Problem Set #4
Due: 2:30pm on Monday, May 9th

For each problem, **<u>explain/justify how you obtained your answer</u>** in order to obtain full credit. In fact, most of the credit for each problem will be given for the derivation/model used as opposed to the final answer. Make sure to describe the distribution and parameter values you used, where appropriate. It is fine for your answers to include summations, products, factorials, exponentials, or combinations, unless you are specifically asked for a computed numeric answer.

1. A web site historically receives requests at an average rate of 12 per minute, where each request is equally likely to come from a human or from a bot. Compute the conditional probability that at most 5 requests come from humans in a particular minute if 7 requests from bots are received in that same minute.

2. On average 5.5 users sign-up for an on-line social networking site each minute. What is the probability that:

   a. More than 7 users will sign-up for the social networking site in the next minute?

   b. More than 13 users will sign-up for the social networking site in the next 2 minutes?

   c. More than 15 users will sign-up for the social networking site in the next 3 minutes?

3. A robot is located at the <u>center</u> of a square world that is 10 kilometers on each side. A package is dropped off in the robot's world at a point (x, y) that is uniformly (continuously) distributed in the square. If the robot's starting location is designated to be (0, 0) and the robot can only move up/down/left/right parallel to the sides of the square, the distance the robot must travel to get to the package at point (x, y) is |x| + |y|. Let D = the distance the robot travels to get to the package. Compute E[D].

4. The joint probability density function of continuous random variables X and Y is given by:

   $$f_{X,Y}(x,y) = c\left(x^2 + \frac{xy}{2}\right) \quad \text{where } 0 < x < 1, \ 0 < y < 2$$

   a. What is the value of c in order for $f_{X,Y}(x,y)$ to be a valid probability density function?

   b. Are X and Y independent? Explain why or why not.

   c. What is the marginal density function of X?

   d. What is P(X > Y)?

   e. What is P(Y > 0.5 | X > 0.5)?

   f. What is E[X]?

5. The joint probability density function of continuous random variables X and Y is given by:

$$f_{X,Y}(x,y) = c\frac{y}{x} \quad \text{where} \quad 0 < y < x < 1$$

a. What is the value of c in order for $f_{X,Y}(x,y)$ to be a valid probability density function?

b. Are X and Y independent? Explain why or why not.

c. What is the marginal density function of X?

d. What is the marginal density function of Y?

e. What is E[X]?

f. What is E[Y]?

(Hint: At some point, integration by parts may be your friend on this problem.)

6. Say that of all the students who will ever take CS109, each will buy at most one textbook for the class. 60% will purchase the 9th Edition of Ross's textbook, 25% will purchase the 8th Edition of Ross's textbook, 5% will purchase a textbook not written by Ross, and remaining 10% will not buy any textbook at all. If 20 students are asked which, if any, textbook they purchased, what is the probability that exactly 10 students will have purchased the 9th Edition of Ross's textbook, 3 will have purchased the 8th Edition of Ross's textbook, 2 will have purchased textbooks not written by Ross, and the remaining 5 students will have not purchased any textbook at all?

7. Say we have two independent variables X and Y, such that X ~ Geo(p) and Y ~ Geo(p). Mathematically derive an expression for $P(X = k \mid X + Y = n)$.

8. Choose a number X at random from the set of numbers {1, 2, 3, 4, 5, 6}. Now choose a number at random from the subset no larger than X, that is from {1, ..., X}. Let Y denote the second number chosen.

a. Determine the joint probability mass function of X and Y.

b. Determine the conditional mass function $P(X = j \mid Y = i)$ as a function of i and j.

c. Are X and Y independent? Justify your answer.

9. Let X, Y, and Z be independent random variables, where X ~ $N(\mu_1, \sigma_1^2)$, Y ~ $N(\mu_2, \sigma_2^2)$, and Z ~ $N(\mu_3, \sigma_3^2)$. Let $W = aX - bY + c^2Z$, where a, b, and c are real-valued constants. What is the distribution (along with parameter values) for W? Show how you derived your answer.

10. Let X_i = the number of weekly visitors to a web site in week i, where X_i ~ N(2200, 52900) for all i. Assume that all X_i are independent of each other.

a. What is the probability that the total number of visitors to the web site in the next two weeks exceeds 5000.

b. What is the probability that the weekly number of visitors exceeds 2000 in at least 2 of the next 3 weeks?

11. Consider a series of strings that independently get hashed into a hash table. Each such string can be sent to any one of $k + 1$ buckets (numbered from 0 to k). Let index i denotes the i-th bucket. A string will independently get hashed to bucket i with probability p_i, where $\sum_{i=0}^{k} p_i = 1$. Let N denote the number of strings that are hashed until one is hashed to any bucket other than bucket 0. Let X be the number of that bucket (i.e. the bucket not numbered 0 that receives a string).

   a. Find $P(N = n)$, $n \geq 1$.

   b. Find $P(X = j)$, $j = 1, 2, \ldots, k$. *(Hint: Think conditional probability)*

   c. Prove that N and X are independent.

12. Let X_1, X_2, and X_3 be independently and identically distributed exponential random variables, all with the same parameter λ (i.e., X_1, X_2, and $X_3 \sim \text{Exp}(\lambda)$).

   a. What is $P(\min(X_1, X_2, X_3) \leq a)$?

   b. What is $P(\max(X_1, X_2, X_3) \leq a)$?

13. Say we have a coin with unknown probability X of coming up heads when flipped. However, we believe (subjectively) that the prior probability (before seeing the results of any flips of the coin) of X is a Beta distribution, where $E[X] = 0.5$ and $\text{Var}(X) = 1/36 \approx 0.02778$.

   a. What are the values of the parameters a and b (where a, $b > 1$) of the prior Beta distribution for X?

   b. Now say we flip the coin 13 times, obtaining 8 heads and 5 tails. What is the form (and parameters) of the posterior distribution of (X | 13 flips resulting in 8 heads and 5 tails)?

   c. What is $E[X \mid 12$ flips resulting in 8 heads and 4 tails]?

   d. What is $\text{Var}(X \mid 12$ flips resulting in 8 heads and 4 tails)?

14. Consider a bit string of length n, where each bit is independently generated and has probability p of being a 1. We say that a *bit switch* occurs whenever a bit differs from the one preceding it in the string (if there is a preceding bit). For example, if $n = 5$ and we have the bit string 11010, then there are 3 bit switches. Find the expected number of bit switches in a string of length n.

   (Hint: You might find it helpful to use a set of indicator variables that are defined in terms of whether a bit switch occurred in each *position* of the string. And in case you're wondering why we care about bit switches, the number of bit switches in a string can be one indicator of how compressible that string might be – for example, if the bit string represented a file that we were trying to ZIP. Also, remember back to the previous PSET problem on indicator variables.)

15. Say we have an array of n doubles, `arr[n]` (indexed from 0 to $n - 1$), which contains uniformly generated *non-negative* real values (where each value in the array is unique). What is the expected number of times that "max update" (as noted by the comment in the code) is executed in the function below (assuming the function is passed the array `arr` and its size `n`). Give an expression (not a big-Oh running time) for the expectation, and explain how you derived your answer.

```
double max(double arr[], int n) {
   double max = -1;         // note: all elements in arr[] are > -1.
   for(int i = 0; i < n; i++) {
      if (arr[i] > max) {
         max = arr[i];     // max update: (max = arr[i])
      }
   }
   return max;
}
```