Problem Set 11 Solutions

Due: Tuesday, November 25, 7pm

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, and by luck we only had to survey about 14 people before we found a match.

However, before we reached a match amongst the surveyed people, we had already found other people in the rest of the class who had 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 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 2. [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 3. [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: $\Pr\left(A \cap \bar{B}\right) = \Pr(A) - \Pr(A \cap B)$.

Solution.

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

The last equality holds since the probability of any event equals 1 minus the probability of its complement. Thus, we have shown that $\Pr\left(A \cap \bar{B}\right) = \Pr(A) \cdot \Pr\left(\bar{B}\right)$, 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 4. [20 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 to 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, $[C \equiv 1 \mod 2]$.

(a) [5 pts] Show that none of these six variables is independent of C.

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

Solution. If $C = 3$, then the values of all the other variables are determined, namely $H_1 = H_2 = H_3 = M = S = 1$.

Therefore, for all six variables V, $\Pr(V = 1 \mid C = 3) = 1$, but $\Pr(V = 1) \neq 1$. So

$$\Pr(V = 1 \mid C = 3) \neq \Pr(V = 1)$$

for all six variables, V, which shows that none of them is independent of C. ■

(b) [5 pts] Show that M and S are pairwise independent.

Solution. To see that M and S are pairwise independent, we check each of the cases.

$$\begin{aligned}
\Pr(S = 0 \text{ and } M = 0) &= \Pr(HHT, HTH, THH) \\
&= \frac{3}{8} = \frac{1}{2} \cdot \frac{3}{4} = \Pr(S = 0) \cdot \Pr(M = 0) \\
\Pr(S = 0 \text{ and } M = 1) &= \Pr(TTT) \\
&= \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = \Pr(S = 0) \cdot \Pr(M = 1) \\
\Pr(S = 1 \text{ and } M = 0) &= \Pr(TTH, THT, HTT) \\
&= \frac{3}{8} = \frac{1}{2} \cdot \frac{3}{4} = \Pr(S = 1) \cdot \Pr(M = 0) \\
\Pr(S = 1 \text{ and } M = 1) &= \Pr(HHH) \\
&= \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{4} = \Pr(S = 1) \cdot \Pr(M = 1).
\end{aligned}$$

■

(c) [10 pts] Show that H_1, H_2, H_3, and S are 3-wise independent, but not mutually independent.

Solution. Since H_1, H_2, H_3 determine all the other variables, then by the same kind of argument used in part (a), no set of four or more variables including these three can be mutually independent.

The variables H_1, H_2, H_3 are 3-wise independent by definition of "flipping independently." Now consider the three variables H_1, H_2, S.

$$\begin{aligned}
&\Pr(S = a \text{ and } H_1 = b \text{ and } H_2 = c) \\
&= \Pr((b + c + H_3 \equiv a \pmod 2) \text{ and } H_1 = b \text{ and } H_2 = c) \quad ([S = a] ::= [H_1 + H_2 + H_3 \equiv a \pmod 2]) \\
&= \Pr(H_3 = rem(a - b - c, 2) \text{ and } H_1 = b \text{ and } H_2 = c) \\
&= \Pr(H_3 = rem(a - b - c, 2)) \cdot \Pr(H_1 = b) \cdot \Pr(H_2 = c) \qquad \text{(independence of the flips)} \\
&= (1/2)^3 \\
&= \Pr(S = a) \cdot \Pr(H_1 = b) \cdot \Pr(H_2 = c) \qquad (\text{since } \Pr(S = a) = 1/2).
\end{aligned}$$

Likewise for S and any other two H_i's. So H_1, H_2, H_3, and S are 3-wise independent because any three of them are mutually independent. ■

Problem 5. [15 points] A drunken sailor wanders along main street, 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 which 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	?	?	?	?	?	?	?	?	?

Solution.

	location								
	-4	-3	-2	-1	0	1	2	3	4
initially					1				
after 1 step				1/2	0	1/2			
after 2 steps			1/4	0	2/4	0	1/4		
after 3 steps		1/8	0	3/8	0	3/8	0	1/8	
after 4 steps	1/16	0	4/16	0	6/16	0	4/16	0	1/16

■

(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?

Solution. If he takes i steps to the right, then he takes $t - i$ steps to the left. Since steps left and right cancel, he nets $i - (t - i) = 2i - t$ steps to the right, ending at location $2i - t$.

The number of paths is the number of length-t sequences with i "right"s, which is $\binom{t}{i}$.

Each path is equally likely, so he takes the given path with probability $1/2^t$. Thus, he ends at the location $2i - t$ with probability

$$2^{-t}\binom{t}{i}.$$

■

(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.

Solution. From part (b), we have $\Pr(L = 2x - t) = 2^{-t}\binom{t}{x}$ for $0 \le x \le t$, so:

$$\begin{aligned} PDF_B(x) &= \Pr(B = x) \\ &= \Pr((L+t)/2 = x) \\ &= \Pr(L = 2x - t) \\ &= \frac{1}{2^t}\binom{t}{x}, \end{aligned}$$

which is the binomial distribution.

Intuitively, B represents the probability of taking exactly x steps to the right, which is the same as flipping a fair coin that comes up heads exactly x times. Notice that the probability that $B = x$ is the same as the probability that the path ends at location $L = 2x - t$, so our answer matches that from part (b). ■

Problem 6. [20 points]

Suppose n balls are thrown randomly into n boxes, so each ball lands in each box with uniform probability. Also, suppose the outcome of each throw is independent of all the other throws.

(a) [5 pts] Let X_i be an indicator random variable whose value is 1 if box i is empty and 0 otherwise. Write a simple closed form expression for the probability distribution of X_i. Are $X_1, X_2, \ldots, X_n$ independent random variables?

Solution. Box i is empty iff all n balls land in other boxes. The probability that a ball will land in another box in $(n-1)/n = 1-(1/n)$, and since the balls are thrown independently, we have

$$\Pr(X_i = 1) = \left(1 - \frac{1}{n}\right)^n. \tag{1}$$

The X_i's are not independent. For example,

$$\Pr(X_1 = X_2 = \cdots = X_n = 1) = 0 < \prod_{i=1}^{n} \Pr(X_i = 1).$$

■

(b) [5 pts] Show that

$$\Pr(\text{at least } k \text{ balls fall in the first box}) \le \binom{n}{k}\left(\frac{1}{n}\right)^k.$$

Solution. Let S be a set of k of the n balls, and let E_S be the event that each of these k balls falls in the first box. Since the probability that a ball lands in this box is $1/n$, and the throws are independent, we have

$$\Pr(E_S) = \left(\frac{1}{n}\right)^k. \tag{2}$$

The event that *at least* k balls land in the first box is the union of all the events E_S. There are $\binom{n}{k}$ subsets, S, of k balls, so by the Union Bound,

$$\Pr(\text{at least } k \text{ balls fall in the first box}) \le \binom{n}{k} \cdot \Pr(E_S).$$

Using the value for $\Pr(E_S)$ from (2) in the preceding inequality yields the required bound.
■

(c) [5 pts] Let R be the maximum of the numbers of balls that land in each of the boxes. Conclude from the previous parts that

$$\Pr(R \ge k) \le \frac{n}{k!}.$$

Solution. Note that $R \geq k$ exactly when some box has at least k balls. Since the bound on the probability of at least k balls in the first box applies just as well to any box, we can apply the Union Bound to having at least k balls in at least one of the n boxes:

$$\Pr(R \geq k) \leq n \Pr(\text{at least } k \text{ balls fall in the first box}).$$

So from the previous problem part, we have

$$\begin{aligned}
\Pr(R \geq k) &\leq n \binom{n}{k} \left(\frac{1}{n}\right)^k \\
&= n \left(\frac{n(n-1)\cdots(n-k+1)}{k!\, n^k}\right) \\
&= \frac{n}{k!} \left(\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}\right) \\
&\leq \frac{n}{k!}
\end{aligned}$$

■

(d) [5 pts] Conclude that

$$\lim_{n \to \infty} \Pr(R \geq n^\epsilon) = 0$$

for all $\epsilon > 0$.

Solution. Using Stirling's formula, and the upper bound from the previous part, we have

$$\Pr(R \geq k) \leq \frac{n}{k!} \sim \frac{n}{\sqrt{2\pi k}(k/e)^k} \leq \frac{n}{(k/e)^k} = \frac{ne^k}{k^k} = \frac{e^{k+\ln n}}{e^{k \ln k}}.$$

Now let $k = n^\epsilon$. Then the exponent of e in the numerator above is $n^\epsilon + \ln n$, and the exponent of e in the denominator is $n^\epsilon \ln n^\epsilon$. We conclude

$$\Pr(R \geq n^\epsilon) \leq \frac{e^{n^\epsilon + \ln n}}{e^{n^\epsilon \ln n^\epsilon}} \to 0$$

as n approaches ∞. ■