MATH 5310: Probability
Fall 2024
Homework 1
Due before 11:59pm on Tuesday 9/10

Part1:

1. What is the most rigorous mathematical definition for probability that you know or have learned?

2. How would you describe the purpose of probability to a friend that has a limited math and stats background?

3. Give examples of the following three sample spaces
   (i) discrete and of infinite dimension,
   (ii) finite and continuous
   (iii) something of interest to you

4. The statistical inference method of *hypothesis testing* is based on framing a scientific question as deciding between one of two competing hypotheses called the *null* and *alternative* hypotheses. The usual approach with hypothesis testing is have the null represent no effect or no change resulting from an intervention or treatment and the alternative hypothesis would be that the intervention or treatment had an effect or caused a change. For example, if a researcher wanted to use hypothesis testing to show that free lunch improved grades in students, they would set the null hypothesis as "free lunch does not improve grades" and set the alternative as "free lunch does improve grades".

Next, the researcher collects data and decides which hypothesis better matches the data. However, to try to reduce the effect of random noise and random chance leading you to falsely claim a treatment has an effect when it really doesn't, a research will assume the null is true unless the data favors the alternative so strongly that there is at most a small percent (for example, at most 5%) chance that the magnitude of the effect could have been from unlucky sampling. For example, if schools with free lunch had a very small increase in grades, this might just be because of random noise (samples naturally vary some); however, if the grades increased dramatically, this is very unlikely to have been because of random sampling - that is, it would be really hard to choose such a biased or skewed sample by random chance, meaning it is much more likely the large effect was real and hence the alternative hypothesis is very likely true. The logic is to assume an intervention, policy, drug, etc. has no effect until the data is overwhelming showing that it has an effect that would be hard to see by random chance; that is, you are observing an effect of a size that is too large to have reasonably been just unlucky sampling if the null (no effect) was actually true.

There is no guarantee in sampling that you get the perfect data. You can't sample in a way to make sure your data always leads you to make the right conclusions. For example, simply by random chance you sometimes get data that causes you to reject the null when in fact it was true - this is called a Type I error. When designing their study and analytical approach to make conclusions, the researcher chooses the level of risk of a Type I error that they can accept, and

this is what is known as the significance threshold or level (e.g. the 5% mentioned above). Even more, if a researcher is conducting many hypothesis simultaneously (e.g. answering multiple scientific questions with different hypothesis tests), there is an increased risk of making a Type I error; that is, doing more and more tests gives random chance more and more tries to give a researcher data that causes them to falsely conclude that an alternative hypothesis is correct when it is not. For example, let's assume a researcher is conducting two hypotheses tests, and chooses 0.05 or 5% as the acceptable risk (significance level) of a Type I error. If the truth is that both null hypothesis are true, the probability of correctly concluding this is 0.95*0.95 or 0.9025. That is, there is only a 1-0.9025 or 0.0975 probability of not making at least one Type I error. This is what is known as the Family-Wise Error Rate (FWER) for a Multiple Testing setting. Error here refers to Type I errors, and Multiple Testing refers to doing multiple simultaneous hypothesis tests.

To correct for the inflated Type-I error rates, which means the FWER is greater than the implied desired acceptable level used in the individual tests (in the example above 0.0975>0.05), a researcher can use what is known as the Bonferroni correction. Mathematical, the procedure divides up the significance level, representing the acceptable chance of a Type-I error (5% in our example), by the number of tests being conducted (5%/2=2.5% in our example). Then, each individual hypothesis test is done using this new "Bonferroni adjusted" significance level instead of the original one. Here, a researcher would then look at the two hypothesis tests individually and reject the null if and only if there was a 2.5% or less chance of an observed effect (e.g. a large magnitude of change observed in the data) being the result of random sampling.

(i) Use Bonferroni's Inequality (Example 1.2.10) to show the Bonferroni Correction method will result in a FWER of at most a 5% in the two hypothesis test scenario described above.
(ii) The usual Bonferroni Correction assumes that the hypotheses tests are independent of each other. Why is this important to your calculation in (i)?

Part 2:

Casella and Berger Problems:
1.4 (a) and (b) only
1.8
1.11 (c) only
1.18 or 1.23
1.31