Bayesian Data Analysis Final Exam

Pick questions with a total of 110 points out of the following 12 questions to answer. Except for question 2, which accounts for 20 points, each of the other questions accounts for 10 points. There are 10 bonus points in this exam. If you answer questions with more than 110 points, the answers for questions exceeding 110 points will not be graded.

1.
   A. (5 points) Briefly explain the 5 steps of Bayesian Data Analysis.
   B. (5 points) The concept of Bayesian analysis was actually developed in middle of 18th century. However, it was not popular until early 1990s. Provide two reasons why it was not widely accepted in the old days.

2. Human males have one X-chromosome and one Y-chromosome, whereas females have two X-chromosomes, each chromosome being inherited from one parent. Hemophilia is a disease that exhibits X-chromosome-linked recessive inheritance, meaning that a male who inherits the gene that causes the disease on the X-chromosome is affected, whereas a female carrying the gene on only one of her two X-chromosomes is not affected. The disease is generally fatal for women who inherit two such genes, and this is rare, since the frequency of occurrence of the gene is low in human populations.

   Consider a woman who has an affected brother, which implies that her mother must be a carrier of the hemophilia gene with one 'good' and one 'bad' hemophilia gene. We are also told that her father is not affected; thus the woman herself has a fifty-fifty chance of having the gene. The unknown quantity of interest, the state of the woman, has just two values: the woman is either a carrier of the gene ($\theta = 1$) or not ($\theta = 0$). Based on the information provided thus far, compute $\Pr(\theta = 1)$ and $\Pr(\theta = 0)$.

   The data used to update the prior information consist of the affection status of the woman's sons. Suppose she has two sons, neither of whom is affected. Let $y_i = 1$ or 0 denote an affected or unaffected son, respectively. The outcomes of the sons are exchangeable and, conditional on the unknown θ, are independent; we assume the sons are not identical twins.

   A. (8 points) Compute $\Pr(y_1 = 0, y_2 = 0 | \theta = 1)$ and $\Pr(y_1 = 0, y_2 = 0 | \theta = 0)$
   B. (7 points) Compute $\Pr(\theta = 1 | y_1 = 0, y_2 = 0)$.

C. (7 points) Suppose that the woman has a third son, who is also unaffected. Compute
$\Pr(\theta = 1 | (y_1 = 0, y_2 = 0, y_3 = 0)$.

3. In the Politian example, Suppose the current position probabilities are the target probabilities, i.e., w = [… , P(θ −1), P(θ), P(θ +1), …]/Z, where $Z = \sum_\theta p(\theta)$ is the normalizer for the target distribution. Consider the θ component of wT, where T represents the matrix of probability of transitioning from θ to other positions. We want to demonstrate that the θ component of wT is the same as the θ component of w, for any component θ. In other words, we want to show that:

$$\begin{aligned}\sum_r w_r T_{r\theta} &= P(\theta-1)/Z \cdot 0.5\min\left(\frac{P(\theta)}{P(\theta-1)}, 1\right) \\ &+ P(\theta)/Z \cdot \left(0.5\left[1-\min\left(\frac{P(\theta-1)}{P(\theta)}, 1\right)\right] + 0.5\left[1-\min\left(\frac{P(\theta+1)}{P(\theta)}, 1\right)\right]\right) \\ &+ P(\theta+1)/Z \cdot 0.5\min\left(\frac{P(\theta)}{P(\theta+1)}, 1\right) \qquad (7.4)\end{aligned}$$

will be $P(\theta)/Z$ for all possible four cases. Please simply check the following two cases hold:

A. (5 points) $P(\theta-1) > P(\theta) > P(\theta+1)$.

B. (5 points) $P(\theta-1) < P(\theta)$ and $P(\theta) > P(\theta+1)$.

4.

A. (5 points) Briefly explain what conjugate prior is and list two of its issues.

B. (5 points) Given a conjugate prior of beta(5, 2) for the coin bias. Assume the coin is thrown 7 times with 2 heads and 5 tails, provide the function of the posterior probability distribution.

5. (10 points) Assume that you are going to collect a large amount of MCMC samples from a Bernoulli likelihood, provide two ways to speed up JAGS, and briefly explain each of them.

6.

A. (5 points) Suppose a colleague tells you she read a research report in which “there were about 50 patients and two-thirds of them survived at least one year after surgery.” Suppose you want to use this information in an informal prior for subsequent data. (You’ll get the exact numbers later.) What are the corresponding *a* and *b* shape parameters?

B. (5 points) Suppose you are given a six-sided dice, and you want to find out whether it's biased by assessing the probability of rolling a 1-dot outcome. You know the die just came out of a brand new package from a reputable manufacturer, so you believe that the most probable probability of a 1-dot outcome is 1/6. Suppose it would take 50 rolls to begin to sway you from your prior belief that the die is fair. What are the corresponding *a* and *b* shape parameters?

7.

A. (5 points) Figure 1 shows the batting average of two particular players, one 1st base and the other a pitcher. They both have 1 hit out of 5 tries, but their batting averages are very different. Explain briefly what reason may cause these different results?

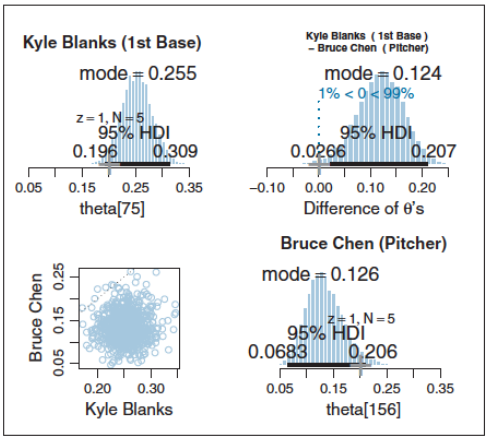

Figure 1. Batting average between Kyle Blanks and Bruce Chen in MLB.

B. (5 points) Assume that both Kyle and Bruce have a greater number of at bat opportunities (i.e., N becomes greater than 300) and their average battings remain close to 0.20, will that lead to both players having the close posterior distributions? Explain your answer.

8. Figure 2 shows the combinations of two nominal variables.

A. (5 points) Do the two nominal variables (i.e., x_1 and x_2) have additive effects or non-additive effects on the predicted variable (i.e., y)? Explain why.

B. (5 points) Based on Figure 2, give the mathematics Generalized Linear Model representation (i.e., formula) of the relationship between the two nominal variables and the predicted variable.

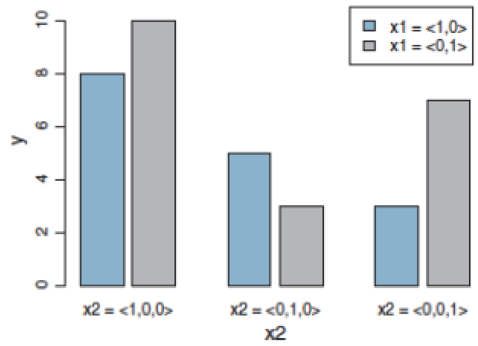

Figure 2. Combinations of two nominal variables.

9. In GLM, we normally represent the predicted variable y as a sample drawn from a probability density function pdf. For instance, $y \sim \text{pdf}(\mu, [\text{scale,shape,etc.}])$, where $\mu = f(\text{lin}(x))$.
   A. (5 points) Explain why do not we simply let $y = f(\text{lin}(x))$.
   B. (5 points) If y is of type "Count" (i.e., non-negative integer like 0, 1, 2, …), which distribution will be a typical pdf probability density function?
10.
   A. (5 points) Brief explain ROPE.
   B. (5 points) What is Meehl's Paradox? What is the connection between ROPE and Meehl's Paradox?
11.
   A. (5 points) Briefly explain HDI, CI, and ETI.
   B. (5 points) Which one out of HDI, CI, and ETI is a better analytical tool for data science? Briefly explain why?
12. (10 points) Now you have taken this course for one full semester, please provide as many points as possible (at least two to earn full credit) that you think this course can be further improved when it is offered again in the future.
   Hint: I will NOT deduct point for any input you give to me, so do not be afraid to let me know what you think.