STAT400: Midterm Review

Fall 2019

Exam-related

1. What are the difference between distribution functions in R, such as `runif()` and `dunif()` and why we need to distinguish them?
2. Will we likely have to find our own equation to do importance sampling or will they be provided since we cannot graph them to see how they look with our $h(x)$?
3. If we need to know things like the median statistic, or a percentile of distributions will the equation for it be provided like they have been in the homework, or should we include those on our sheets?
4. For the accept-reject method, when we are finding an envelope, since we cannot graph will we have a way to decide what types of equations we can have or should we just try to have an idea of what each distribution sort of looks like?

1 dunif gives the density function
runif generates a sample from the Uniform dsn.

2 It's reasonable for you to choose a candidate, but I won't expect you to know how to plot them.

3 everything we have learned is fair game.

4 See 2.

Sampling:

1. Accept-reject vs. Inverse transform method of sampling.
2. How to identify the best envelope and envelope constant while building an accept-reject algorithm?

3. Homework 4, question 3:

   A discrete random variable has pmf

x	0.0	1.0	2.0	3.0	4.0
f	0.1	0.2	0.2	0.2	0.3

   Use the inverse transform method to generate a random sample of size 1000 from the distribution of X. Construct a relative frequency table and compare the empirical with the theoretical probabilities. Repeat using the `R` `sample` function.

We want to sample $X_1, \dots, X_n \overset{iid}{\sim} f$.

① $\boxed{\text{Accept Reject}}$:

1. Choose $g()$ (a density) and c a constant such that

$$c \cdot g(x) \geq f(x) \quad \forall\, x \in \mathbb{R}$$

(this implies $\{x : g(x) > 0\} \supseteq \{x : f(x) > 0\}$.

2. Sample $Y \sim g$ and $U \sim \text{Unif}[0,1]$.
3. If $U \leq f(Y)/c \cdot g(Y)$, accept Y.
4. repeat 1.-3. until you have accepted n values.

$\boxed{\text{Inverse Transform}}$:

CTS

1. Find $F^{-1}(u)$ by setting $F(x) = u$ and solving for x
2. Sample $U_1, \dots, U_n \sim \text{Unif}[0,1]$.
3. $X_i = F^{-1}(U_i)$ for $i = 1, \dots, n$.

Discrete

1. Generate $U \sim \text{Unif}[0,1]$
2. Find x_i where $F(x_{i-1}) < U \leq F(x_i)$, where $x_1 < \dots < x_{i-1} < x_i < \dots$ are ordered discontinuity pts.
3. Repeat 1.-2. n times.

3

x	$f(x) = P(X=x)$	$F(x) = P(X \le x)$
0	0.1	0.1
1	0.2	0.3
2	0.2	0.5
3	0.2	0.7
4	0.3	1

Plan:

1. $n = 1000$
2. Sample $U_1, \dots, U_n \sim \text{Unif}[0,1]$.
3. pick $x_j^{(i)}$ s.t. $F(x_{j-1}^{(i)}) < U_i \le F(x_j^{(i)})$ for each $i = 1, \dots, n$

Monte Carlo:

?

1. What are the 3 different Monte Carlo techniques?
2. What is the general order of steps for importance sampling?
3. Method 2 when using Monte Carlo integration from the standard normal cdf and want to estimate $\Phi(x)$?

   Estimating the cdf of a normal distribution. Use $m = 1000$.

$$\Phi(x) = \int_{-infty}^{0} \phi(t)dt + \int_{0}^{x} \phi(t)dt$$

   Let $Y \sim Unif(0, x)$.

4. Homework 6, question 4, part d

   Estimating the cdf of a normal distribution. Use $m = 1000$. For each method, compute a 95% confidence interval for $\Phi(2)$. Summarise your findings. Which CI is narrower and why does that matter?

5. Homework 6, question 3

   Develop two Monte Carlo integration approaches to estimate $\int_0^5 x^2 \exp(-x)dx$. (You must use different distributions in the two approaches). Check your answer using the `integrate()` function.

GOAL: want to evaluate $\theta = \int_a^b h(x)\,dx$.

2 $\boxed{\text{Step 1}}$ derive

a) Find f (a density) and g (a function) such that

$$\theta = E[g(X)] \text{ where } X \sim f.$$

b) If importance, choose ϕ (a density) w/ support that covers the support of f.

$$\Rightarrow \theta = E\left[\frac{g(X) \cdot f(X)}{\phi(X)}\right] \text{ where } X \sim \phi.$$

$\boxed{\text{STEP 2}}$ Plan.

Algorithm:

Monte Carlo

1. For $m = 1000$, Sample $X_1, \dots, X_m \overset{iid}{\sim} f$
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} g(X_i)$

w/ Importance Algorithm:

1. For $m = 1000$, Sample $X_1, \dots, X_m \overset{iid}{\sim} \phi$
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} \frac{g(X_i)\, f(X_i)}{\phi(X_i)}$.

$\boxed{\text{STEP 3}}$ write code.

3. Want to estimate $\theta = \frac{1}{2} + \int_0^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$

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

$\boxed{\text{STEP 1a}}$

$$\Rightarrow \int_0^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \int_0^x \underbrace{\frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, x}_{g(x)} \cdot \underbrace{\frac{1}{x}}_{f(x)}\, dy = E\left[\frac{x}{\sqrt{2\pi}} e^{-Y^2/2}\right]$$

$Y \sim \text{Unif}[0, x]$.

$\boxed{\text{STEP 2}}$

1. For $m = 1000$, sample $Y_1, \dots, Y_n \sim \text{Unif}[0, x]$.
2. evaluate $\hat{\underline{\Phi}}(x) = \frac{1}{2} + \frac{1}{m} \sum_{i=1}^{m} \left\{ \frac{x}{\sqrt{2\pi}} e^{-Y_i^2/2} \right\}$

4. 95% CI for $\hat{\theta}$: $\hat{\theta} \pm 1.96 \sqrt{\text{Var}(\hat{\theta})}$

So, for $\hat{\underline{\Phi}}(2)$, $\hat{\underline{\Phi}}(2) \pm 1.96 \sqrt{\hat{\text{Var}}(\hat{\underline{\Phi}}(2))}$

where $\text{Var}(\hat{\underline{\Phi}}(2)) = \frac{1}{m}\left[\frac{1}{n} \sum_{i=1}^{m} \left[g(X_i) - Eg(X)\right]^2\right]$, $X_i \overset{iid}{\sim} f$

don't know, use $\hat{\theta}$ instead.

⑤ $\theta = \int_0^5 x^2 e^{-x}\, dx$

Approach 1:

Let $f \sim \text{Unif}[0,5] \Rightarrow f(x) = \begin{cases} \frac{1}{5} & x \in [0,5] \\ 0 & \text{o.w.} \end{cases}$

$$\theta = \int_0^5 \underbrace{x^2 e^{-x} 5}_{g} \cdot \underbrace{\frac{1}{5}}_{f}\, dx = E\left[5X^2 e^{-X}\right], \quad X \sim \text{Unif}[0,5]$$

Plan:

1. Sample $X_1, \ldots, X_m \overset{iid}{\sim} \text{Unif}[0,5]$
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} 5 \cdot X_i^2 e^{-X_i}$

Approach 2:

Let $f \sim \exp(1) \Rightarrow f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{o.w.} \end{cases}$

$$\theta = \int_0^5 x^2 e^{-x}\, dx = \int_0^\infty \underbrace{x^2 I(x \leq 5)}_{g(x)} \underbrace{e^{-x}}_{f(x)}\, dx = E\left[X^2 I(X \leq 5)\right]$$

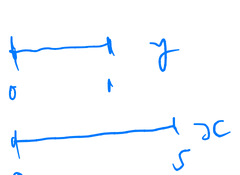
$X \sim \exp(1)$.

Plan:

1. Sample $X_1, \ldots, X_n \overset{iid}{\sim} \text{Exp}(1)$
2. $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} X_i^2 I(X_i \leq 5)$.

Approach 3:

Let $f \sim \text{Unif}[0,1]$. Then $\theta = \int_0^5 x^2 e^{-x}\, dx$, need change of variable.



$\frac{y-0}{1} = \frac{x-0}{5}$

$\Rightarrow 5y = x$

$5\,dy = dx$

$$= \int_0^1 (5y)^2 e^{-5y} \cdot 5\, dy$$

$$= E\left[5(5Y)^2 e^{-5Y}\right], \quad Y \sim \text{Unif}[0,1]$$

Plan: 1. Sample $Y_1, \ldots, Y_n \sim \text{Unif}[0,1]$

2. $\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} 5 (5Y_i)^2 e^{-5Y_i}$