CS112: Modeling Uncertainty in Information Systems
Homework 2

Due Monday, April 30, 2pm (in class)

Please refer to the course academic integrity policy for collaboration rules. In particular, be sure to include a list of anyone with whom you have discussed the assignment. 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 or random variables that you use in your solution!!

1. Consider three discrete random variables, X, Y, and Z. Suppose you are given $p_X(x), p_{Y|X}(y|x)$, and $p_{Z|Y}(z|y)$ for all values x, y, and z that the random variables can take. For each part of this problem, state whether or not it is possible to compute the PMF with the information that is given. In cases where it is possible to compute the PMF, give an expression for the PMF in terms of $p_X, p_{Y|X}$, $p_{Z|Y}$, or other quantities that you have already derived. In cases where it is not possible, state what additional information you would need to compute the PMF.

   (a) $p_Y(y)$ (b) $p_Z(z)$ (c) $p_{X,Y}(x,y)$ (d) $p_{Y,Z}(y,z)$ (e) $p_{X,Z}(x,z)$
   (f) $p_{X|Y}(x|y)$ (g) $p_{X|Z}(x|z)$ (h) $p_{Y|Z}(y|z)$

2. Your computer has been acting very strangely lately, and you suspect that it might have a virus on it. Unfortunately, all 10 of the different virus detection programs you own are outdated.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 0.75 chance of believing that your computer as infected, and a 0.25 chance of thinking your computer is fine. On the other hand, if your computer does not have a virus, each program has a 0.65 chance of believing that your computer is fine, and a 0.35 chance of wrongly thinking your computer is infected. Given that your computer has a 0.8 chance of being infected with some virus, and given that you will believe your virus protection programs only if 8 or more of them agree, find the probability that your detection programs will lead you to the right answer.

3. Joe plays the lottery on any given week with probability p, independently of whether he played on any other week. Each time he plays, he has a probability q of winning, again independently of everything else. During a fixed time period of w weeks, let X be the number of weeks that he played the lottery and Y be the number of weeks that he won.

   You can leave your answers in terms of exponents, factorials, etc.

   (a) What is the probability that Joe played the lottery on any particular week, given that he did not win on that week?
   (b) Find the conditional PMF $p_{Y|X}(y|x)$.

(c) Find the joint PMF $p_{X,Y}(x, y)$. (It might help to start by expressing the joint PMF in terms of the conditional probabilities that you derived above.)

4. Alice and Bob play a "sudden-death" chess match. The match consists of at least one game. The first player to win a game wins the match. Each game is won by Alice with probability p, is won by Bob with probability q, and is a draw with probability 1 - p - q. The outcome of each game is independent of the other games.

   (a) What is the probability that Alice wins the match? (HINT: For any $c \in [0, 1]$, $\sum_{i=0}^{\infty} c^i = \frac{1}{1-c}$)
   (b) What is the PMF of the match length?

5. You have created a hot new web app. You decide to monetize your app by displaying ads. There are two types of ads that you could display, Type A or Type B. You receive 10 cents each time a user clicks on an ad of Type A, and 5 cents each time a user clicks on an ad of Type B. You may only show one ad to each user.

   If an older user is shown an ad of Type A, he will click on it with probability 0.3. He will click on an ad of Type B with probability 0.1. Younger users will click on an ad of Type A with probability 0.1, and will click on an ad of Type B with probability 0.4. Half of the users of your web app are young and half are old.

   (a) If you have to show the same type of ad to everyone, would you expect to make more money by showing ads of Type A or ads of Type B? To receive full credit, you must provide a justification for your answer. Be sure to clearly define any random variables or events that you use in your solution.
   (b) Suppose you have some demographic information about users of your app, and are able to choose an ad to display based on this information. Show that you can make even more money by showing the ads of Type A to the older users and ads of Type B to the younger users.

6. Consider tossing 2 fair coins (i.e., for each coin the probability of getting heads is 1/2). The outcome of each coin is independent. Define the following random variables:

   $A = 1$ if the first coin is heads, -1 if the first coin is tails.
   $B = 1$ if the second coin is heads, -1 if the second coin is tails
   $C = A + B$
   $D = A \times B$

   (a) What are the expected values of A, B, C, and D?
   (b) What are the variances of A, B, C, and D?
   (c) Are A and C independent? To receive full credit, you must provide a justification for your answer.
   (d) Are A and D independent? Again, to receive full credit, you must provide a justification for your answer.

7. Let X_1 and X_2 be independent, geometric random variables with parameters p_1 and p_2 respectively. You might find it useful to define $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. Find an expression in terms of p_1 and p_2 (and/or q_1 and q_2) for the probability that $X_1 \leq X_2$. **(Reduce your answer so it does not contain any summations.)** (HINT: see hint for 4.a)