CS 70 Discrete Mathematics and Probability Theory
Spring 2018 Satish Rao and Babak Ayazifar **HW 9**

Sundry

Before you start your homework, write down your team. Who else did you work with on this homework? List names and email addresses. (In case of homework party, you can also just describe the group.) How did you work on this homework? Working in groups of 3-5 will earn credit for your "Sundry" grade.

Please copy the following statement and sign next to it:

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up.

I certify that all solutions are entirely in my words and that I have not looked at another student's solutions. I have credited all external sources in this write up. (*signature here*)

1 Indicator Variables

(a) After throwing n balls into m bins at random, what is the expected number of bins that contains exactly k balls?

(b) Alice and Bob each draw k cards out of a deck of 52 distinct cards with replacement. Find k such that the expected number of common cards that both Alice and Bob draw is at least 1.

(c) How many people do you need in a room so that you expect that there is going to be a shared birthday on a Monday of the year (assume 52 Mondays in a year and 365 days in a year)?

Solution:

(a) Let $X_i = 1$ if bin i contains exactly k balls and $X_i = 0$ otherwise.

$$\mathbb{E}[X_i] = \binom{n}{k}\left(\frac{1}{m}\right)^k\left(\frac{m-1}{m}\right)^{n-k} = \binom{n}{k}\frac{(m-1)^{n-k}}{m^n}$$

$$\mathbb{E}[X] = \sum_{i=1}^{m}\binom{n}{k}\frac{(m-1)^{n-k}}{m^n} = \binom{n}{k}\frac{(m-1)^{n-k}}{m^{n-1}}$$

(b) Let $X_i = 1$ if card i is chosen by both Alice and Bob and $X_i = 0$ otherwise.
After drawing k cards, the probability that any given card appears at least once $1-(51/52)^k$

so

$$\mathbb{E}[X_i] = \left(1-\left(\frac{51}{52}\right)^k\right)\cdot\left(1-\left(\frac{51}{52}\right)^k\right)$$
$$\mathbb{E}[X] = \sum_{i=1}^{52}\left(1-\left(\frac{51}{52}\right)^k\right)^2 = 52\cdot\left(1-\left(\frac{51}{52}\right)^k\right)^2.$$

Setting $\mathbb{E}[X]=1$, we have $k=7.69\approx 8$.

(c) For $i<j$, let $X_{i,j}=1$ if i,j share a birthday and $X_{i,j}=0$ otherwise. Then, the total number of shared birthdays is $X=\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}X_{i,j}$, where k is the total number of people in the room. There is $52/365$ chance that person i has a birthday on a Monday and $1/365$ chance that person j has same birthday as person i so

$$\mathbb{E}[X] = \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\frac{52}{365}\cdot\frac{1}{365} = \binom{k}{2}\frac{52}{365^2} = \frac{k(k-1)}{2}\cdot\frac{52}{365^2}.$$

We want $\mathbb{E}[X]=1$ so $k\geq 73$.

2 Sinho's Dice

Sinho rolls three fair-sided dice.

(a) Let X denote the maximum of the three values rolled. What is the distribution of X(that is, $\mathbb{P}[X=x]$ for $x=1,2,3,4,6$)? You can leave your final answer in terms of "x'. [*Hint*: Let X denote the maximum of the three values roled. First compute $\mathbb{P}[X\leq x]$ for $x=1,2,3,4,5,6$].

(b) Let Y denote the minimum of the three values rolled. What is the distribution of Y?

Solution:

(a) Let X denote the maximum of the three values rolled. We are interested in $\mathbb{P}(X=x)$, where $x=1,2,3,4,5,6$. First, define X_1, X_2, X_3 to be the values rolled by the first, second, and third dice. These random variables are i.i.d. and uniformly distributed between 1 and 6 inclusive.

Following the hint we first compute $\mathbb{P}[X\leq x]$ for $x=1,2,3,4,5,6$:

$$\mathbb{P}(X\leq x) = \mathbb{P}(X_1\leq x)\mathbb{P}(X_2\leq x)\mathbb{P}(X_3\leq x) = \left(\frac{x}{6}\right)\left(\frac{x}{6}\right)\left(\frac{x}{6}\right) = \left(\frac{x}{6}\right)^3$$

Then, observing that $\mathbb{P}(X = x) = \mathbb{P}(X \le x) - \mathbb{P}(X \le x - 1)$:

$$\mathbb{P}(X = x) = \left(\frac{x}{6}\right)^3 - \left(\frac{x-1}{6}\right)^3 = \frac{3x^2 - 3x + 1}{216} = \begin{cases} \frac{1}{216}, & x = 1 \\ \frac{7}{216}, & x = 2 \\ \frac{19}{216}, & x = 3 \\ \frac{37}{216}, & x = 4 \\ \frac{61}{216}, & x = 5 \\ \frac{91}{216}, & x = 6 \end{cases}$$

One can confirm that $\sum_{x=1}^{6} \mathbb{P}(X = x) = 1$.

(b) Similarly to the previous part, we first compute $\mathbb{P}[Y \ge y]$.

$$\mathbb{P}(Y \ge y) = \mathbb{P}(X_1 \ge y)\mathbb{P}(X_2 \ge y)\mathbb{P}(X_3 \ge y) = \left(\frac{6-(y-1)}{6}\right)\left(\frac{6-(y-1)}{6}\right)\left(\frac{6-(y-1)}{6}\right) = \left(\frac{7-y}{6}\right)^3 .$$

Then, observing that $\mathbb{P}(Y = y) = \mathbb{P}(Y \ge y) - \mathbb{P}(Y \ge y - 1)$:

$$\mathbb{P}[Y = y] = \left(\frac{7-y}{6}\right)^3 - \left(\frac{6-y}{6}\right)^3 .$$

3 Linearity

Solve each of the following problems using linearity of expectation. Explain your methods clearly.

(a) In an arcade, you play game A 10 times and game B 20 times. Each time you play game A, you win with probability $1/3$ (independently of the other times), and if you win you get 3 tickets (redeemable for prizes), and if you lose you get 0 tickets. Game B is similar, but you win with probability $1/5$, and if you win you get 4 tickets. What is the expected total number of tickets you receive?

(b) A monkey types at a 26-letter keyboard with one key corresponding to each of the lower-case English letters. Each keystroke is chosen independently and uniformly at random from the 26 possibilities. If the monkey types 1 million letters, what is the expected number of times the sequence "book" appears?

(c) A building has n floors numbered $1, 2, \dots, n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each gets off at a uniformly random one of the n floors (independently of everybody else). What is the expected number of floors the elevator stops at (not counting the ground floor)?

(d) A coin with heads probability p is flipped n times. A "run" is a maximal sequence of consecutive flips that are all the same. (Thus, for example, the sequence $HTHHHTTH$ with $n = 8$ has five runs.) Show that the expected number of runs is $1 + 2(n-1)p(1-p)$. Justify your calculation carefully.

Solution:

(a) Let A_i be the indicator you win the ith time you play game A and B_i be the same for game B. The expected value of A_i and B_i are

$$\mathbb{E}[A_i] = 1 \cdot \frac{1}{3} + 0 \cdot \frac{2}{3} = \frac{1}{3},$$
$$\mathbb{E}[B_i] = 1 \cdot \frac{1}{5} + 0 \cdot \frac{4}{5} = \frac{1}{5}.$$

Let T_A be the random variable for the number of tickets you win in game A, and T_B be the number of tickets you win in game B.

$$\begin{aligned}\mathbb{E}[T_A + T_B] &= 3\,\mathbb{E}[A_1] + \cdots + 3\,\mathbb{E}[A_{10}] + 4\,\mathbb{E}[B_1] + \cdots + 4\,\mathbb{E}[B_{20}] \\ &= 10\left(3 \cdot \frac{1}{3}\right) + 20\left(4 \cdot \frac{1}{5}\right) = 26\end{aligned}$$

(b) There are $1,000,000 - 4 + 1 = 999,997$ places where "book" can appear, each with a (non-independent) probability of $1/26^4$ of happening. If A is the random variable that tells how many times "book" appears, and A_i is the indicator variable that is 1 if "book" appears starting at the ith letter, then

$$\begin{aligned}\mathbb{E}[A] &= \mathbb{E}[A_1 + \cdots + A_{999,997}] \\ &= \mathbb{E}[A_1] + \cdots + \mathbb{E}[A_{999,997}] \\ &= \frac{999,997}{26^4} \approx 2.19.\end{aligned}$$

(c) Let A_i be the indicator that the elevator stopped at floor i.We know the elevator will only stop at floor i if at least one person gets out.

$$\mathbb{P}[A_i = 1] = 1 - \mathbb{P}[\text{no one gets off at } i] = 1 - \left(\frac{n-1}{n}\right)^m.$$

If A is the number of floors the elevator stops at, then

$$\begin{aligned}\mathbb{E}[A] &= \mathbb{E}[A_1 + \cdots + A_n] \\ &= \mathbb{E}[A_1] + \cdots + \mathbb{E}[A_n] = n \cdot \left[1 - \left(\frac{n-1}{n}\right)^m\right].\end{aligned}$$

(d) .

Let A_i be the indicator for the event that a run starts at the i-th toss. Let $A = A_1 + \cdots + A_n$ be the random variable for the number of runs total. Obviously, $\mathbb{E}[A_1] = 1$. For $i \neq 1$, we know that a run starts at position i if the i-th toss differs from the $(i-1)$-th toss and, thus,

$$\begin{aligned}\mathbb{E}[A_i] &= \mathbb{P}[A_i = 1] \\ &= \mathbb{P}[i = H \mid i-1 = T] \cdot \mathbb{P}[i-1 = T] + \mathbb{P}[i = T \mid i-1 = H] \cdot \mathbb{P}[i-1 = H] \\ &= p \cdot (1-p) + (1-p) \cdot p \\ &= 2p \cdot (1-p).\end{aligned}$$

This gives

$$\begin{aligned}\mathbb{E}[A] &= \mathbb{E}[A_1 + A_2 + \cdots + A_n] \\ &= \mathbb{E}[A_1] + \mathbb{E}[A_2] + \cdots + \mathbb{E}[A_n] = 1 + 2(n-1)p(1-p).\end{aligned}$$

4 Random Coloring

Consider a graph $G(V,E)$ with m edges and a set of q colors. Our goal is to prove that there exists a vertex-coloring of the graph using these q colors so that at most $\frac{m}{q}$ edges are monochromatic. An edge is monochromatic if both its vertices are assigned the same color.

(a) Suppose we color the graph randomly. That is, for each vertex $v \in V$ we choose a color uniformly at random and assign it to v. Prove that the expected number of monochromatic edges is $\frac{m}{q}$.

(b) Conclude that there exists a vertex-coloring of G so that at most $\frac{m}{q}$ edges are monochromatic. (Hint: Suppose that every coloring of G is such so that at least $\frac{m}{q} + 1$ edges are monochromatic and reach a contradiction.)

Solution:

(a) Let X be the random variable that equals the number of monochromatic edges. For an edge e let X_e be indicator random variable that equals 1 if X_e is monochromatic or 0 otherwise. Notice that $X = \sum_{e \in E} X_e$. Thus, using linearity of expectation we have that:

$$\mathbb{E}[X] = \sum_{e \in E} \mathbb{E}[X_e] = m\mathbb{P}[X_e = 1] = m\frac{1}{q} \ ,$$

since for any given vertex e, there are q out of q^2 ways in which it can be monochromatic.

(b) Let $\mathscr{C}$ denote the set of all possible colorings of G. Assume that every coloring in $\mathscr{C}$ is such so that at least $\frac{m}{q} + 1$ edges are monochromatic. Moreover, for a coloring $c \in \mathscr{C}$ let $M(c)$ denote the number of monochromatic edges in c.

Notice now that coloring each vertex of G uniformly at random is equivalent to choosing an element of $\mathscr{C}$ uniformly at random. Thus, the expectation of X can be written as

$$\begin{aligned}\mathbb{E}[X] &= \sum_{c\in\mathscr{C}} \mathbb{P}[c]M(c) \\ &= \frac{1}{|\mathscr{C}|}\sum_{c\in\mathscr{C}} M(c) \\ &\geq \frac{1}{|\mathscr{C}|}\left(\frac{m}{q}+1\right)|\mathscr{C}| \\ &\geq \frac{m}{q}+1 \\ &> \frac{m}{q} .\end{aligned}$$

According to part (a) we have reached a contradiction.

5 Long Streaks

Suppose we flip a coin n times to obtain a sequence of flips $X_1, X_2, \ldots, X_n$. A *streak* of flips is a consecutive subsequence of flips that are all the same. For example, if X_3, X_4, and X_5 are all heads, there is a streak of length 3 starting at the third flip.

(a) Let n be a power of 2. Show that the expected number of streaks of length $\log_2 n+1$ is $1-\frac{\log_2 n}{n}$ and notice that for large n it tends to 1.

(b) Given a sequence of flips of length n, suppose we break it up into disjoint blocks of $\lfloor \log_2 n - 2\log_2\log_2 n\rfloor$ consecutive flips. Show that for every block the probability that is a streak is at least $2\frac{(\log_2 n)^2}{n}$.

(c) Show that, for sufficiently large n, the probability that there is no streak of length at least $\lfloor \log_2 n - 2\log_2\log_2 n\rfloor$ is less than $1/n$. (Hint: Use the previous part. You might find useful the fact that for every $x>0$ we have $(1-\frac{1}{x}) < \mathrm{e}^{-x}$).

Solution:

(a) Let Y_k denote the indicator r.v. of whether the subsequence starting at X_k of length $\log_2 n+1$. Since the flips are independent, we have

$$\mathbb{E}[Y_k] = 2(\frac{1}{2})^{\log_2 n+1} = \frac{1}{n}$$

Notice the 2 at the front is because a streak can be either all heads or all tails. Also $\frac{1}{2}^{\log_2 n+1} = \frac{1}{2^{\log_2 n+1}} = \frac{1}{2n}$. Now the total number of streaks of length $\log_2 n+1$ is the sum of $Y_1,\ldots,Y_{n-(\log_2 n+1)+1}$, thus the expected number of streaks is

$$\mathbb{E}[\sum_{i=1}^{n-\log_2 n} Y_i] = \sum_{i=1}^{n-\log_2 n}\mathbb{E}[Y_i] = (n-\log_2 n)\frac{1}{n} = 1-\frac{\log_2 n}{n} .$$

(b) The probability of having a streak for an individual block is

$$2(\frac{1}{2})^{\lfloor \log_2 n - 2\log_2 \log_2 n \rfloor} \geq 2(\frac{1}{2})^{\log_2 n - 2\log_2 \log_2 n} = 2\frac{(\log_2 n)^2}{n} \quad .$$

(c) The probability of each block being a streak are independent since the blocks are disjoint, and for there to be no streak of length at least $\lfloor \log_2 n - 2\log_2 \log_2 n \rfloor$, a necessary (though not sufficient) condition is that all the disjoint blocks are not streaks, which happens with probability

$$\begin{aligned}
(1 - 2\frac{(\log_2 n)^2}{n})^{\frac{n}{\lfloor \log_2 n - 2\log_2 \log_2 n \rfloor}} &< (1 - 2\frac{(\log_2 n)^2}{n})^{\frac{n}{\log_2 n}} \\
&< \mathrm{e}^{-2\frac{(\log_2 n)^2}{n}\frac{n}{\log_2 n}} \\
&= \frac{1}{e}^{2\log_2 n} \\
&< \frac{1}{e}^{\ln n} \\
&= \frac{1}{n} \quad .
\end{aligned}$$