Problem Set 11

Problem 1. **[20 points]** You are organizing a neighborhood census and instruct your census takers to knock on doors and note the sex of any child that answers the knock. Assume that there are two children in a household and that girls and boys are equally likely to be children and to open the door.

A sample space for this experiment has outcomes that are triples whose first element is either B or G for the sex of the elder child, likewise for the second element and the sex of the younger child, and whose third coordinate is E or Y indicating whether the *e*lder child or *y*ounger child opened the door. For example, (B, G, Y) is the outcome that the elder child is a boy, the younger child is a girl, and the girl opened the door.

(a) [5 pts] Let T be the event that the household has two girls, and O be the event that a girl opened the door. List the outcomes in T and O.

(b) [5 pts] What is the probability $\Pr(T \mid O)$, that both children are girls, given that a girl opened the door?

(c) [10 pts] Where is the mistake in the following argument for computing $\Pr(T \mid O)$?

> If a girl opens the door, then we know that there is at least one girl in the household. The probability that there is at least one girl is
>
> $$1 - \Pr(\text{both children are boys}) = 1 - (1/2 \times 1/2) = 3/4.$$
>
> So,
>
> $$\begin{aligned} &\Pr(T \mid \text{there is at least one girl in the household}) \\ &\quad = \frac{\Pr(T \cap \text{there is at least one girl in the household})}{\Pr\{\text{there is at least one girl in the household}\}} \\ &\quad = \frac{\Pr(T)}{\Pr\{\text{there is at least one girl in the household}\}} \\ &\quad = (1/4)/(3/4) = 1/3. \end{aligned}$$
>
> Therefore, given that a girl opened the door, the probability that there are two girls in the household is 1/3.

Problem 2. [20 points] Professor Plum, Mr. Green, and Miss Scarlet are all plotting to shoot Colonel Mustard. If one of these three has both an *opportunity* and the *revolver*, then that person shoots Colonel Mustard. Otherwise, Colonel Mustard escapes. Exactly one of the three has an *opportunity* with the following probabilities:

$$\Pr\{\text{Plum has opportunity}\} = 1/6$$
$$\Pr\{\text{Green has opportunity}\} = 2/6$$
$$\Pr\{\text{Scarlet has opportunity}\} = 3/6$$

Exactly one has the *revolver* with the following probabilities, regardless of who has an opportuntity:

$$\Pr\{\text{Plum has revolver}\} = 4/8$$
$$\Pr\{\text{Green has revolver}\} = 3/8$$
$$\Pr\{\text{Scarlet has revolver}\} = 1/8$$

(a) [5 pts] Draw a tree diagram for this problem. Indicate edge and outcome probabilities.

(b) [5 pts] What is the probability that Colonel Mustard is shot?

(c) [5 pts] What is the probability that Colonel Mustard is shot, given that Miss Scarlet does not have the revolver?

(d) [5 pts] What is the probability that Mr. Green had an opportunity, given that Colonel Mustard was shot?

Problem 3. [15 points] In lecture we discussed the Birthday Paradox. Namely, we found that in a group of m people with N possible birthdays, if $m \ll N$, then:

$$\Pr\{\text{all } m \text{ birthdays are different}\} \sim e^{-\frac{m(m-1)}{2N}}$$

To find the number of people, m, necessary for a half chance of a match, we set the probability to 1/2 to get:

$$m \sim \sqrt{(2\ln 2)N} \approx 1.18\sqrt{N}$$

For $N = 365$ days we found m to be 23.

We could also run a different experiment. As we put on the board the birthdays of the poeple surveyed, we could ask the class if anyone has the same birthday. In this case, before we reached a match amongst the surveyed people, we would already have found other people in the rest of the class who have the same birthday as someone already surveyed. Let's investigate why this is.

(a) [5 pts] Consider a group of m people with N possible birthdays amongst a larger class of k people, such that $m \leq k$. Define $\Pr\{A\}$ to be the probability that m people all have different birthdays *and* none of the other $k - m$ people have the same birthday as one of the m.

Show that, if $m \ll N$, then $\Pr\{A\} \sim e^{\frac{m(m-2k)}{2N}}$. (Notice that the probability of no match is $e^{-\frac{m^2}{2N}}$ when k is m, and it gets smaller as k gets larger.)

Hints: For $m \ll N$: $\frac{N!}{(N-m)!N^m} \sim e^{-\frac{m^2}{2N}}$, and $(1 - \frac{m}{N}) \sim e^{-\frac{m}{N}}$.

(b) [10 pts] Find the approximate number of people in the group, m, necessary for a half chance of a match (your answer will be in the form of a quadratic). Then simplify your answer to show that, as k gets large (such that $\sqrt{N} \ll k$), then $m \sim \frac{N \ln 2}{k}$.

Hint: For $x \ll 1$: $\sqrt{1-x} \sim (1 - \frac{x}{2})$.

Problem 4. [10 points] We're covering probability in 6.042 lecture one day, and you volunteer for one of Professor Leighton's demonstrations. He shows you a coin and says he'll bet you \$1 that the coin will come up heads. Now, you've been to lecture before and therefore suspect the coin is biased, such that the probability of a flip coming up heads, $\Pr\{H\}$, is p for $1/2 < p \leq 1$.

You call him out on this, and Professor Leighton offers you a deal. He'll allow you to come up with an algorithm using the biased coin to *simulate* a fair coin, such that the probability you win and he loses, $\Pr\{W\}$, is equal to the probability that he wins and you lose, $\Pr\{L\}$. You come up with the following algorithm:

1. Flip the coin twice.
2. Based on the results:
   - $TH \Rightarrow$ you win [W], and the game terminates.
   - $HT \Rightarrow$ Professor Leighton wins [L], and the game terminates.
   - $(HH \vee TT) \Rightarrow$ discard the result and flip again.
3. If at the end of N rounds nobody has won, declare a tie.

As an example, for $N = 3$, an outcome of HT would mean the game ends early and you lose, $HHTH$ would mean the game ends early and you win, and $HHTTTT$ would mean you play the full N rounds and result in a tie.

(a) [5 pts] Assume the flips are mutually independent. Show that $\Pr\{W\} = \Pr\{L\}$.

(b) [5 pts] Show that, if $p < 1$, the probability of a tie goes to 0 as N goes to infinity.

Problem 5. [20 points]

(a) [5 pts] Suppose A and B are *disjoint* events. Prove that A and B are *not independent*, unless $\Pr(A)$ or $\Pr(B)$ is zero.

(b) [5 pts] If A and B are independent, prove that A and $\bar{B}$ are also independent.
Hint: $\Pr\left(A \cap \bar{B}\right) = \Pr(A) - \Pr(A \cap B)$.

(c) [5 pts] Give an example of events A,B, C such that A is independent of B, A is independent of C, but A is not independent of $B \cup C$.

(d) [5 pts] Prove that if C is independent of A, and C is independent of B, and C is independent of $A \cap B$, then C is independent of $A \cup B$.
Hint: Calculate $\Pr(A \cup B \mid C)$.

Problem 6. [15 points] Three very rare DNA markers were found in the DNA collected at a crime scene. Only one in every $1,000$ people has marker A, one in every $3,000$ people has marker B, and one in every $5,000$ people has marker C. Joe the plumber was arrested and accused of committing the crime, because he had all those markers present in his DNA. The prosecutor argues that the chances of any person having all three DNA markers is

$$\frac{1}{1000} \cdot \frac{1}{3000} \cdot \frac{1}{5000} = \frac{1}{15,000,000,000}$$

, which is more than 1 over the number of people in the world. This, plus the fact that Joe the plumber lives only 100 miles away from the crime scene must clearly mean that he is guilty. Having taken 6.042, you should be suspicious of this reasoning.

(a) [2 pts] What assumption has the prosecutor made (even though he hasn't realized it) about the presence of the 3 markers in human DNA?

(b) [4 pts] What would be the probability of a person having all three markers assuming that the markers appear pairwise independently? Under this assumption, can it be stated with such certainty that Joe the plumber commited the crime?

(c) [4 pts] What can you say about the probability of a person having all three markers if there is no independence between the markers?

(d) [5 pts] In fact, it turns out that neither of the above assumptions is correct. A researcher from MIT (who was actually in your recitation section for 6.042 back in the day) has discovered that while markers B and C appear independently, the probability of having marker B if you have marker A is $\frac{1}{2}$ and the probability of having marker C if you have marker A is $\frac{1}{3}$. The defence attorney now argues that the probability of a randomly selected person having all three markers is

$$Pr[A \cap B \cap C] = Pr[A] \cdot Pr[B|A] \cdot Pr[C|A] = \frac{1}{1000} \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6,000}.$$

Called as a witness, the MIT researcher points out that this is not necessarily true and that in fact he himself does not know what the probability is. What is wrong with the defence attorney's reasoning? (We assume that the MIT researcher published correct information and that, since he took 6.042, he knows what he is talking about.)