CS177, Homework 6
Due Date: Wednesday, May 21st

Please hand in a hardcopy of your homework at the beginning of class. Also upload your MATLAB functions to the appropriate EEE folder

Reading

Olofsson Section 2.3,2.4,2.7,4.3

Problem 1: Gambler's Ruin

In this problem you see how well the theoretical results we derived in class for the "Gambler's Ruin" problem match simulated results.

There are two players A and B that play a sequence of independent games (i.e. coin flips), betting 1 dollar in each round. The probability of player A winning an individual game is p (thus B wins with probability $1-p$) There are n dollars between the two players. Player A starts with i dollars and player B starts with the remaining $n-i$. The series ends when either player wins all the money.

Simulate the following four cases:

- p=0.5, i = 4, n = 20
- p=0.5, i = 10, n = 20
- p=0.55, i = 4, n = 20
- p=0.55, i = 10, n = 20

For each setting of the parameters, simulate 5,000 series (each series ending when A or B has all the money). For each of these series, record whether or not A wins and the length of the series. Based on these simulation results, compute and report on the (1) empirical estimate of the probability that A wins the series (2) the average length of the series and (3) the distribution of series lengths (you can display the distribution as a histogram using the `hist.m` function in MATLAB). In class, we gave an explicit formula for computing (1). Briefly comment on how your empirical estimate compares to this theoretical result we derived in class. What effect do you think i and p have on the distribution of series lengths?

NOTE: You are welcome and encouraged to use your code from last week for simulating Markov chains but you will need to modify it to deal with the absorbing states in the right way.

Problem 2: Variance

Let $f(x)$ be the probability density function for a real-valued variable X. In class we gave a formula for the variance of X, namely:

$$Var[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx$$

In class we showed that expectation is linear. Variance is not linear but it is handy to have a related formula. Please prove thatif a, b are constants then in general:

$$Var[aX + b] = a^2 Var(X)$$

Please show your work step-by-step.

Problem 3: Gaussian Distribution

Write a MATLAB function called `gauss_plot.m` which computes and plots the pdf of a Gaussian (Normal) random variable with parameters μ and σ. Your function should take three parameters:

- mu: the mean parameter for the pdf (a scalar)
- sigma: the standard deviation parameter for the pdf (a scalar)
- x: an `Nx1` vector of values where the pdf should be evaluated, e.g. x = [0.0,0.1,0.2,...20].

You should turn in

1. A copy of your documented MATLAB code for `gauss_plot.m`
2. Graphs of $f(x)$ for parameters $\mu = 100, \sigma = 10$ and $\mu = 100, \sigma = 30$. You should plot these with a 1000 values for x ranging from 50 to 150. You can easily generate such an x vector in MATLAB using $x = 50 : 0.1 : 150$
3. For each graph, manually shade in the region that corresponds to $P(110 \leq X \leq 120)$
4. For the two distributions, compute the probability $P(110 \leq X \leq 120)$. You can do this using either by modifying your code to numerically estimate the integral of $f(x)$ or by using the cdf table in the book appendix. In either case, please explain in detail how you calculated your answer.

Problem 4: Disk Usage

A lucky startup company has 1 million customers sign up to use their online photo storage service. A random customer i uses an amount of disk space X_i which obeys a Gaussian (normal) distribution with mean $\mu = 500$ Mbytes and standard deviation of $\sigma = 50$ Mbytes. Given this information, calculate the minimum amount of disk space the company should purchase if the company wants to be 99.9% sure that it will have disk space for all 1 million customers.

Problem 5: Central Limit Theorem

In this problem you will explore the Central Limit Theorem via simulation. You do not need to hand in the MATLAB scripts that you write for this problem.

1. Generate 100,000 uniform random numbers between 0 and 1 using the MATLAB function `rand.m`. You will probably want to pass parameters to `rand` so that it returns a whole vector of random numbers rather than writing a for loop (see `help rand` in MATLAB for more info). Use the `hist.m` function to plot a histogram of these numbers on the interval [0,1] with bins of width 0.01. Comment briefly on what you see and turn in the plot.

2. Let S_n be the sum of n random numbers where each number comes from the uniform distribution on $[0, 1]$. To study the properties of S_n you should generate K different sums. E.g., for $K = 10$ and $n = 10000$ you will compute 10 different sums where each sum is the sum of 10000 uniformly random numbers. You can do this efficiently in MATLAB by generating a `n x K` matrix of uniform random numbers and then summing the columns. Generate a set of values for each of the following choices of K and n

   - K=10, n = 100,000
   - K=100000 and n = 10
   - K=100 and n = 100
   - K = 10000 and n = 1000

   In each case

   - Calculate the theoretical mean of the K random sums (via the Central Limit Theorem) and also the empirical mean (the average of the numbers you generated). Comment on the differences between the two.
   - Plot a histogram of the values of S_n for the different cases and turn in the 4 histograms. Make sure and use an appropriate number of bins for each histogram (I'd suggest 5 bins for K=10, 10 bins for K=100, and 30 bins for the two largest Ks).
   - Comment on the shape of the histograms. Do they look like Normal distributions? If not, why not.