Problem Set 10

Due: Monday, December 3

Problem 1. [**5 points**] *Finalphobia* is a rare disease in which the victim has the delusion that he or she is being subjected to an intense mathematical examination.

- A person selected uniformly at random has finalphobia with probability $1/100$.
- A person with finalphobia has shaky hands with probability $9/10$.
- A person without finalphobia has shaky hands with probability $1/20$.

What is the probablility that a person selected uniformly at random has finalphobia, given that he or she has shaky hands?

Problem 2. [**9 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 described by triples of letters: the first letter is either B or G for the sex of the elder child, likewise for the second letter denoting the sex of the younger child, and the third letter 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) [3 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) [3 pts] What is the probability $\Pr(T \mid O)$, that both children are girls, given that a girl opened the door?

(c) [3 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,

$$\begin{aligned}
&\Pr(T \mid \text{there is at least one girl in the household}) \\
&= \frac{\Pr(T \cap \text{there is at least one girl in the household})}{\Pr\{\text{there is at least one girl in the household}\}} \\
&= \frac{\Pr(T)}{\Pr\{\text{there is at least one girl in the household}\}} \\
&= (1/4)/(3/4) = 1/3.
\end{aligned}$$

Therefore, given that a girl opened the door, the probability that there are two girls in the household is 1/3.

Problem 3. [15 points] With the Birthday Paradox, we found that in a group of m people with N possible birthdays, if $m << N$, 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 matching birthday, 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 write the birthday of each student surveyed on the white board, we could ask the class if anyone has the same birthday; and it's very likely that someone in the class has the same birthday as a surveyed person before we could find two surveyed students with the same birthday. Let's investigate why this would be the case.

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

(b) [5 pts] Find m, the approximate number of people in the group 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 $\sqrt{N} << k$, $m \sim \frac{N \ln 2}{k}$.

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

(c) [5 pts] Suppose there are 100 people in a room. Assume that their birthdays are mutually independent and uniformly distributed. Again let A be the event that two people have the same birthday. As stated in the textbook, $\Pr(A) > 0.99$. Suppose we fix a particular person in the class—call her "Jane"—and then ask everyone in the room *except Jane* when their birthday is. Let B be the event that all of those 99 birthdays are different. What is the actual probability that Jane has the same birthday as another person in the room? In other words, what is $\Pr(A \mid B)$?

Problem 4. [16 points] Let A, B, and C be events of some experiment.

(a) [2 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) [2 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] Prove the following statement if it is true or give a counterexample if it is false: if A is independent of B, and A is independent of C, then A is independent of $B \cup C$.

(d) [5 pts] Prove the following statement if it is true or give a counterexample if it is false: ff A is independent of B, and A is independent of C, then A is independent of $B \cap C$.

(e) [2 pts] Prove that if C is independent of A, 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 5. [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 that he'll bet you \$1 on the coin coming up heads. Now, you've been to lecture before and therefore suspect the coin to be 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 of you winning and him losing, $\Pr\{W\}$, 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. Base the results on:
   - $TH \Rightarrow$ 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 by the end of N rounds nobody has won, declare a tie.

As an example, for $N = 3$, an outcome of HT would mean that the game ends early and you lose; $HHTH$ would mean that the game ends early and you win; and $HHTTTT$ would mean that 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 $0 < p < 1$, the probability of a tie goes to 0 as N goes to ∞.

Problem 6. [9 points] Three very rare DNA markers were found in the DNA collected at a crime scene. Only one in every $1,000$ persons has marker A, one in every $3,000$ persons has marker B, and one in every $5,000$ persons 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

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

She notes that it is more than 1 over the number of people in the world; plus the fact that Joe the plumber lives only 100 miles away from the crime scene, she impresses upon the jury that these numbers must clearly mean that Joe the plumber is guilty. Having taken 6.042, you are suspicious of this reasoning.

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

(b) [2 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 committed the crime?

(c) [2 pts] What can you say about the probability of a person having all three markers, if there is no independence among the markers?

(d) [3 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 defense 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 argument is not necessarily valid and that in fact he himself does not know what the probability is. What is wrong with the defense attorney's reasoning? (We assume that the MIT researcher published the correct information and that, since he took 6.042, he knows what he is talking about.)

Problem 7. [14 points] Independently flip three fair coins (with "fair" meaning "equally likely to come up with a head or a tail").

- Let H_i be the indicator variable for a head occurring on the ith flip, for $i = 1, 2, 3$;
- C be the random variable for the number of heads flipped, $H_1 + H_2 + H_3$;
- M be the indicator variable for the event that all three coins match: $[H_1 = H_2 = H_3]$;
- and S be the indicator variable for the event that an odd number of heads are flipped, $[C \equiv 1 \pmod 2]$.

(a) [2 pts] Show that none of these six variables is independent of C.

Hint: Consider the case when $C = 3$.

(b) [4 pts] Show that M and S are pairwise independent.

(c) [8 pts] Show that H_1, H_2, H_3, and S are 3-wise independent, but not mutually independent.

Problem 8. [15 points] An over-caffeinated sailor of Tech Dinghy wanders along Seaside Boulevard, which conveniently consists of the points along the x axis with integral coordinates. In each step, the sailor moves one unit left or right along the x axis. A particular ***path*** taken by the sailor can be described by a sequence of "left" and "right" steps. For example, $\langle left, left, right \rangle$ describes the walk that goes left twice then goes right.

We model this scenario with a random walk graph whose vertices are the integers and with edges going in each direction between consecutive integers. All edges are labelled 1/2.

The sailor begins his random walk at the origin. This is described by an initial distribution that labels the origin with probability 1 and all other vertices with probability 0. After one step, the sailor is equally likely to be at location 1 or -1, so the distribution after one step gives label 1/2 to the vertices 1 and -1 and labels all other vertices with probability 0.

(a) [5 pts] Give the distributions after the 2nd, 3rd, and 4th step by filling in the table of probabilities below, where omitted entries are 0. For each row, write all the nonzero entries so they have the same denominator.

	location								
	-4	-3	-2	-1	0	1	2	3	4
initially					1				
after 1 step				1/2	0	1/2			
after 2 steps			?	?	?	?	?		
after 3 steps		?	?	?	?	?	?	?	
after 4 steps	?	?	?	?	?	?	?	?	?

(b) [5 pts]

1. What is the final location of a t-step path that moves right exactly i times?
2. How many different paths are there that end at that location?
3. What is the probability that the sailor ends at this location?

(c) [5 pts] Let L be the random variable giving the sailor's location after t steps, and let $B ::= (L+t)/2$. Use the answer to part (b) to show that B has an unbiased binomial density function.

Problem 9. [7 points] We have two coins: one is a fair coin and the other is a coin that produces heads with probability $\frac{3}{4}$. One of the two coins is picked, and this coin is tossed n times. Explain how to calculate the number of tosses to make us 95% confident of which coin was chosen. You do not have to calculate the minimum value of n, though we'd be pleased if you try.