Basic Algorithms — Fall 2020 — Problem Set 6
Due: Wed, Nov 25, 11am

1. **Birthday paradox.** Let $X_1, \ldots, X_n$ be random variables that are uniformly and independently distributed over the set $\{1, \ldots, m\}$. Assume $n \leq m$. Let $p_{n,m}$ be the probability that $X_i = X_j$ for some $i \neq j$.

   (a) Show that
   $$p_{n,m} \leq \frac{n(n-1)}{2m}.$$
   Hint: union bound (and you really only need pairwise independence).
   Note: This says that if $n \leq \sqrt{m}$, then the probability that there is a collision among the X_i's (i.e., two X_i's taking the same value) is at most $1/2$.

   (b) Show that
   $$1 - p_{n,m} = \prod_{i=1}^{n} \left(1 - \frac{i-1}{m}\right).$$
   Hint: For this, you need to use the assumption that the X_i's are mutually independent. Specifically, use the fact that for every n-tuple of values $(s_1, \ldots, s_n) \in \{1, \ldots, m\}^n$, we have
   $$\Pr[(X_1, \ldots, X_n) = (s_1, \ldots, s_n)] = \frac{1}{m^n}.$$

   (c) Using the part (b), along with the handy inequality $1 + x \leq e^x$ (which holds for all real numbers x), show that
   $$p_{n,m} \geq 1 - e^{-n(n-1)/2m}.$$

   (d) Using part (c), show that if $n \geq \sqrt{2\ln(2)m} + 1$, then $p_{n,m} \geq 1/2$.
   Note: This says we only need to have $n \approx 1.177\sqrt{m}$ in order for there to be a collision among the X_i's with probability *at least* $1/2$. This is a special case of the "birthday paradox", which says that if there are 23 people in a room, it is more likely than not that two people in the room share the same birthday (you can plug $n = 23$ and $m = 365$ directly into the inequality in part (c) to see this).

2. **Computing expectations.** Let $X_1, \ldots, X_n$ be uniformly and independently distributed over the set
   $$\{-3, -2, -1, 0, 1, 2, 3\}.$$
   Let $S := X_1 + \cdots + X_n$. Compute $\mathrm{E}[S^2]$ as a function of n.

   Hints: linearity of expectation, product rule for independent random variables.

3. **Tossing coins.** You toss a coin until you get a total of k heads. What is the expected number of coin tosses, as a function of k?

   Hint: Let X be a random variable representing total number of coin tosses; write as a sum of random variables $X = X_1 + \cdots + X_k$, and use linearity of expectation. The random variable X_i is defined to be the number of coin tosses you make to get the ith head, after you have already gotten $i - 1$ heads. Each X_i has the same distribution. What is it?

4. **Doubling down.** Recall the dice game played between Alice and Bob: Alice rolls two dice, and tells Bob the sum; then Bob guesses a number from 1 to 6. Bob wins the game if his guess appears on either of the two dice.

   Now suppose Alice and Bob play for money. If Bob wins, he wins a dollar, and if he loses, Alice wins a dollar. We can model this using a random variable W representing Bob's winnings, where $W = 1$ if Bob wins, and $W = -1$ if Bob loses. We saw that using an optimal strategy, Bob wins with probability $5/9$. This means
   $$\mathrm{E}[W] = (1) \cdot \Pr[W = 1] + (-1) \Pr[W = -1] = (1)(5/9) + (-1)(4/9) = 1/9.$$

   To make the game more interesting, Alice allows Bob to "double down": this means that *after* Alice tells Bob the sum, Bob is allowed to double his bet from one dollar to two dollars, if he so chooses. So, if Bob doubles down, then W takes the values ± 2, and if Bob does not double down, then W takes the values ± 1 as before.

Give an optimal strategy for Bob (for both his guess and the double-down decision) and compute $\mathrm{E}[W]$ for this strategy.

Hint: If Z is a random variable representing the sum of the two dice, compute $\mathrm{E}[W]$ using the law of total expectation:

$$\mathrm{E}[W] = \sum_{\ell=2}^{12} \mathrm{E}[W \mid Z = \ell] \Pr[Z = \ell].$$

5. **Wheel of fortune (a couple of spins).** You have a wheel that you can spin, labeled with numbers $1, \ldots, n$. For example, here is a picture of a wheel with $n = 10$:

Each time you spin the wheel, it comes up on a number that is uniformly distributed over $\{1, \ldots, n\}$, and independently of all other spins.

(a) What is the expected value of the sum of two spins? Express your answer as a function of n.

(b) What is the expected value of the product of two spins? Express your answer as a function of n.

6. **Wheel of fortune (keep on spinning).** Consider again the wheel in Exercise 5. You spin the wheel once, and that comes up on a number, say X. You then continue spinning the wheel, stopping when when one of the additional spins comes up on a number that is *at most* X. Let T be the number of additional spins.

Show that $\mathrm{E}[T] = \ln(n) + O(1)$.

Hint: use the law of total expectation to write

$$\mathrm{E}[T] = \sum_{i=1}^{n} \mathrm{E}[T \mid X = i] \Pr[X = i].$$

First argue that for each $i = 1, \ldots, n$, the distribution of the random variable T, given that $X = i$, is a geometric ditribution with a particular success probability p_i. What is p_i (expressed as a function of i and n)?

Note: remember that the above "implicit big-O" notation means that you must show that

$$|\mathrm{E}[T] - \ln(n)| \le c$$

for some positive constant c and all sufficiently large n.

7. **Wheel of fortune (all the numbers).** Consider again the wheel in Exercise 5. You spin the wheel as many times as it takes until every number on the wheel comes up at least once. Let X be the total number of spins.

Show that $\mathrm{E}[X] = n \ln(n) + O(n)$.

Hint: using an idea similar to that used in Exercise 3, express X as a sum $X = X_1 + \cdots + X_n$. The random variable X_i is defined to be the number of spins you make to get the ith distinct number, after you have already gotten $i - 1$ distinct numbers. What is the distribution of X_i?

Note: remember that the above "implicit big-O" notation means that you must show that

$$|\mathrm{E}[X] - n \ln(n)| \le cn$$

for some positive constant c and all sufficiently large n.

8. **Wheel of fortune (the smallest number).** Consider again the wheel in Exercise 5. You spin the wheel t times. Let M be the smallest number that comes up on any of these t spins. Your goal is to show that

$$\mathrm{E}[M] = n/(t+1) + O(1).$$

(a) Show that for $j = 1, \ldots, n$, we have $\Pr[M \geq j] = (n - j + 1)^t / n^t$.

(b) Using the *tail sum formula* for expectation, along with part (a), show that

$$\mathrm{E}[M] = \frac{1}{n^t} \sum_{i=1}^{n} i^t.$$

(c) Approximating a sum by an integral, use part (b) to show that

$$\mathrm{E}[M] = \frac{n}{t+1} + O(1).$$

Note: remember that the above "implicit big-O" notation means that you must show that

$$\left| \mathrm{E}[M] - \frac{n}{t+1} \right| \leq c$$

for some positive constant c and all sufficiently large n.