CS177, Homework 1
Due Date: Wednesday, April 9th

Reading

Reading for this week will review basic material on probability and discrete random variables. From Olofsson:

- Chapter 1, Sections 1.1-1.4: Review of basic notions and axioms for probability, sample spaces, events
- Chapter 2, Sections 2.1-2.2: Discrete random variables
- Chapter 2, Sections 2.4: Expectation and variance for discrete variables (i.e. p 97-102, 106-109)
- Chapter 2, Sections 2.5.1-2.5.2: Indicators and binomial distribution, useful for the homework

For now we will focus on discrete random variables in order to get quickly to interesting models and applications. We will revisit continuous variables and distributions later in the term.

MATLAB

Almost every assignment this quarter will involve some exercise in MATLAB so it is worth getting familiar during the first week. MATLAB is easy to learn and work with since it is interpreted and interactive. If you haven't used it before, spend an hour or two walking thru one of the tutorials linked from the course webpage and make sure to attend discussion section.

Submitting Homework

Please hand in hardcopies of solutions to all problems in class. If problems require MATLAB code, please include a printout of your code with your other solutions. Code should be clearly documented. If solutions ask for graphs, please make sure graphs have labels indicating which problem they go with. For MATLAB functions you are asked to write, please also submit the .m files electronically via the CS177 folder on EEE prior to the class period when they are due.

Problems

Please document clearly how you arrived at your answer. Solutions that skip intermediate work will lose points.

Problem 1: Sample Spaces

For each problem, define the sample space S and state its cardinality (finite,countably infinite, or uncountably infinite). If S is finite, indicate the number of elements in $|S|$

1. Toss 3 coins and count the number of tails.
2. Record the number of ants eaten by a contestant in an ant eating contest
3. Add up the years printed on 3 coins that were manufactured in 2000-2008
4. Toss a dart at a board which is 1 meter by 1 meter square
5. Count the number of web pages that have a single outgoing link

Problem 2: Events

In your sock drawer there are light and dark socks. Suppose you reach in and pull out some socks. Let L_k denote the event that the kth sock is light and D_k indicate that the kth sock is dark. Describe each of the events below in terms of L_k and D_k. For example, pulling out 2 socks, both of which are dark would be indicated by $D_1 \cap D_2$

1. You pull out three socks and at least one is dark
2. You pull out two socks and they happen to be matching colors
3. You pull out three socks and exactly one sock is dark
4. You pull out five socks and two successive socks are the same color
5. If you grab a single sock out of the drawer, the probability the sock is dark is 0.5 which we write as $P(D_1) = 0.5$. Compute the probability $P(L_1 \cap L_2 \cap L_3)$ of pulling out 3 light colored socks (Hint: assume there are enough socks in the drawer that removing a couple won't significantly affect the relative proportion of dark and light, that is $P(D_k) = 0.5$ for small values of k)
6. If you grab three socks, what is the probability there will be a matched pair?

Problem 3: Pair of socks paradox

Francis reasons as follows about socks. "If I pull out 3 socks, at least 2 will be alike. Suppose that they are dark. The third one will also be dark with probability $\frac{1}{2}$. Similarly, if the pair is light, the probability that the remaining one will be light is also $\frac{1}{2}$, so the probability of all three socks being the same color must be $\frac{1}{2}$." What is wrong with this argument? What is the actual probability of pulling out three socks of the same color?

Problem 4: Routing packets

Suppose packets are being routed from one computer to another and we have a choice of sending the packet over one of two networks. In network A, the packet takes $1, 2, 3$ or 4 hops with equal probability before getting to the destination. Each hop takes 2 milliseconds. In network B, the packet takes $7, 8$ or 9 hops but each hop only takes 1 millisecond. Let X be a random variable which represents the number of milliseconds it takes to transmit the packet from one computer to the other.

1. Suppose we send packets over network A, what is the probability mass function (pmf) of X?

2. What is the expected time $E[X]$ for a packet sent over network A? How about network B?

3. What is the variance $Var[X]$ for a packet sent over network A? What is the variance for network B?

4. If for each packet, we choose one of the two networks with equal probability, what is the pmf of X for this combined network?

5. The time it takes a packet to traverse a network is referred to as latency. Describe an application where having low expected network latency is important? Describe an application where having low variance in the network latency would be important? Which network (A,B,combined) would you choose in each case?

Problem 5: Lightning

Thanks to the pleasant weather, there is a one in a million chance each year that your computer is struck by lightning in Los Angeles.

1. If there are two million computers in LA, what is the probability than one is struck?

2. How many computers would there need to be in order for the probability to be $\frac{1}{2}$?

Problem 6: The binomial distribution (MATLAB)

Write a MATLAB function to plot the binomial distribution (binomial pmf) for each of the following parameters:

- n = 20, p = 0.1
- n = 20, p = 0.75
- n = 1000, p = 0.5
- n = 1000, p = 0.9

Generate a graph for each (4 graphs total) and comment briefly on the differences between the 4.

To do this you will need to write a MATLAB function `binomial_pmf.m`. A template is available on the class website which contains most of the function, but you will need to fill in the details.

For the cases with $n = 1000$ it is impractical to compute the binomial coefficients directly. Instead you can use the following approximation to the binomial which is accurate for large n:

$$p(i) \approx \frac{1}{\sqrt{2\pi np(1-p)}} e^{-(i-np)^2/(2np(1-p))}$$

You will find a function `binomial_pmf_approx.m` on the class website which implements this approximation. Use the approximate function for cases where $n = 1000$.

To generate nice plots in MATLAB, take a look at the script `plot_demo.m` on the website. You should familiarize yourself with the functions being called there and make sure you know how they work. You can access help for MATLAB functions by typing `help <functionname>` at the MATLAB prompt, e.g. `help bar` to learn more about the bar plot.

Problem 7: The binomial cdf (MATLAB)

Based on your code for the previous question, write a function called `binomial_cdf.m` that calculates the cumulative distribution function (cdf) for the binomial. This function will be invoked as: `y = binomial_cdf(n,p)`. The return value y should be a vector of length $(n + 1)$ containing the binomial cdf values evaluated from 0 up to n.

To generate the cdf, your function will first need to call `binomial_pmf.m` to get the values of the distribution. If $np > 5$ or $n(1 - p) > 5$ you should call the approximate function `binomial_pmf_approx.m` instead. The next step is to compute the partial sums from 0 up to i for each setting of i. You can do this with a for loop or the MATLAB function `cumsum` which should be faster.

Plot the cdf for the same 4 settings of n and p as in the previous problem.

Problem 8: Binomial failures

A hard-drive manufacturer sells 10,000 hard-drives each year. Each drive has an independent probability failure of $p = 0.1$ within the 1 year warranty period. Let N be the number of hard-drives that fail. When necessary, use the MATLAB functions you've written to compute answers to the following questions:

1. What is $E[N]$, the expected number of hard-drives that will have to be replaced in the 1 year period?
2. Compute the probability that 1200 or more of the original 10000 hard-drives fail?
3. Compute the probability that fewer than the expected number of drives fail? i.e. $P(N < E[N])$
4. Suppose the manufacturer can use components which double the price of the hard drive but also make it more reliable. How reliable (what value of p) would the new hard drives need to have in order for the manufacturer to break even, on average, if they make the switch?
5. What is the least reliable component (largest value of p) the manufacturer could use that would still guarantee the probability of more than 500 drives being returned would be less than 0.0001?