CS112: Modeling Uncertainty in Information Systems
Homework 1

Due Monday, April 16, 2pm (in class)

Please refer to the course academic integrity policy for collaboration rules. Remember that you will be graded on both the correctness and the clarity of your solutions. You will not receive credit for a solution if the grader can't read your writing or understand your argument. Only a subset of the problems will be graded for credit. Show all of your work. Start early!

IMPORTANT: For each problem, be sure to clearly define any events that you use in your solution!!

1. Suppose that 40 percent of the email you receive is spam. With probability 0.3, these emails contain the word "buy" in the subject. Non-spam emails may also contain that keyword but with probability 0.01. Given that you receive an email containing "buy" in its subject, what is the probability that it is NOT a spam email?

2. The disc containing your favorite MP3 song on it has been scratched very badly. The disc was mixed up with two other scratched discs that were lying around. It is equally likely that any of the three discs holds your favorite song. A DJ friend of yours offers to take a look since he claims to recover songs from any disc with an 80% chance (assuming the song is there).

   (a) What is the probability that your friend will find the song if he searches all three discs?
   (b) Given that he searches on disc 1 but cannot recover your song, what is the probability that your song is on disc i for $i = 1, 2, 3$?
   (c) Given that he searches on disc 1 but cannot recover your song, what is the probability that he would be able to find the song by searching disc 2? (HINT: Given that the song is on a particular disc i, the event that the DJ finds it on disc 1 is independent of the event that the DJ finds it on disc 2.)
   (d) Now suppose your DJ friend has searched both disc 1 and disc 2 and has not yet found the song. What is the probability it is on disc i for $i = 1, 2, 3$? (HINT: You might need the hint from part (c))

3. We are told that events A and B are independent. In addition, events A and C are independent. Is it ALWAYS true that A is independent of $B \cup C$? Provide a proof or counterexample (A proof should be rigorous with each step justified, and a counterexample should include well-specified probabilities for any events of interest).

4. Let A and B be events such that $A \subset B$. Is it possible for A and B to be independent? Provide an example where they are, or prove that this cannot occur.

5. (a) Let X and Y be independent events. Prove that $P(X \cup Y) = P(X) + P(Y) - P(X)P(Y)$.
   (b) Use the result from part a, the algebra of sets, and the definition of independence to show that if events A, B, and C are independent, then A and $B \cup C$ are independent. Be sure to justify every step of your proof.

6. Suppose that you are trying to guess your friend's Gmail password. You know that it contains either 6 or 7 characters, and that each character it contains is either a lowercase letter $\{a, ..., z\}$ or a numerical digit $\{0, ..., 9\}$. As far as you know, all strings that satisfy these conditions are equally likely to be his password.

   For this problem, you may leave your answers in terms of arithmetic operations, such as addition, multiplication, and factorials (but not "choose" notation).

   (a) What is the probability that your friend's password contains all unique characters, i.e., that no character in the password is repeated?

   (b) You sit down at the computer and start guessing strings of length 6 or 7 containing only lowercase letters and digits uniformly at random and independently, without keeping track of the strings you have already guessed. What is the probability that you guess the correct password at least once in your first k tries?

   (c) Suppose your friend tells you that the password contains exactly two numerical digits. What is the conditional probability that both digits are 0?

7. A parking lot consists of a single row containing n parking spaces for some reasonably large n (say, $n \geq 5$). Mary arrives when all spaces are free. Tom is the next person to arrive. Each person makes an equally likely choice among all available spaces at the time of arrival.

   (a) Describe a natural sample space for this experiment mathematically. (Remember that a sample space is defined as a *set* of possible outcomes.)

   (b) Using the sample space that you defined, obtain the probability that the parking spaces selected by Mary and Tom are not directly next to each other

   (c) Obtain the probability that the parking spaces selected by Mary and Tom are at most 2 spaces apart (that is, have at most 1 empty space between them)