

# STAT400: Midterm Review

Fall 2020

## Sampling

- non uniform envelope curve*
1. Would we be able to go over how to find the envelope for the accept reject method?
  2. How to find the envelope value?
  3. Concerning Accept-Reject: How can we be sure that our envelope function is greater than the function we are sampling from, for infinite support?
  4. Can you talk more about choosing a constant to multiply by when using accept-reject and not using the uniform as an envelope function? I would like to see outliers where choosing a very large C and letting it run for longer would still not work (I know these situations exist).
  5. Can you do an example of the accept reject method where you cannot use the uniform(0, 1) trick using the maximum and instead have to guess and check for the value of c for the envelope?
  6. Can you review finding an envelope function in cases where the support is not  $0 \leq x \leq 1$ ? Is there a good, systematic way to go about finding an envelope if you don't immediately recognize the shape of the distribution or if you can't solve for the maximum?
  7. For the accept reject ratio graph (image pasted): I know that y is g distributed,  $f(y)$  is just a point on the green  $f(x)$  curve,  $e(y)$  is just a point on the purple  $e(x)$  curve, and both y and x share the x-axis, but I don't understand what the red points are. Neither of them can be points on a density, can they? I know that  $g(x)$  is just another curve, but it doesn't look like the red dots are height values in the  $g(x)$  density. So what are the red points?

## Accept/Reject (Q1-6)

### Algorithm

1. Find suitable  $g$  and  $c$

- $g$  needs to be <sup>①</sup> easy to sample from and <sup>②</sup> easily calculate  $g(x)$ , support of <sup>③</sup>  $g$  includes support of  $f$ .
- $c \cdot g(x) = c \cdot g(x)$  where  $c(x) \geq f(x) \forall x$ .

2. Sample  $Y \sim g$ .

3. Sample  $U \sim \text{Unif}[0, 1]$

4. If  $U < f(Y)/c(Y)$ , accept  $Y$

repeat 2-4 until done.

Assume support of  $f$  is not finite.  $\mathcal{X} = \mathbb{R}$ .

① pick  $g$  from a table of distributions

↳ plot  $x$  vs.  $f$  to see if there are asymptotes on  $\mathcal{X}$

↳ paying special attention to support of  $f$  and support choices for  $g$ .

② find  $c$  s.t.  $c \cdot g(x) \geq f(x) \forall x$ .

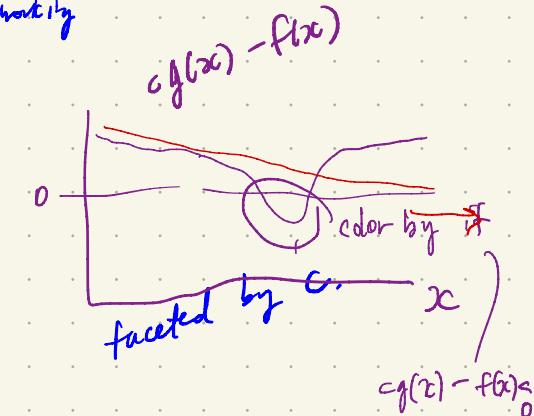
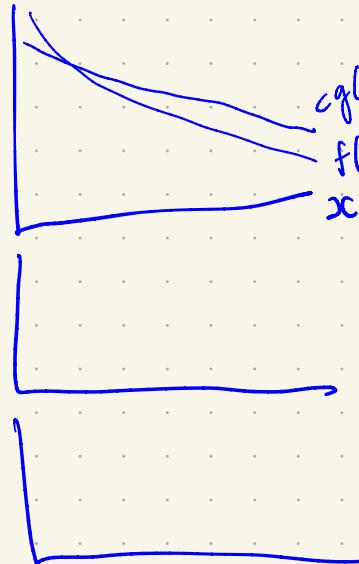
↳ try to solve for  $c$   
this can be hard.

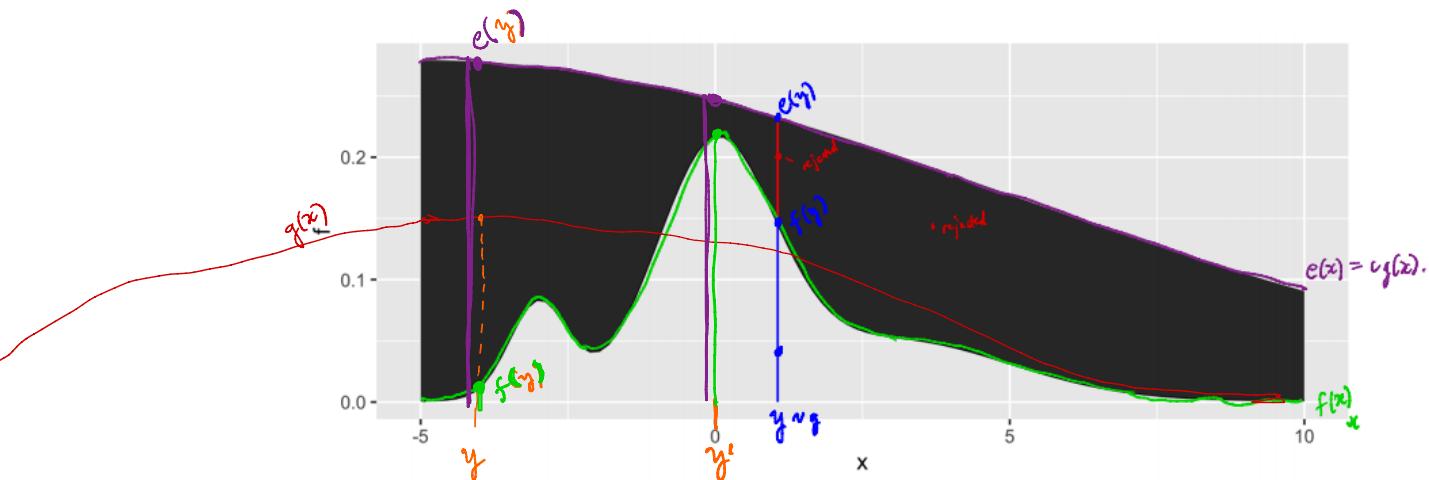
↳ make plots.

very wide of  $x$ 's to get a sense of potentially working

info:

- ① pick many  $c$ 's
- ② pick many  $x$ 's
- ③ for each  $c$ , plot



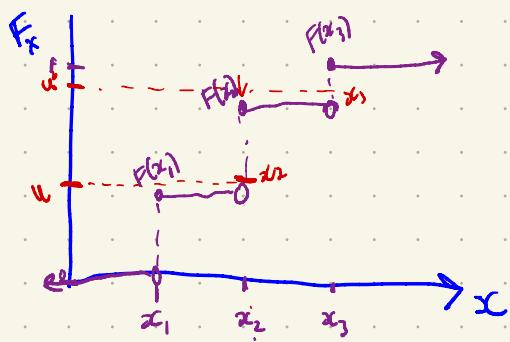


8. I have some practical questions regarding the accept-reject method and what you have previously stated about the support of the envelope encompassing the support of the pdf we're trying to sample from.
- In a lecture, you gave an example of a uniform(-10,10) distribution not being appropriate for a normal(0,2) distribution since the support for the normal distribution goes from -infinity to infinity while the uniform distribution goes from -10 to 10. However, if we for example consider dnorm(100,0,2), R rounds the pdf down to 0 anyway. Wouldn't it be the case then, that any time we use a normal pdf as an envelope, we are essentially truncating the distribution we are attempting to sample from (due to rounding)?
  - If this is the case wouldn't it be more realistic to use something like a uniform(-100,100) distribution (efficiency notwithstanding), where the pdf is calculated as nonzero for larger values as opposed to most distributions with infinite support?
  - And if we were to use a uniform distribution in such a manner on the exam, would it be technically incorrect even if it produces a better approximation?
  - On a related note, I do not remember if this was mentioned in a lecture, but instead of using a single uniform distribution, can we use multiple (in a manner reminiscent of Riemann Sums) to improve efficiency?
9. In HW 5 when we took many samples, increased sample size, or did both, was the many samples part showing unbiasedness, and was the increased sample size part showing consistency? *Both!*
10. If there is extra time can you go over how to sample from a discrete dsn? Page 4 on the Simulating notes, we skipped over how to do it the long way, but I am curious.

## Discrete RV's $\rightarrow$ inverse pdf

1. Generate r.v.  $U \sim \text{Unif}(0,1)$
2. select  $x_i$  where  $F_x(x_{i-1}) < U \leq F_x(x_i)$

$$x \leftarrow 1:3$$
$$p \leftarrow [0.1, 0.2, 0.7]$$



# Monte Carlo Integration

1. Could we go over tips for finding  $g(x)$  and  $f(x)$  in monte carlo?
2. Can you go into more detail about the CLT's use in Monte Carlo integration, particularly why the R-calculated variance (i.e.  $\text{var}(\text{sample})$ ) of the Monte Carlo sample is different from the variance formula we used on our most recent homework?
3. Concerning Monte Carlo Integration:
  - a. Is the indicator method of Monte Carlo integration always best when we can't use the uniform distribution (because of an infinite support or whatever)? *not always, but usually easy to conceptualize.*
  - b. Is it valid to use transformation or accept-reject methods in order to sample from a density we are using in a Monte Carlo integral? Mainly I'm concerned about increased variance from the sampling making the estimated integral value less usable. *Case by case (depends on shape of  $d\sigma$  vs.  $h(x)$ ).*
  - c. I know it is okay to reuse the same samples (samples used in estimating the estimator) when estimating the variance of a Monte Carlo estimator, but is it equally okay to generate new samples? *Yes*
  - d. Slightly unrelated, is variance of the estimated variance a thing? Is variance of the estimated variance of the estimated variance a thing? *[Is this what moment generating functions do?]* *NO*
4. Can you review how to choose  $\phi(x)$  in importance sampling? I know we have certain requirements for  $\phi(x)$  for importance sampling to work but is there a good way to go about finding functions that work?
5. On page 18 in the importance sampling notes, it says "A rare draw from  $\phi$  with much higher density under  $f$  than under  $\phi$  will receive a huge weight and inflate the variance of the estimate". Would you be able to clarify this, perhaps using a visualization?
6. we could look into another example of importance sampling where the support goes to infinity, for some reason I have some trouble working with those, especially the derivation.

## ① Tips for finding $f$ and $g$ :

- once we choose  $f$ ,  $g$  is fixed (just have to derive it).

→ need to choose  $f$  so that  $\mathcal{X}$  contains limits of integration.

MC Integration of  $\int_a^b h(x)dx$

① select  $f \rightarrow g$   
so that

$$\int_a^b h(x)dx = E[g(X)],$$

$X \sim f$

$$\Rightarrow \int_a^b h(x)dx = \int_{\mathcal{X}} g(x)f(x)dx$$

$\mathcal{X} = \text{support of } f$

## ② Derive estimator

$\hat{\theta}$  estimates  $\theta = E[g(X)]$ .

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^m g(X_i), X_i \sim f$$

## ③ do it in R

sample  $X_i \sim f$

get  $\hat{\theta}$

② CLT:  $X_1, \dots, X_n \stackrel{iid}{\sim} f$  then  $\frac{\bar{X}_n - E\bar{X}_n}{\sqrt{\text{Var } \bar{X}_n}} \xrightarrow{d} N(0, 1)$ .

$$E\bar{X}_n = EX,$$

$$\text{Var } \bar{X}_n = \frac{1}{n} \text{Var } X$$

Monte Carlo  
Integration. ↓

$X_1, \dots, X_m \stackrel{iid}{\sim} f$

↓ transform  
 $g(X_1), \dots, g(X_n) \stackrel{iid}{\sim} f'$  CLT

// MC estimator.

$$\frac{\frac{1}{m} \sum_{i=1}^m g(X_i)}{\sqrt{\text{Var} \left[ \frac{1}{m} \sum_{i=1}^m g(X_i) \right]}} \xrightarrow{d} N(0, 1).$$

$\frac{1}{m} \text{Var } g(X_i)$

⇒ get CI's for  $\theta = \int h(x)dx$

$$\text{Var } \hat{\theta} = \frac{1}{m} \text{Var } g(X_i)$$

⇒ use MC integration to estimate.

R:  $\text{var}(\text{sample}) = \text{Var } X_i$

= if  $g = \text{identity}$ .

Q 4-5 requirements for  $\phi$ :  $f(x)/\phi(x)$  bounded  
 $\phi$  heavier tails than  $f \leftarrow$  good.  
 $\mathbb{X}_\phi \geq \mathbb{X}_f$

Usually perform importance sampling after we have already done MC Integration.

$\Rightarrow$  If  $\phi = f$  then this is the same as <sup>original</sup> MC integration.

$\Rightarrow$  plot  $h(x)$  over limits of integration or subst.  
 get a sense of shape.

large-ish  $\phi$  when  $h \downarrow$

look at  $h/\phi$  want close to constant.

Support requirements  
 → choose candidates  
 → make plots.

① MC Integration  
 → have  $f, g$   $\exists$  exist here.

$$\theta = E[g(X)] \times f$$

importance  
 ② pick  $\phi$

$$\theta = E\left[g(Y) \cdot \frac{f(Y)}{\phi(Y)}\right] \quad Y \sim \phi$$

## Other Questions

1. Will we ever have to prove why certain methods work or should we just know how to implement them?

NO, but if interested in doing these, let me know.