Problem Set 11

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

(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] 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 $\Pr(W) = \Pr(L)$.

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

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.

(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 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 $C = 3$.

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

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

(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 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?

(b) [5 pts] Show that

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

(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

$$\Pr\left(R \geq k\right) \leq \frac{n}{k!}.$$

(d) [5 pts] Conclude that

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

for all $\epsilon > 0$.