FINAL EXAM

ISyE6420
Fall 2021

Released December 9, 12:00 am – due December 12, 11:59 pm. This exam is not proctored and not time limited except the due date. Late submissions will not be accepted.

Use of all available electronic and printed resources is allowed except direct communication that violates Georgia Tech Academic Integrity Rules. However, internet search on course-related topics is not allowed during the exam period. Furthermore, public piazza posting about the exam questions is not permitted. If you need any clarification, please use private posting that are visible only to the instructors.

Name ____________________________________

Problem	1	2	3	Total
Score	/16	/17	/17	/50

1. Heart Disease. One of the earliest studies on heart disease started in 1960 and used 3,154 healthy men, aged between 39 and 59, from the San Francisco area. At the start of the study, all were free of heart disease. Eight and half years later, the study recorded whether these men now suffered from heart disease. The file `heart.csv` contains the data on the coronary heart disease (chd) along with other variables that might be related to the chance of developing this disease: age (in years), height (in inches), weight (in pounds), sdp (systolic blood pressure in mm Hg), dbp (diastolic blood pressure in mm Hg), chol (cholestrol), and cigs (number of cigaretts smoked per day). Please note that there are a few missing values in the dataset. Please also encode the 'chd' variable with 'yes' as 1 and 'no' as 0.

1. Fit a Bayesian logistic regression model with chd as the response and the other seven variables as the predictors. Use non-informative priors. A suggested OpenBUGS distribution for the missing chol[] data is dnorm(200,0.01) but you are free to try others as well. Please run 10,000 MCMC iterations after a burn-in of 1000.
2. Plot the posterior densities of the coefficients of the seven predictor variables.
3. Provide 95% credible intervals for the coefficients of the seven predictor variables.
4. Find the posterior distribution of the probability of getting the heart disease for a man with median values for the seven predictors. (Plot the density.)

2. Pulp brightness. Consider the following experiment, which was performed at a pulp mill. Plant performance is based on pulp brightness as measured by a reflectance meter. Each of the four shift operators (denoted by A, B, C, and D) made five pulp handsheets from unbleached pulp. Reflectance was read for each of the handsheets using a brightness tester. The data is given in the table below:

Treatment			
A	B	C	D
59.8	59.8	60.7	61.0
60.0	60.2	60.7	60.8
60.8	60.4	60.5	60.6
60.8	59.9	60.9	60.5
59.8	60.0	60.3	60.5

Solve the problem as a Bayesian one-way ANOVA. Use STZ constraints on treatment effects.

1. Do the operators differ in making the pulp handsheets and reading their brightness? Look at the 95% credible sets for the differences between treatment effects.
2. Find the 95% credible set for the contrast $\mu_1 - \mu_2 - \mu_3 + \mu_4$, where μ_1, μ_2, μ_3, and μ_4 are the mean pulp brightness for the operators A, B, C, and D, respectively.

3. Enzyme. The data in the file `enzyme.csv` gives the initial rate of reaction of an enzyme (y) and the substrate concentration (x). Consider the following nonlinear regression model:

$$y = \frac{\theta_1 x}{\theta_2 + x} + \epsilon,$$

where $\epsilon_i \sim^{iid} N(0, \sigma^2)$, $\theta_1 > 0$, and $\theta_2 > 0$. Assume noninformative priors for θ_1, θ_2, and σ^2 (in BUGS, we will specify a prior for the precision $\tau = \frac{1}{\sigma^2}$). For these constraints, the recommended priors in BUGS are $Gamma(0.01, 0.01)$.

1. Plot the marginal posterior densities of θ_1, θ_2, and σ^2. Use $\theta_1 = 200$, $\theta_2 = 0.1$, and $\sigma^2 = 100$ (equivalently, $\tau = 0.01$) for initializing the MCMC chain.

2. Compute 95% credible intervals for each of the three parameters.

3. Plot the posterior predictive distribution of y when $x = 1.0$ and provide its 95% credible intervals.