CS112: Computer System Modeling Fundamentals
Homework 2

Due Tuesday, April 26, 4pm (in class)

Please refer to the course academic integrity policy for collaboration rules, and be sure to make use of the problem solving guidelines covered in section. Remember that you will be graded on both the correctness and the clarity of your solutions. Start early!

1. Let N be an integer-valued random variable that takes on only positive values. Show that

$$\mathbf{E}[N] = \sum_{i=1}^{\infty} \mathbf{P}(N \geq i).$$

2. Let $X_1, ..., X_n$ be independent, identically distributed random variables with common mean and variance. Find the values of c and d that will make the following formula true:

$$\mathbf{E}[(X_1 + ... + X_n)^2] = c\mathbf{E}[(X_1)^2] + d(\mathbf{E}[X_1])^2$$

3. Your computer has been acting very strangely lately, and you suspect that it might have a virus on it. Unfortunately, all 12 of the different virus detection programs you own are outdated. You know that if your computer does have a virus, each of the programs, independently of the others, has a 0.8 chance of believing that your computer as infected, and a 0.2 chance of thinking your computer is fine. On the other hand, if your computer does not have a virus, each program has a 0.9 chance of believing that your computer is fine, and a 0.1 chance of wrongly thinking your computer is infected. Given that your computer has a 0.65 chance of being infected with some virus, and given that you will believe your virus protection programs only if 9 or more of them agree, find the probability that your detection programs will lead you to the right answer.

4. Alvin shops for probability books for K hours, where K is a random variable that is equally likely to be 1, 2, 3, or 4. The number of books N that he buys is random and depends on how long he shops according to the conditional PMF

$$p_{N|K}(n|k) = \frac{1}{k}, \quad \text{for } n = 1, ..., k.$$

   (a) Find the joint PMF of K and N .

   (b) Find the marginal PMF of N.

   (c) Find the conditional PMF of K given that $N = 2$.

   (d) Find the conditional mean and variance of K, given that he bought at least 2 but no more than 3 books.

   (e) The cost of each book is a random variable with mean \$30. What is the expected value of his total expenditure? Hint: Condition on the events $N = 1, ..., N = 4$, and use the total expectation theorem.

5. A particular binary data transmission and reception device is prone to some error when receiving data. Suppose that each bit is read correctly with probability p. Find a value of p such that when 100,000 bits are received, the expected number of errors is at most 850.

6. Suppose that you have designed a very cool personal website and you are interested in the types of visitors you have. To facilitate this, you have written a program to track the top-level domain (TLD) of each visitor to your website. For example, some visitors are students at other universities, so you get visitors with TLD of "edu"; some visitors are people in large companies who are interested in your resume, so you get visitors with TLD of "com".

   There are n different TLDs ("edu","com","org", etc.). Assume each visitor has exactly one TLD and TLDs for different visitors are mutually independent. In addition, assume it is equally likely for the next visitor of your website to be from any one of n TLDs. Find the expected number of visitors you need so that you have at least one visitor from each TLD.

   Hint: Let X be the number of visitors needed. Represent X as a sum of geometric random variables.

7. Consider n independent tosses of a die. Each toss has a probability $p_i > 0$ of resulting in i for $i = 1, 2, \ldots, 6$. Let X_i be the number of tosses that result in i. Show that X_1 and X_2 are negatively correlated (i.e., a large number of ones suggests a smaller number of twos).

8. A computer system consists of n subsystems, each of which has a lifetime which follows an exponential distribution with parameter λ_i, for $i = 1, 2, \ldots, n$. Each subsystem is independent but the whole computer system fails if any of the subsystems fails. Show that the lifetime of the computer system follows an exponential random variable with parameter $\sum_{i=1}^{n} \lambda_i$.

9. There are n hard disks. Let $X_1, \ldots, X_n$ be independent random variables representing the lifetime (in years) of the disks. Each X_i has a uniform distribution over $(0, 1)$ for $i = 1 \ldots n$. Let $M = max\{X_1, \ldots, X_n\}$, the random variable representing the lifetime of the hard disk that fails last. Show that the cumulative distribution function (CDF) of M is given by $F_M(x) = x^n$, $0 \leq x \leq 1$. What is the probability density function (pdf) of M?

10. This problem is about the performance of different scheduling schemes for read/write operations of a disk. Assume there are n independent requests for disk read/write which are uniformly distributed among all disk cylinders. For simplicity, we assume the requests follow a continuous uniform distribution over $(0, C)$, for some $C > 0$. Also we assume that initially, the read/write head of the disk is at position 0, and after serving these n read/writes the read/write head will move back to position 0.

    (a) The first scheme is to collect all the n requests, sort them according to location (cylinder), then serve the requests with one scan. With this scheme, the distance that the read/write head moves is 2 times the maximal distance from 0 among all n requests. Compute the expectation of this distance.

    (b) The second scheme is *first come first serve.* With this scheme, the distance that the read/write head moves consists of 3 parts:

        i. From location 0 to the location of the first job.

ii. From the location of the 1^{st} job to the 2^{nd} job, from the 2^{nd} to the 3^{rd}, $\ldots$, from the $n-1^{st}$ to the n^{th}.

iii. From the location of the n^{th} job back to location 0.

Compute the expectation of this distance (the sum of all 3 parts).

Think about which scheme you might prefer when n is large. (You do not have to write this down.)