Quiz Section #4: Supplementary Exercises

CSE 312: Foundations of Computing II

1. Questions about the Naive Bayes Classifier:
   (a) Naive Bayes assumes conditional independence of words in an email, given that we know the label (ham/spam) of the email. Why is that assumption necessary to make Naive Bayes work?
   (b) Is the conditional independence assumption actually true in the real world? That is, are the occurrences of words in an email independent of each other, if we know the label of the email? Explain.
   (c) Do you expect the Naive Bayes Classifier to correctly classify all emails in a test set? Explain why or why not.
   (d) If you were a spammer and you knew we used Naive Bayes to filter spam, how would you change your emails to try to get past the filter?

2. Let the random variable X be the sum of two independent rolls of a fair 3-sided die. (If you are having trouble imagining what that looks like, you can use a 6-sided die and change the numbers on 3 of its faces.)
   (a) What is the probability mass function of X?
   (b) Find $E[X]$ directly by applying the definition of expectation to the result from part (a).
   (c) Find $E[X]$ again using linearity of expectation.
   (d) Check that your answers to parts (b) and (c) are the same. Which way of computing the expectation was simpler, (a)+(b), or (c)?

3. Let the random variable X be the number of heads in n independent flips of a fair coin.
   (a) What is the probability mass function of X?
   (b) Find $E[X]$ directly by applying the definition of expectation to the result from part (a).
   Hint: prove and use the identity $i\binom{n}{i} = n\binom{n-1}{i-1}$.

4. You have 10 pairs of socks (so 20 socks in total), with each pair being a different color. You put them in the washing machine, but the washing machine eats 4 of the socks chosen at random. Every subset of 4 socks is equally probable to be the subset that gets eaten. Let X be the number of complete pairs of socks that you have left.
   (a) What is the probability mass function of X?
   (b) Find $E[X]$ directly by applying the definition of expectation to the result from part (a). Give your answer exactly as a simplified fraction.
   (c) Find $E[X]$ again using linearity of expectation. Give your answer exactly as a simplified fraction.
   (d) Check that your answers to parts (b) and (c) are the same. Which way of computing the expectation was simpler, (a)+(b), or (c)?

5. Find the expected number of bins that remain empty when m balls are distributed into n bins randomly and independently. For each ball, each bin has an equal probability of being chosen. (Notice that two bins being empty are not independent events: if one bin is empty, that decreases the probability that the second bin will also be empty. This is particularly obvious when $n = 2$ and $m > 0$.)

6. At a reception, n people give their hats to a hat-check person. When they leave, the hat-check person gives each of them a hat chosen at random from the hats that remain. What is the expected number of people who get their own hats back? (This is closely related to, but much simpler than, the challenge problem from the worksheet from quiz section #2. Notice that the hats returned to two people are not independent events: if a certain hat is returned to one person, it cannot also be returned to the other person.)

7. (This exercise is the same as Exercise 2, but with an ordinary 6-sided die rather than a 3-sided die.) Let the random variable X be the sum of two independent rolls of a fair 6-sided die.

   (a) What is the probability mass function of X?

   (b) Find $E[X]$ directly by applying the definition of expectation to the result from part (a).

   (c) Find $E[X]$ again using linearity of expectation.

   (d) Check that your answers to parts (b) and (c) are the same. Which way of computing the expectation was simpler, (a)+(b), or (c)?