STAT400: Midterm Review

Fall 2021

Sampling

accept-reject.

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.

Mixtures

def'n: A r.v. Y is a discrete mixture if $F_Y(y) = \sum \theta_i \underbrace{F_{X_i}}_{\text{cdf.}}(y)$ for some sequence $X_1, X_2, \ldots$ s.t. $\theta_i > 0$ and $\sum \theta_i = 1$.

If $X_1, X_2 \ldots$ are cts, if $\underbrace{f_Y}_{\text{pdf}}(y) = \sum \theta_i f_{X_i}(y)$.

For 2 r.v.'s:

$$f_Y(y) = \theta f_{X_1}(y) + (1-\theta) f_{X_2}(y).$$

NOTE:

mixtures $\neq$ sums of r.v.'s.

How to simulate from a mixture:

1) Simulate $Z \sim \text{Bern}(\theta)$
2) If $Z = 1$, $Y \sim f_{X_1}$

   $Z = 0$, $Y \sim f_{X_2}$

Accept-Reject

Goal: Want to sample from f.

Algorithm:

1) Find a suitable proposal density g and envelope e
   - g must be easy to sample from and have support that includes support of f, i.e. $\underbrace{\{x : g(x) > 0\}}_{\mathcal{X}_g} \supset \underbrace{\{x : f(x) > 0\}}_{\mathcal{X}_f}$
   - e must be s.t. $e(x) = \underset{\text{constant}}{c} \cdot g(x)$ and $e(x) \geq f(x)$ $\underset{\text{"For all"}}{\forall} x \in \mathcal{X}_f$.
2) Sample $Y \sim g$.
3) Sample $U \sim \text{Unif}(0,1)$.
4) If $U < f(Y)/e(Y)$ accept Y (else reject).
5) Repeat from 2 until desired sample size accepted.

while loop.

example Sample from Frechet (3) distribution.

$f(x) = 3x^{-4} e^{-x^{-3}}$ $x > 0$. Want to sample from this dsn.

$\mathcal{X}_f = \{x : x > 0\} = [0, \infty)$.

let g = student's t dsn w/ 1 df $\Rightarrow$ $c = 7$ works.

Plan:

1. Sample $Y \sim t_1$
2. Sample $U \sim \text{Unif}(0,1)$.
3. If $U < f(y)/e(y)$ accept else reject
   - ↳ f is Frechet (3) pdf
   - $e = 7 \cdot t_1$ pdf.
4. repeat until accepted n
   - ↳ use while loop.

Monte Carlo Integration

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

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 ← Yes 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? Yes + Consistency.

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. Any one method that works is great. (some may have lower variance).

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?

   should be the same.
   We choose the 1 that is easiest.

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

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? MC no importance.

7. How to find truncated distribution again (option 2b for importance sampling)? remember $\int_{\mathcal{X}} f(x)\,dx = 1$.

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

   $\phi(x) > 0$ when $f(x) > 0$ and $\frac{f(x)}{\phi(x)}$ large only when $g(x)$ small
   i.e. $\frac{f(x)}{\phi(x)} \cdot g(x) \approx$ constant.

10. Go step by step through an importance sampling problem. Any type of problem is good. Just getting practice would be helpful for me. See homework 7 solution

11.

Monte Carlo Integration

Goal: evaluate $\theta = \int h(x)\,dx$.

① Derive

a) Find appropriate f and g to rewrite $\theta = E[g(X)]$, $X \sim f$

① choose f to have same support as limits of integration

② transform limits of integration to support of f. (change of variable in integral).

③ Use an indicator function as part of g to make limits equal the support.

b) For importance sampling: Find ϕ s.t. $\phi(x) > 0$ when $f(x) > 0$.

and s.t. $\frac{f(x)}{\phi(x)}$ is large only when $g(x)$ is small.

i.e. $\frac{f(x)}{\phi(x)} \cdot g(x)$ is approximately constant.

② Plan.

For MC:

1. Sample $X_1, \dots, X_m \sim f$
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} g(X_i)$.

For importance sampling:

1. Sample $Y_1, \dots, Y_m \sim \phi$
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} g(Y_i) \underbrace{\frac{f(Y_i)}{\phi(Y_i)}}_{\text{importance weights.}}$

③ Do it.

Hmwk 6 problem 3

$$\theta = \int_0^{0.5} e^{-x}\, dx$$

support = limits of integration.

① let $f \sim Unif(0, 0.5) \Rightarrow f(x) = \begin{cases} 2 & x \in [0, 0.5] \\ 0 & o.w. \end{cases}$

1) rewrite θ as expected value.

$$\theta = \int_0^{0.5} e^{-x}\, dx = \int_0^{0.5} e^{-x} \cdot \frac{2}{2}\, dx = \int_0^{0.5} \frac{e^{-x}}{2} \cdot \underbrace{2}_{f(x)}\, dx = E\left[\frac{e^{-X}}{2}\right], \ X \sim Unif(0, 0.5)$$

$$\Rightarrow g(x) = \frac{e^{-x}}{2}$$

2) Plan.

1. Sample $X_1, \ldots, X_m \sim Unif(0, 0.5)$.
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} \underbrace{\frac{1}{2} \cdot e^{-X_i}}_{g(X_i)}$

optimal 3. $\hat{Var}(\hat{\theta}) = \frac{1}{m^2} \sum_{i=1}^{m} \left[\left(\frac{1}{2} \cdot e^{-X_i} - \hat{\theta}\right)^2\right]$.

② let $f \sim Exp(1). \Rightarrow f(x) = e^{-x} \quad x \geq 0.$

support $= [0, \infty) \neq$ limits of integration!

1) rewrite θ as expected value.

$$\theta = \int_0^{0.5} e^{-x}\, dx = \int_0^{\infty} \mathbb{I}[x < 0.5] \cdot \underbrace{e^{-x}}_{f(x)}\, dx = E\left[\mathbb{I}[X < 0.5]\right], \ X \sim Exp(1).$$

$$\Rightarrow g(x) = \mathbb{I}(x < 0.5).$$

2) Plan.

1. Sample $X_1, \ldots, X_m \sim Exp(1)$
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} \left[\mathbb{I}(X_i < 0.5)\right]$

optimal 3. $\hat{Var}\, \hat{\theta} = \frac{1}{m^2} \sum_{i=1}^{m} \left[\left(\mathbb{I}(X_i < 0.5) - \hat{\theta}\right)^2\right]$.

③ let $f \sim Unif(0, 1) \Rightarrow f(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & o.w. \end{cases}$

1) rewrite θ as expected value.

$$\theta = \int_0^{0.5} e^{-x}\, dx \overset{C.o.V.}{=} \int_0^1 e^{-(\frac{1}{2}y)} \cdot \frac{1}{2}\, dy = E\left[\frac{1}{2} e^{-(\frac{1}{2}Y)}\right], \ Y \sim Unif(0, 1).$$

let $y = 2x \Rightarrow x = \frac{1}{2}y$

$dx = \frac{1}{2}\, dy.$

$$\Rightarrow g(y) = \frac{1}{2} e^{-(\frac{1}{2}y)}$$

2) Plan

1. Sample $X_1, \ldots, X_m \sim Unif(0, 1)$.
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} e^{-\frac{1}{2}X_i}$

optional 3. $\hat{Var}(\hat{\theta}) = \frac{1}{m^2} \sum_{i=1}^{m} \left[\left(\underbrace{\frac{1}{2} e^{-\frac{1}{2}X_i}}_{g(X_i)} - \hat{\theta}\right)^2\right]$.

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?

You have already picked f in MC. and tried importance sampling later to do better.

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?

Not always but common choice.

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

biasing dsn and reweighting estimate by importance weights.

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. contact me to discuss offline.

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.

unlikely events show up less => take sample mean at those values there are less points in the mean.

$$\mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m} g(X_i)\right) = \underbrace{\frac{1}{m}\mathrm{Var}\, g(X_1)}_{\text{small } m,\ \text{quantity } \uparrow} \qquad X_1, \ldots, X_m \text{ iid}$$

Other Questions

all class material is fair game.

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. $\hat{\theta}$ is unbiased if $E(\hat{\theta}) = \theta$.
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?

   Not write, but maybe pseudocode.
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? everything up to & including Oct 19.
8. What do you recommend studying for for the exam?

   homeworks, class examples.
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.

   anything you don't want to memorize that isn't an example.
10. Will we be allowed a calculator to use on the exam/will a calculator be needed?

    No and No.