STAT4102 Final/Extra Credit Options

Option 1: Take the final as on the syllabus. Do nothing extra.

Option 2: Take the final as on the syllabus. Additionally, choose either #3 or #4 below to do for extra credit, to increase a midterm exam score by up to one letter grade.

Option 3: Do all 6 below instead of taking the final. Your score on #5, #6 and the best three of the other four will count as your final exam grade. The other will be extra credit towards a midterm as in Option 2. Due by 3:30 on May 15. Drafts turned in for comments by class time on May 5 will be returned no later than in class on May 9.

Notes on Academic Integrity

This needs to be your own work, and no one else's. However, working together to understand the material and check your understanding is okay. Similar to the homework, a general rule is that verbal discussion and working out of similar problems together is okay, but that copying exact phrasings and mathematics is not. I can almost always tell when you're just parroting something, anyway.

As for checking your understanding, you should work out your calculations and write your explanations; once you have it written out, it is okay to ask a friend to read it. It's okay for the friend to fix minor math and english errors; it would not be okay for the friend to completely redo a mathematical section or explanation. Think proofreading, not rewriting. If it's significantly wrong, discuss a different problem together until you think you understand, then go back to your own work. If you are going to be working with others in this manner, choosing different data sets/distributions will help keep problems to a minimum.

I recognize that this may leave some grey areas. Please ask if you have any questions. The main thing I want you to keep in mind is that you need to fully understand everything you turn in; you should feel like you could easily recreate anything if I were to ask you to.

1. Point Estimation

For a random sample of size n from a distribution of your choice with one parameter, find the likelihood function and a sufficient statistic for that parameter. Give a reason why sufficient statistics are called sufficient. Find the MLE and MOM estimators for that parameter. Are they unbiased? Consistent? Efficient? Also explain why they are good properties for a point estimator. Finally, find a UMVUE estimator for the parameter, using the Rao-Blackwell theorem. Explain why it's possible for the UMVUE estimator to not have the lowest MSE among all possible estimators. (Parts of this may be difficult or impossible for certain distributions, so choose carefully. It's okay to choose an "easy" one.)

2. Large Sample Tests

Find a data set (not used in the book or in class) where a large-sample test about the differences between two populations (either means or proportions) is appropriate. Explain what the Central Limit Theorem says and why this makes the large-sample test appropriate for this data set. Write out sensible hypotheses and perform the test (at a chosen alpha level) by finding the rejection region in terms of the difference in sample means or proportions. Be sure to write out your conclusion in terms of the original problem. Choose a value in the alternate hypothesis which seems of practical significance to you, and calculate the power for this value. (You may need to make some assumptions about the standard error; that's okay.) Using the rejection region you found originally, what sample size is necessary to make the power 0.9? (This could be a larger or a smaller sample size depending on the power with your existing sample size.)

3. Neyman-Pearson

For a sample of one data point from a distribution of your choice, set up a simple hypothesis and use the Neyman-Pearson Lemma to find the rejection region for a particular alpha value of your choice. Explain why the Neyman-Pearson Lemma is used in this case, what it guarantees, and give a explanation (not a proof) why this Lemma makes sense. Extend your simple hypothesis to two different composite hypotheses, one for which Neyman-Pearson can be used to show that your original rejection region is UMP and one where it is not. (You may have to choose your original hypothesis carefully to make sure that both of these are possible.) Carefully explain why the one is UMP and the other is not. For the one that is not, find another rejection region that has more power than your original rejection region for a given particular value in the alternate hypothesis. (The chosen distribution should NOT be one that we have done a Neyman-Pearson test for in class or on the practice or actual exam.)

4. Likelihood Ratio

Find a data set (not used in the book or in class) where there is a hypothesis of interest about three or more parameters. Two possible options are 1) a data set where the hypothesis of interest is either independence or homogeneity or 2) data sets from three or more normal distributions where you're testing equality of means. Write out the hypotheses and test it using the likelihood ratio test at a chosen alpha level. Be sure to explain each step along the way; how you find the likelihood function, how you get the maximums, how you determine the degrees of freedom, etc. Also explain why the likelihood ratio test in general seems sensible. (A resource for possible data sets is http://lib.stat.cmu.edu/DASL/.)

5. Nonparametric

Find a data set (not used in the book or in class) for which you think a nonparametric test would be appropriate. Explain why. Set up the desired hypothesis and perform the nonparametric test at a chosen alpha level as well as the the corresponding parametric test. Explain carefully what the nonparametric is testing and how that differs from the corresponding parametric test. Write up the results of both of your tests, being sure to note if they agreed or disagreed. If they agreed, why do you think that happened? Do you still think the nonparametric test was appropriate? If they didn't agree, explain which test you prefer and why. (A resource for possible data sets is http://lib.stat.cmu.edu/DASL/.)

6. Bayesian

Find (or create) a data point from a binomial distribution with an unknown probability p. (It should have a sample size n of at least 30.) Create a prior distribution (using a Beta distribution) for p, and explain why you chose it. Then derive the posterior distribution for p and find a 90% credible interval using this posterior distribution. Additionally, find a 90% confidence interval using large sample methods. In your answer, be sure to clearly describe how these two intervals should be interpreted and to explain the fundamental differences between Bayesian and Frequentist statistics.