OR 664 / SYST 664 / CSI 674: Midterm Exam

due March 23, 2020 at 11:59 PM

Each question is worth ten points, for a total of 100 points for the midterm exam. Show all your reasoning. You will receive at least some credit for any question if you make an honest attempt and explain what you were thinking. The exam is open book and open notes, but you must work by yourself. Any collaboration with others or use of the work of other students is an honor code violation. I will be available by phone or email to answer clarification questions.

1. A security screening system photographs people entering a facility, matches them to a database of terrorist suspects, and alerts security to stop an individual for further scrutiny if a match is found. The system has a miss probability of 12% and a false alarm probability of 4%, where a miss is defined as failing to issue an alert for a person in the database, and a false alarm is defined as issuing an alert for a person who is not in the database. Assume that true positives and true negatives cost nothing, and the cost of a miss is 150 times the cost of a false alarm. Let p be the prior probability that a person is in the database. Plot the expected loss of three policies: (1) stop everyone for questioning; (2) stop no one for questioning; and (3) stop someone for questioning if an alert is issued. For what range of p is each policy optimal? Comment on your answer.

2. Historically, one in every 50 fourth-graders at an elementary school fails to reach minimum threshold on a state-mandated reading exam. The school is planning a study to evaluate a new reading curriculum. A school administrator assesses a 55% chance that the new curriculum would reduce the failure rate for the exam. That is, there is a 55% chance that if the new curriculum were implemented, fewer than 1 in 50 students would fail the exam. Suppose the administrator also assesses a 45% chance that the failure rate would be cut in half and a 30% chance that the fail rate is improved by a factor of 10. That is, there is a 45% chance that fewer than 1 in 100 students would fail the exam, and a 30% chance that fewer than 2 in 1000 students would fail the exam under the new curriculum. The administrator also assesses a 35% chance that the failure rate would more than double, or at least 4% of the students would fail the exam. If you were to use a Beta distribution to fit these judgments, what parameters would you use? Do you think it provides a good fit? Justify your answer.

3. A shape recognition system classifies objects as round, rectangular, or irregular. It correctly classifies round objects 80% of the time, rectangular objects 85% of the time, and irregular objects 70% of the time. Incorrectly classified objects are equally likely to be classified as either of the two incorrect object types. Assume in a given environment that 15% of the objects are round, 25% of the objects are rectangular, and 60% of the objects are irregularly shaped. Find the joint distribution of object shapes and classification results. If the system reports that an object is rectangular, what is the posterior probability that the object has each of the three shapes given the system report?

4. A drug for treating depression is undergoing clinical trials. A total of 120 patients were enrolled in a randomized, double-blind, placebo-controlled study. Half of the patients were randomly selected to receive the drug; the remaining patients were given a placebo. Questionnaires were administered at the start of the study and 30 days after treatment

began. One of the questions on the post-study questionnaire asked whether patients experienced improvement in their mood since the start of the study. The table below shows responses to this question.

	No Improvement	Improvement	Total
Treatment	18	42	60
Placebo	26	34	60
Total	44	76	120

Let Y_1 and Y_2 be the number of patients in the treatment and placebo groups who experienced improvement. Assume Y_1 and Y_2 are independent random variables with Binomial(60, Θ_i) distributions, for i=1,2. Assume Θ_1 and Θ_2 are independent with a Beta distribution with shape parameters ½ and ½ (this is the Jeffreys prior distribution). Find the joint posterior distribution for Θ_1 and Θ_2. Name the distribution type and its hyperparameters. Plot the posterior density functions for Θ_1 and Θ_2 on the same axes. Comment on your results.

5. Generate 5000 random pairs (θ_{1i}, θ_{2i}), i=1, …, 5000 from the joint posterior distribution for (Θ_1, Θ_2). Use this random sample to estimate the posterior probability that the rate of improvement is higher for treatment than for placebo. Does your analysis support the hypothesis that the drug alleviates symptoms of depression? Explain clearly your process for generating the sample. Discuss your analysis and results.

6. Suppose 25 new patients are given the drug. Using the posterior distribution from problem 4 as the prior distribution for the probability of reporting improvement, find the predictive distribution for the number of patients who will report improvement 30 days after receiving the drug. State the distribution type and parameters. Find the posterior probability that 20 or more patients will report improvement in 30 days.

7. A researcher is performing a study of plant growth. To model the number of plants of a given species that will be found on a 10 m^2 plot of land, she uses a Poisson distribution with unknown rate parameter Λ that depends on the species and the soil conditions. Based on her knowledge of plants and soil conditions, she assesses a conjugate Gamma prior distribution with shape 2 and scale 6 for the mean number of plants growing on a 10m^2 plot of land of a given type. She plans to count the number of plants growing on five plots of this given type. What is her predictive distribution for the total number of plants she will find growing on the five plots? Find the probability that she will count fewer than a total of 75 plants growing on the five plots.

8. The researcher goes out into the field and counts the plants she finds on the five plots. She counts a total of 72 plants. Find her posterior distribution for the Poisson parameter Λ. Name the type of distribution and hyperparameters. Find a 95% credible interval for Λ.

.

9. Management at a call center is investigating the call load in order to find an efficient staffing policy. Assume that time intervals between calls are exponentially distributed. Assume the mean time between calls Θ is constant during the mid-morning period. Assume an inverse Gamma prior distribution with shape α =4 and scale β= 0.0015 for Θ. The following sequence of call times was collected during mid-morning, measured in seconds after the start of data collection: 168, 314, 560, 754, 1215, 1493, 1757, 1820, 1871, 1982, 2134, 2430, 3187, 3388, 3485. Find the posterior distribution for Θ. Find the prior and posterior mean and standard deviation for Θ. Discuss. (*Note: Because of the memoryless property of the exponential distribution, you can treat the time until the first call as having an exponential distribution.*)

10. Do you think the model of independent and identically distributed exponential observations is a good model for the data of Problem 9? Explain your reasoning.