Final Exam

ISyE6420
Spring 2025

Released April 25, 6:00 PM ET – due April 27, 6:00 PM ET. This exam is not proctored and not time limited except the due date. Late submissions will not be accepted.

Use of all course materials, including the class Github site, the textbook, and any personal notes are allowed. Internet search using exam-related keywords is not allowed during the exam period, nor is any communication with other students relating to the exam. You should start with methods from our course materials to solve all problems and fully explain your reasoning. Your answers must be fully supported by your work. If you need any clarification on the questions, please use the private posting function on Ed Discussion so that your post is only visible to the instructors.

Discussing the exam questions with anyone outside of the course staff and/or posting any portion of the exam to non-GT sites are considered serious violations of the Georgia Tech honor code. AI tools, like ChatGPT/Copilot and other similar ones, are not allowed for any purpose.

All questions on this exam must be solved with a Probabilistic Programming Language (PPL); for example, OpenBUGS or PyMC. Further, frequentist answers will not be accepted for credit.

Please read and sign (or e-sign) the following honor pledge and submit a copy along with your answers.

I pledge on my honor that I have completed the exam on my own and I have not used any unauthorized materials or taken anyone's help for completing this exam.

Name _____

Problem	1	2	3	Total
Score	/35	/30	/35	/100

Spring 2025 Final Exam Problem 1. A questionnaire was distributed to 12,231 passengers on randomly selected buses. One of the questions was "How satisfied are you with the punctuality of this bus?" and the answer was classified into five categories: very dissatisfied, dissatisfied, neutral, satisfied, and very satisfied. The delay of the bus was also recorded at 0, 2, 5 or 7 minutes delay. The data is given in the file `bus.csv`. Answer the following questions.

1. Fit a multinomial logistic regression to the data with the same prior from the lecture and provide the posterior summary output.
2. Find the probability that a customer is satisfied or very satisfied when the delay is 1 minute.
3. Give the odds ratio (and credible set) that a passenger would report "Satisfied" vs. "Very Dissatisfied" per minute of delay.

Spring 2025 Final Exam Problem 2. The dataset `pima.csv` contains information on 768 adult female Pima Indians living near Phoenix. Almost half of the data have missing values. Answer the following questions after standardizing the data (using 1 standard deviation to scale).

1. Replace the missing values with the mean of the corresponding variable and fit a logistic regression using `test` as the response variable and the remaining eight variables as input. Which variables appear to be significant?
2. Fit the logistic regression again, still working with the standardized data. This time, assume a standard normal model ($N(0,1)$) to impute the missing values. Are there any differences with the coefficients from the previous result?
3. Using the model from part 2 as a base, perform Stochastic Search Variable Selection (SSVS) to choose the best set of predictors. Report the best 5 models by posterior model probability. **Hint:** divergences are common here. You can set `target_accept=0.99` in PyMC to mitigate them.

Spring 2025 Final Exam Problem 3. The file `calibration.csv` contains the data for an ultrasonic calibration, where the response variable is ultrasonic response (y), and the predictor variable is metal distance (x). The following nonlinear model is postulated:

$$y = \frac{\exp(-\beta_1 x)}{\beta_2 + \beta_3 x} + \epsilon,$$

where $\epsilon|\sigma^2 \sim^{iid} N(0, \sigma^2)$ and $\beta_1, \beta_2, \beta_3 > 0$.

1. Fit the nonlinear regression model to the data with noninformative priors (use initial values $\beta_1 = 0.2$, $\beta_2 = 0.01$, $\beta_3 = 0.01$, and $\sigma = 1$).

2. Predict the ultrasonic response for 100 values of $x \in [0, 6]$. Show the curve on the scatter plot of x versus y along with its 95% credible intervals.

3. Find the posterior mean and 95% credible intervals of the metal distance that can achieve an ultrasonic response of 20 on average.

Hint for Part 3:

Part 3 must be completed outside the model with a standard optimizer such as `scipy.optimize.newton` (useful because it is vectorized) or R's `uniroot`. Once the roots are found, summarize them with the posterior mean and a 95% credible interval.