Assignment 7.

Given April 19, due April 2.

Objective: Basic Monte Carlo.

Note that this assignment has many parts. Plan your work to maximize the modularity of your code, the independent testing of procedures, and code sharing. For example, a procedure that creates and plots a histogram and compares it with a theoretical density would be reused.

1. Use each of the following three techniques to generate 10^5 independent samples from the density $f(x) = xe^{-x}$. In each case, verify that your generator is correct by creating a histogram of the samples using a reasonable bin width and comparing the actual bin counts with the predicted counts. Make a plot.

a. If X_1 and X_2 are independent standard exponentials (density e^{-x}), then $X = X_1 + X_2$ has the density f. Do the convolution to show this.

b. Compute the cumulative distribution function $F(x)$. For a standard uniform random variable, T, get X by solving $F(X) = T$. You will have to do the solution numerically. Do this using Newton's method. Choose a heuristic initial guess that makes sense.

c. Generate X by rejection from an exponential with density $f_0(x) = \lambda e^{-\lambda x}$.

i. Why is it impossible to reject from the standard exponential ($\lambda = 1$)?

ii. Try to find the λ that maximizes the acceptance probability, either analytically (which is possible) or numerically.

2. Write a program that uses the Box Muller method to generate n iid $\mathcal{N}(0,1)$ random variables from n iid standard uniforms. Use the histogram method to check your work.

3. For some number l, the random variable Y is defined as follows. There are l iid Bernoullis $Z_1, \ldots, Z_l$, with $Pr(Z = 1) = p$ and $Pr(Z = 0) = 1 - p$. There are l standard normals $X_1, \ldots, X_l$. Then $Y = Z_1X_1 + \cdots + Z_lX_l$. If we think of the k with $Z_k = 1$ as hits, then we just add in a standard normal for each hit. We want to estimate $v = E(Y \mid Y \geq a)$.

a. Estimate v for $l = 5$, $p = .2$, $a = .5$ using vanilla Monte Carlo from the definition. Create an error bar. How many samples does it take to get 1% accuracy with 95% confidence? There are various ways to do this.

i. One is to use Bayes' rule and estimate $Pr(Y \geq a)$ also. This is simple but has the drawback that v is given by a ratio of two uncertain numbers so it's error bars are uncertain. How would you deal with this?

ii. Another way is to sample the random variable $Y \mid Y \geq a$ by generating Y samples and keeping only those greater than a. This is a rejection technique. To make the error bar problem simple, you can generate a fixed number of these so there is no uncertainty in the sample size.

b. (This is long, hard, and open ended. Attempt it only as time permits.) Estimate v for $p = .2$, $l = 30$, and as large values of a as possible, consistent with getting error bars that indicate that the uncertainty in v is small relative to v. You might try importance sampling where you replace p with $p' > p$ and $\mathcal{N}(0,1)$ with $\mathcal{N}(\mu,1)$ with $\mu > 0$. Use trial and error to try to find useful p' and μ values.