Problem Set 12

Due: Monday, December 5, 7:30 pm

Reading Assignment: Sections 16.1 - 16.5

Problem 1. **[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 2. **[10 points]** Suppose that events A, B, C, D are mutually independent. Prove whether each of the following statements is true or give a counterexample if it is false.

(a) [5 pts] $A \cap C$ is independent of $\bar{B} \cap D$ where $\bar{B}$ denotes the complement of B.

(b) [5 pts] $A \cap \bar{B} \cap C$ is independent of $\bar{B} \cup \bar{C}$.

Problem 3. [20 points]

You roll 2 five-sided die. The sides of each die are numbered from 1 to 5. All sides of each die have an equal probability of appearing, and the two die rolls are independent. Let event A be that sum of the results of the rolls of the two die is 10, and let B be the event that the sum of the results of the rolls of the two die is 8.

(a) [5 pts] Is A independent of the event that at least one of the die rolls resulted in a 5?

(b) [5 pts] Is A independent of the event that at least one of the die rolls resulted in a 3?

(c) [5 pts] Is B independent of the event that both die rolls resulted in the same number?

(d) [5 pts] Is B independent of the event that at least one of the die rolled a 3?

Problem 4. [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 5. [20 points]

Professor Moitra has a deck of 52 randomly shuffled playing cards, 26 red, 26 black. He proposes the following game: he will continually draw a card off the top of the deck, turn it face up so that you can see it and then put it aside. At any point while there are still cards left in the deck, you may say "stop" and he will flip over one last card. If that next card turns up black you win and otherwise you lose. Either way, the game ends.

(a) [4 pts] Show that if you say "stop" before you have seen any cards, you then have probability $1/2$ of winning the game.

(b) [4 pts] Suppose you don't say "stop" before the first card is flipped and it turns up red. Show that you then have a probability of winning the game that is greater than $1/2$.

(c) [4 pts] If there are r red cards left in the deck and b black cards, show that the probability of winning if you say "stop" before the next card is flipped is $b/(r+b)$.

(d) [8 pts] Either,

1. come up with a strategy for this game that gives you a probability of winning strictly greater than $1/2$ and prove that the strategy works, or,
2. come up with a proof that no such strategy can exist.

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 defense 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.)