Problem Set 12 Solutions

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 $(1 - \frac{m}{N}) \sim e^{-\frac{m}{N}}$.

Solution. We know:

$$\Pr\{A\} = \frac{N(N-1)\ldots(N-m+1) \cdot (N-m)^{k-m}}{N^k}$$

since there are N choices for the first birthday, $N - 1$ choices for the second birthday, etc., for the first m birthdays, and $N - m$ choices for each of the remaining $k - m$ birthdays. There are total N^k possible combinations of birthdays within the class.

$$\Pr\{A\} = \frac{N(N-1)\ldots(N-m+1)\cdot(N-m)^{k-m}}{N^k}$$

$$= \frac{N!}{(N-m)!}\left(\frac{(N-m)^{k-m}}{N^k}\right)$$

$$= \frac{N!}{(N-m)!N^m}\left(\frac{N-m}{N}\right)^{k-m}$$

$$= \frac{N!}{(N-m)!N^m}\left(1-\frac{m}{N}\right)^{k-m}$$

$$\sim e^{-\frac{m^2}{2N}}\cdot e^{-\frac{m}{N}(k-m)} \qquad\qquad \text{(by the Hint)}$$

$$= e^{\frac{m(m-2k)}{2N}}$$

∎

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

Solution. Setting $\Pr\{A\} = 1/2$, we get a solution for m:

$$1/2 = e^{\frac{m(m-2k)}{2N}}$$

$$-2N\ln 2 = m^2 - 2km$$

$$0 = m^2 - 2km + (2N\ln 2)$$

$$m = \frac{2k \pm \sqrt{(2k)^2 - 4(2N\ln 2)}}{2}$$

Simplifying the solution under the assumption of large k, we find:

$$m = \frac{2k - \sqrt{4k^2 - 8N\ln 2}}{2} \qquad\qquad \text{(taking the lower positive root)}$$

$$= k - k\sqrt{1 - \frac{2N\ln 2}{k^2}}$$

$$\sim k - k\left(1 - \frac{2N\ln 2}{2k^2}\right) \qquad\qquad \text{(by the Hint)}$$

$$= \frac{N\ln 2}{k}$$

∎

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.

Solution. This statement is true. Intuitively, knowing that events A and C occurred should tell you nothing about events $\bar{B}$ and D occurring. More formally,

$$p(A \cap \bar{B} \cap C \cap D) + p(A \cap B \cap C \cap D) = p(A \cap C \cap D)$$
$$\Rightarrow p(A \cap \bar{B} \cap C \cap D) = p(A) \cdot p(C) \cdot p(D)(1 - p(B))$$
$$\Rightarrow p(A \cap \bar{B} \cap C \cap D) = p(A \cap C) \cdot p(\bar{B} \cap D)$$

∎

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

Solution. This statement is false. Intuitively, knowing that event $A \cap \bar{B} \cap C$ occurred implies that C occurred and so this influences the probability that $\bar{B} \cup \bar{C}$ occurs. As a counterexample, consider 4 independent coin tosses, and let A be the event that the first coin is a heads, B the event that the second coin is a heads, C be the event that the third coin is a heads, and D be the event that the fourth coin is a heads. Now $p(A \cap \bar{B} \cap C) = \frac{1}{8}$ and $p(\bar{B} \cup \bar{C}) = 1 - \frac{1}{4} = \frac{3}{4}$. Finally,

$$p((A \cap \bar{B} \cap C) \cap (\bar{B} \cup \bar{C})) = p((A \cap \bar{B} \cap C \cap \bar{B}) \cup (A \cap \bar{B} \cap C \cap \bar{C}))$$
$$= p(A \cap \bar{B} \cap C)$$
$$= \frac{1}{8}$$

Now since $p((A \cap \bar{B} \cap C) \cap (\bar{B} \cup \bar{C})) \neq p(A \cap \bar{B} \cap C) \cdot p(\bar{B} \cup \bar{C})$, these events are not independent.

∎

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?

Solution. No. $p(A) = \frac{1}{25}$. Let F be the event that at least one of the die rolls resulted in a 5. Then $p(F) = \frac{5}{25} + \frac{5}{25} - \frac{1}{25} = \frac{9}{25}$. Finally $p(A \cap F) = \frac{1}{25}$ since if event A happens then both die rolls must have resulted in a 5. Now as $p(A \cap F) \neq p(A) \cdot p(F)$, the events are not independent. ∎

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

Solution. No. Again $p(A) = \frac{1}{25}$. Now let T be the event that at least one of the die rolls resulted in a 3. $p(T) = \frac{9}{25}$ as before. However $p(A \cap T) = 0$ since both die rolls need to result in a 5 for event A to happen. ∎

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

Solution. No. $p(B) = \frac{3}{25}$. Let S be the event that both die rolls resulted in the same number. Then $p(S) = \frac{5}{25}$. Now $p(B \cap S) = \frac{1}{25}$, which is not equal to $\frac{3}{25} \cdot \frac{5}{25}$. ∎

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

Solution. No. Again $p(B) = \frac{3}{25}$. Now let T be the event that at least one of the die rolled a 3. Then $p(T) = \frac{9}{25}$. Now, $p(B \cap T) = \frac{2}{25}$, which is not equal to $\frac{3}{25} \cdot \frac{9}{25}$. ∎

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.

Solution. Since A and B are disjoint,

$$\Pr(A \cap B) = \Pr(\emptyset) = 0.$$

So, $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$ iff $\Pr(A) = 0$ or $\Pr(B) = 0$. ∎

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

Solution.

$$
\begin{aligned}
\Pr(A \cap \bar{B}) &= \Pr(A) - \Pr(A \cap B) && \text{(by the hint)} \\
&= \Pr(A) - \Pr(A) \cdot \Pr(B) && \text{(since } A \text{ and } B \text{ are independent)} \\
&= \Pr(A) \cdot (1 - \Pr(B)) \\
&= \Pr(A) \cdot \Pr(\bar{B}).
\end{aligned}
$$

The last equality holds since the probability of any event equals 1 minus the probability of its complement. Thus, we have shown that $\Pr(A \cap \bar{B}) = \Pr(A) \cdot \Pr(\bar{B})$, which is equivalent to A and $\bar{B}$ being independent. ∎

(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 $B \cup C$.

Solution. The experiment is 2 independent coin flips, letting A be "the 1st flip is heads", B the "the 2nd flip is heads," C is "odd number of heads." Then A is not independent of $B \cup C$ because

$$\Pr\left(A \mid B \cup C\right) = \frac{\Pr\left(A \cap (B \cup C)\right)}{\Pr\left(B \cup C\right)} = \frac{\Pr\left(HH, HT\right)}{\Pr\left(HH, TH, HT\right)} = 2/3 \neq 1/2 = \Pr\left(A\right).$$

∎

(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 $\Pr\left(A \cup B \mid C\right)$.

Solution. Conditional inclusion-exclusion followed by plain inclusion-exclusion provides a quick proof:

$$
\begin{aligned}
\Pr\left(A \cup B \mid C\right) &= \Pr\left(A \mid C\right) + \Pr\left(B \mid C\right) - \Pr\left(A \cap B \mid C\right) &&\text{(by conditional inc-ex)}\\
&= \Pr\left(A\right) + \Pr\left(B\right) - \Pr\left(A \cap B\right) &&\text{(by independence)}\\
&= \Pr\left(A \cup B\right) &&\text{(by regular inc-ex)}
\end{aligned}
$$

∎

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 and 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 choose to stop and he will flip over one last card. If the next card turns up black you win and otherwise you lose. Either way, the game ends.

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

Solution. If we just record the sequence of black and red cards that will be drawn, there are $\binom{51}{25}$ sequences with first card black: 25 positions for the black cards chosen from the 51 remaining positions. Since there are $\binom{52}{26}$ sequences in all, the probability of winning on the first draw is $\binom{51}{25}/\binom{52}{26} = 26/52 = 1/2$. ∎

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

Solution. Suppose you take the next card after that. There are $\binom{50}{25}$ sequences that start with a red card and then a black and there are $\binom{51}{26}$ sequences that start with a red card. So then there is a $\binom{50}{25}/\binom{50}{26} = 26/51 > 1/2$ chance of winning. Any optimum strategy would have to guarantee a probability of winning as least as big as that. ∎

(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 choose to stop before the next card is flipped is $b/(r+b)$.

**Solution.** The probability is $\binom{b+r-1}{b-1}/\binom{b+r}{b} = b/(r+b)$. ∎

(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.

Solution. There is no such strategy. Let $S_{b,r}$ be a strategy that acchieves the best probability of winning when starting with b black cards and r red cards. The claim is that $\Pr(\text{win by playing } S_{b,r}) = b/(r+b)$ for all b, r with at least $b > 0$ or $r > 0$.

Clearly $\Pr(\text{win by playing } S_{1,0}) = 1$ and $\Pr(\text{win by playing } S_{0,1}) = 0$. We prove the rest of the claim by induction on $r + b$. If the strategy $S_{b,r}$ is to take the next card, then $\Pr(\text{win by playing } S_{b,r}) = b/(r+b)$ as claimed. Suppose then that the strategy $S_{r,b}$ is to not take the first card, but to keep playing. Then the first card is going to be red with probability $r/(r+b)$ and it's going to be black with probability $b/(r+b)$. If the first card is black then we now have $b - 1$ black cards remaining and r red cards remaining and we should play $S_{b-1,r}$ which by our inductive hypothesis has a probability of $(b-1)/(b+r-1)$ of winning. On the other hand, if the first card is red, we now have $r - 1$ red cards and b black cards remaining and we should play $S_{b,r-1}$ which by our inductive hypothesis has a probability of $b/(b+r-1)$ of winning. This gives us a tot al probability of

$$\Pr(\text{win by playing } S_{b,r}) = ((b-1)/(b-1+r))(b/(r+b)) + (b/(b+r-1))(r/(b+r)) = b/(b+r),$$

as claimed. This is summarize d in the diagram below.

■
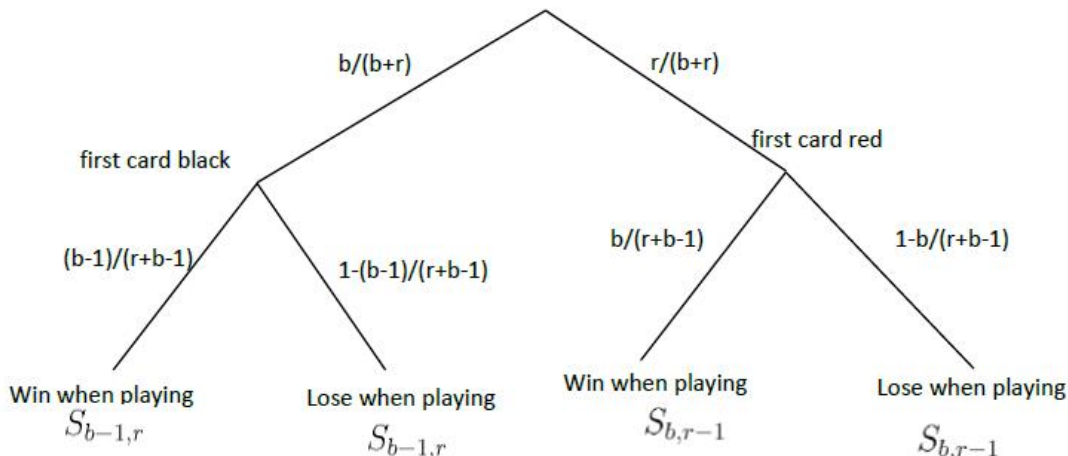
Problem 6. [15 points] Three very rare DNA markers were found in the DNA collected at a crime scene. Only one in every $1,000$ people has marker A, one in every $3,000$ people has marker B, and one in every $5,000$ people has marker C. Joe the plumber was arrested and accused of committing the crime, 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?

Solution. Mutual independence. ■

(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?

Solution.

$$Pr[A \cap B \cap C] \le Pr[B \cap C] = Pr[B] \cdot Pr[C] = \frac{1}{15,000,000}$$

One in $15,000,000$ people might have all the three markers, and the fact that Joe the plumber lives 100 miles away means nothing if the crime was committed in, say, New York City. ■

(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?

Solution. The probability is at most

$$P[C] = \frac{1}{5,000}.$$

■

(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.)

Solution. The researcher said that markers B and C appear independently, but he never said that they are independent if the person has marker A. ∎