STAT400: Midterm Review

Fall 2021

Sampling

1. Review mixture distributions, specifically the set up to approaching a problem?

2. Review how to pick an envelope function when using an accept-reject algorithm to sample from a distribution.

3. Spend a bit of time on the Accept Reject algorithm it would be the most helpful thing for me. Do a full example problem, while skipping most of the r coding and focusing on the parts that would be most relevant to the exam.

Monte Carlo Integration

1. Review a change of variable Monte Carlo problem? Something similar to page 9 of the Monte Carlo integration notes packet.

2. When estimating the variance of a Monte Carlo estimate, we don't have access to the true "theta", so we instead "plug in" the Monte Carlo estimate itself. What licenses us to do this? It seems like this would increase the variance of our estimate of the variance. Perhaps we are not very concerned about this, but I have seen this in other places as well (plugging in an estimate for a parameter when the parameter is unavailable). Are we allowed to do this because of unbiasedness, or some other property?

3. Review selecting f(x) and g(x) for Monte Carlo integration. Doing the homework, making choices did not seem as hard due to the nature of homework reinforcing things learned in class. I am way less confident when trying to do these problems without out the specific section notes alongside. Going over strategies for selecting one of the three methods in chapter 6, section 1.7 and then choosing the correct f and g based on the decision would be great.

4. When we were looking at estimating a CDF, we looked at three different Monte Carlo methods. The first involved sampling from a Unif(0,1) and using a transformation and the second sampled from a Unif(0,x) with no transformation. Based on a simulation I did, I believe both estimators have an equal variance. Is there a reason I would choose one of these methods over the other? Is there any sort of bias associated with the transformation made in the first method? If so, why?

5. Another question that I have is if we can go over the derivations for question 3 on homework 6.

6. In the notes from 10/07 on importance sampling, can you talk more about the meaning of naive monte carlo estimators vs (presumably) non-naive?

7. How to find truncated distribution again (option 2b for importance sampling)?

8. How to we pick a good phi(x) so that our sampling weights are appropriate?

9. Which phi function was best when looking at the graphs

10. Go step by step through an importance sampling problem. Any type of problem is good. Just getting practice would be helpful for me.

11.

When it comes to Monte Carlo integration and importance sampling, both f(x) and the phi are distribution samples. How do you know which distribution to use for which function? And can they be the same distribution?

12. In the importance sampling packet example 2.2, we set the coefficient of variation equal to 5%; will we always set it to this value for these types of problems?

13. how do we correct the issue of not drawing from the target distribution by weighing each role of 1 by 1/3?

14. (Less related to studying for the midterm) In practice, besides the heuristics mentioned in class (ie, that h(x)/phi(x) should be flat), are there any other considerations that go into the selection of an importance function? I found it hard to choose a good importance function on the homework, and I imagine it would be even harder for a real-life problem.

15. Mathematical details of importance sampling. perhaps with the discrete example and then a corresponding continuous example. I understand the algorithm somewhat well, but I would like to understand better why it works. The idea as far as I got from the lecture is that we sample unlikely events to make our estimate converge quicker, but I thought that unlikely events are what messes up our estimate. Maybe it's the trade off in bias and variance that I am not getting a good grip on.

Other Questions

1. Are we expected to study limit theorems for the exam?

2. Are we expected to review integration techniques for the exam?

3. Are we expected to do any complex derivations or integrations or even simpler something like integration by parts or the power rule?

4. A smaller and more specific question that I have is if we will have to do any integration by parts or other similarly difficult forms of integration on the exam. It shouldn't be a problem if I allocate the space on my formula sheet, but I figure it is worth checking.

5. The notes often refer to something being unbiased, can you elaborate? For instance, in chapter 6: monte carlo integration, page 6.

6. Are we going to have to write code at any point in the exam? Or is it mostly just going to be code interpretation from the beginning of the class?

7. What topics are going to be covered on the exam? I know you mentioned since it's not a coding exam then topics related to R will not be covered but could you be more specific?

8. What do you recommend studying for for the exam?

9. Similar to the question above, any recommendations on what to put on the formula sheet? I am a little confused what I should put on it since you will be providing a distribution table and you don't want worked through examples on it.

10. Will we be allowed a calculator to use on the exam/will a calculator be needed?