

## Problem Set 11 Solutions

**Due:** Monday, November 22, 7pm

**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 elder child or younger 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  $O$ .

**Solution.**  $T = \{GGE, GGY\}$ ,  $O = \{GGE, GGY, GBE, BGY\}$  ■

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

**Solution.**  $1/2$  ■

(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,

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

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

**Solution.** The argument is a correct proof that

$$\Pr(T \mid \text{there is at least one girl in the household}) = 1/3.$$

The problem is that the event,  $H$ , that the household has at least one girl, namely,

$$H = \{GGE, GGY, GBE, GBY, BGE, BGY\},$$

is not equal to the event,  $O$ , that a girl opens the door. These two events differ:

$$H - O = \{BGE, GBY\},$$

and their probabilities are different. So the fallacy is in the final conclusion where the value of  $\Pr(T \mid H)$  is taken to be the same as the value  $\Pr(T \mid O)$ . Actually,  $\Pr(T \mid O) = 1/2$ .

■

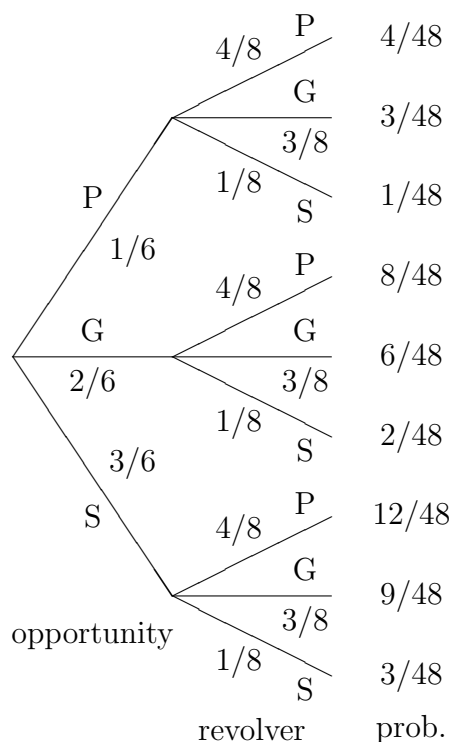
**Problem 2. [20 points]** Professor Plum, Mr. Green, and Miss Scarlet are all plotting to shoot Colonel Mustard. If one of these three has both an *opportunity* and the *revolver*, then that person shoots Colonel Mustard. Otherwise, Colonel Mustard escapes. Exactly one of the three has an *opportunity*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

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

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

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

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



**Solution.**

■

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

**Solution.** Denote each outcome with a pair indicating who has the opportunity and who has the revolver. In this notation, the event that Colonel Mustard is shot consists of all outcomes where a single person has both:

$$\{(P, P), (G, G), (S, S)\}$$

The probability of this event is the sum of the outcome probabilities:

$$\begin{aligned} \Pr \{ \{(P, P), (G, G), (S, S)\} \} &= \Pr \{ (P, P) \} + \Pr \{ (G, G) \} + \Pr \{ (S, S) \} \\ &= 4/48 + 6/48 + 3/48 \\ &= 13/48 \end{aligned}$$

■

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

**Solution.** Let  $S$  be the event that Colonel Mustard is shot, and let  $N$  be the event that

Miss Scarlet does *not* have the revolver. The solution is:

$$\begin{aligned}
 \Pr(S \mid N) &= \frac{\Pr\{S \cap N\}}{\Pr\{N\}} \\
 &= \frac{\Pr\{(P, P), (G, G)\}}{\Pr\{(P, P), (P, G), (G, P), (G, G), (S, P), (S, G)\}} \\
 &= \frac{\frac{4}{48} + \frac{6}{48}}{\frac{4}{48} + \frac{3}{48} + \frac{8}{48} + \frac{6}{48} + \frac{12}{48} + \frac{9}{48}} \\
 &= \frac{5}{21}
 \end{aligned}$$

■

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

**Solution.** Let  $G$  be the event that Mr. Green has an opportunity, and again let  $S$  be the event that Colonel Mustard is shot. Then the solution is:

$$\begin{aligned}
 \Pr(G \mid S) &= \frac{\Pr\{G \cap S\}}{\Pr\{S\}} \\
 &= \frac{\Pr\{(G, G)\}}{\Pr\{(P, P), (G, G), (S, S)\}} \\
 &= \frac{\frac{6}{48}}{\frac{4}{48} + \frac{6}{48} + \frac{3}{48}} \\
 &= \frac{6}{13}
 \end{aligned}$$

■

**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 people 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:

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

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.

$$\begin{aligned} \Pr\{A\} &= \frac{N(N-1)\dots(N-m+1) \cdot (N-m)^{k-m}}{N^k} \\ &= \frac{N!}{(N-m)!} \left( \frac{(N-m)^{k-m}}{N^k} \right) \\ &= \frac{N!}{(N-m)!N^m} \left( \frac{N-m}{N} \right)^{k-m} \\ &= \frac{N!}{(N-m)!N^m} \left( 1 - \frac{m}{N} \right)^{k-m} \\ &\sim e^{-\frac{m^2}{2N}} \cdot e^{-\frac{m}{N}(k-m)} && \text{(by the Hint)} \\ &= e^{\frac{m(m-2k)}{2N}} \end{aligned}$$

■

(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  $m$ :

$$\begin{aligned} 1/2 &= e^{\frac{m(m-2k)}{2N}} \\ -2N \ln 2 &= m^2 - 2km \\ 0 &= m^2 - 2km + (2N \ln 2) \\ m &= \frac{2k \pm \sqrt{(2k)^2 - 4(2N \ln 2)}}{2} \end{aligned}$$

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

$$\begin{aligned}
 m &= \frac{2k - \sqrt{4k^2 - 8N \ln 2}}{2} && \text{(taking the lower positive root)} \\
 &= k - k \sqrt{1 - \frac{2N \ln 2}{k^2}} \\
 &\sim k - k \left(1 - \frac{2N \ln 2}{2k^2}\right) && \text{(by the Hint)} \\
 &= \frac{N \ln 2}{k}
 \end{aligned}$$

■

**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  $\Pr\{W\} = \Pr\{L\}$ .

**Solution.** The probability of you winning is equal to the probability that you win in the first round, plus the probability that nobody won in the first round times the probability that you win in the second round, plus the probability that nobody won in the first two

round times the probability that you win in the third round, etc. The same goes for Professor Leighton. Hence:

$$\begin{aligned}\Pr\{W\} &= \Pr\{TH\} + \Pr\{HH \vee TT\} \Pr\{TH\} + \Pr\{HH \vee TT\}^2 \Pr\{TH\} + \dots \\ &= \Pr\{TH\} \cdot \sum_{i=0}^N \Pr\{HH \vee TT\}^i \\ &= \Pr\{HT\} \cdot \sum_{i=0}^N \Pr\{HH \vee TT\}^i \\ &= \Pr\{L\}\end{aligned}$$

The middle step is possible because  $\Pr\{TH\} = (1-p)p = p(1-p) = \Pr\{HT\}$ . ■

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

**Solution.** The probability of a tie is just the probability that nobody won all  $N$  rounds, namely:

$$\Pr\{tie\} = (\Pr\{HH \vee TT\})^N = (\Pr\{HH\} + \Pr\{TT\})^N = (p^2 + (1-p)^2)^N$$

So the limit as  $N$  goes to infinity is 0, given that  $p$  and therefore  $p^2 + (1-p)^2$  are  $< 1$ . ■

### 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.

**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

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

■

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

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

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

■

**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 committed the crime?



**Solution.**

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

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. ■