j &\sim \text{indep. } N(\psi_j, \sigma_j^2) & j = 1, \dots, 12 \\ \psi_j \mid \psi_0, \sigma_0 &\sim \text{iid } N(\psi_0, \sigma_0^2) & j = 1, \dots, 12 \\ \psi_0 &\sim N(0, 1000^2) \\ \sigma_0 &\sim U(0, 1000)\end{aligned}$$

with ψ_0 and σ_0 independent, and the values σ_j , $j = 1, \dots, 12$, regarded as fixed and known.

- (a) [2 pts] Specify *improper* densities that the proper hyperpriors given above are apparently intended to approximate. (Which parameters are the hyperparameters?)
- (b) [5 pts] Draw a directed acyclic graph (DAG) appropriate for this model. (Use the notation introduced in lecture, including “plates.”) You may draw it neatly by hand or use software.
- (c) [5 pts] Using the template `asgn2template.bug` provided on the course website, form a JAGS model statement (consistent with your graph). Also, set up any R (`rjags`) statements appropriate for creating a JAGS model. Be careful to name your data variables correctly. [Also remember: JAGS “`dnorm`” uses precisions, not variances!]
- (d) [5 pts] Run at least 10,000 iterations of burn-in, then 100,000 iterations to use for inference. For both ψ_0 and σ_0^2 (*not* σ_0), produce a posterior numerical summary and also graphical estimates of the posterior densities. *Explicitly* give the approximations of the posterior expected values, posterior *standard deviations*, and 95% central posterior intervals. (Just showing R output is not enough!)
- (e) Suppose a new case-control study is to be performed, and assume that its log-odds standard error (new σ) will be 0.125. Assume the ψ for the new study is exchangeable with those for the previous studies.
 - (i) [2 pts] Re-draw your DAG, adding new nodes to represent the new $\hat{\psi}$ and new ψ .
 - (ii) [2 pts] Correspondingly modify your JAGS model to answer the following parts. *Show* the modified JAGS and R code and output that you used.
 - (iii) [3 pts] Estimate the posterior mean and posterior *standard deviation*, and form a 95% central posterior predictive interval for the *estimated* log-odds ratio that the new study will obtain. (Remember, this new estimated log-odds ratio will be the new $\hat{\psi}$, not the new ψ .)
 - (iv) [1 pt] Estimate the posterior predictive probability that the new *estimated* log-odds ratio will be at least twice its standard error, i.e., at least two standard errors (2σ) greater than zero. (This is roughly the posterior probability that the new study will find a statistically significant result, and in the positive direction.) Suggestion: Add an *indicator variable* to your JAGS model – one that equals 1 when the event occurs, and 0 otherwise. (What is its mean?)

Use at least 10,000 iterations of burn-in, and 100,000 for inference as before.

Total: 33 pts