CS 70 Spring 2021 — Discrete Mathematics and Probability Theory — DIS 10A

1 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?

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

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

2 Ball in Bins

You are throwing k balls into n bins. Let X_i be the number of balls thrown into bin i.

(a) What is $\mathbb{E}[X_i]$?

(b) What is the expected number of empty bins?

(c) Define a collision to occur when two balls land in the same bin (if there are n balls in a bin, count that as $n-1$ collisions). What is the expected number of collisions?

Solution:

(a) We will use linearity of expectation. Note that the expectation of an indicator variable is just the probability the indicator variable = 1. (Verify for yourself that is true).

$$\mathbb{E}[X_i] = \mathbb{P}[\text{ball 1 falls into bin } i] + \mathbb{P}[\text{ball 2 falls into bin } i] \cdots = \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n}.$$

(b) Let X_i be the indicator variable denoting whether bin i ends up empty. This can happen if and only if all the balls end in the remaining $n-1$ bins, and this happens with a probability of $\left(\frac{n-1}{n}\right)^k$. Hence the expected number of empty bins is

$$\mathbb{E}[X_1 + \ldots + X_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n] = n\left(\frac{n-1}{n}\right)^k$$

(c) The number of collisions is the number of balls minus the number of occupied bins, since the first ball of every occupied bin is not a collision.

$$\begin{aligned}\mathbb{E}[\text{collisions}] &= k - \mathbb{E}[\text{occupied bins}] = k - n + \mathbb{E}[\text{empty locations}] \\ &= k - n + n\left(1 - \frac{1}{n}\right)^k\end{aligned}$$

3 Swaps and Cycles

We'll say that a permutation $\pi = (\pi(1), \ldots, \pi(n))$ contains a *swap* if there exist $i, j \in \{1, \ldots, n\}$ so that $\pi(i) = j$ and $\pi(j) = i$.

(a) What is the expected number of swaps in a random permutation?

(b) In the same spirit as above, we'll say that π contains a *s-cycle* if there exist $i_1, \ldots, i_s \in \{1, \ldots, n\}$ with $\pi(i_1) = i_2, \pi(i_2) = i_3, \ldots, \pi(i_s) = i_1$. Compute the expectation of the number of s-cycles.

Solution:

(a) As a warm-up, let's compute the probability that 1 and 2 are swapped. There are $n!$ possible permutations, and $(n-2)!$ of them have $\pi(1) = 2$ and $\pi(2) = 1$. This means

$$\mathbb{P}[(1,2) \text{ are a swap}] = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.$$

There was nothing special about 1 and 2 in this calculation, so for any $\{i, j\} \subset \{1, \ldots, n\}$, the probability that i and j are swapped is the same as above. Let's write $I_{i,j}$ for the indicator that i and j are swapped, and N for the total number of swaps, so that

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{\{i,j\}\subset\{1,\ldots,n\}} I_{i,j}\right] = \sum_{\{i,j\}\subset\{1,\ldots n\}} \mathbb{P}[(i,j) \text{ are swapped}] = \frac{1}{n(n-1)}\binom{n}{2} = \frac{1}{2}.$$

(b) The idea here is quite similar to the above, so we'll be a little less verbose in the exposition. However, as a first aside we need the notion of a *cyclic ordering* of s elements from a set $\{1, \ldots n\}$. We mean by this a labelling of the s beads of a necklace with elements of the set, where we say that labelings of the beads are the same if we can move them along the string to turn one into the other. For example, $(1,2,3,4)$ and $(1,2,4,3)$ are different cyclic orderings, but $(1,2,3,4)$ and $(2,3,4,1)$ are the same. There are

$$\binom{n}{s}\frac{s!}{s} = \frac{n!}{(n-s)!}\frac{1}{s}$$

possible cyclic orderings of length s from a set with n elements, since if we first count all subsets of size s, and then all permutations of each of those subsets, we have overcounted by a factor of s.

Now, let N be a random variable counting the number of s-cycles, and for each cyclic ordering $(i_1, \ldots, i_s)$ of s elements of $\{1, \ldots, n\}$, let $I_{(i_1,\ldots,i_s)}$ be the indicator that $\pi(i_1) = i_2, \pi(i_2) = i_3, \ldots, \pi(i_s) = i_1$. There are $(n-s)!$ permutations in which $(i_1, \ldots, i_s)$ form an s-cycle (since we are free to do whatever we want to the remaining $(n-s)$ elements of $\{1, \ldots, n\}$), so the probability that $(i_1, \ldots, i_s)$ are such a cycle is $\frac{(n-s)!}{n!}$, and

$$\mathbb{E}[N] = \mathbb{E}\left[\sum_{(i_1,\ldots,i_s) \text{ cyclic ordering}} I_{(i_1,\ldots,i_s)}\right] = \frac{n!}{(n-s)!}\frac{1}{s}\frac{(n-s)!}{n!} = \frac{1}{s}.$$