CS 70 Discrete Mathematics and Probability Theory
Fall 2021 DIS 10A

1 How Many Marbles?

Leanne has 6 marbles, 2 red, 2 blue, and 2 green. She picks three marbles uniformly at random without replacement. Let X denote the number of blue marbles she draws.

(a) What is $\mathbb{P}[X=0]$, $\mathbb{P}[X=1]$, and $\mathbb{P}[X=2]$?

(b) What do your answers you computed in part (a) add up to?

(c) Compute $\mathbb{E}[X]$ from the definition of expectation.

(d) Suppose we define indicators X_i, $1 \leq i \leq 3$, where X_i is the indicator variable that equals 1 if the ith marble is a blue marble and 0 otherwise. Compute $\mathbb{E}[X]$ using linearity of expectation.

(e) Are the X_i indicators independent? Does this affect your solution to part (d)?

Solution:

(a) Calculate each case of $X = 0, 1$, and 2:

We must draw three non-blue marbles in a row, so the probability is

$$\mathbb{P}[X=0] = \frac{4}{6} \cdot \frac{3}{5} \cdot \frac{2}{4} = \frac{1}{5}.$$

Alternatively, every set of three marbles is equally likely, so we can use counting. There are $\binom{6}{3}$ total sets of three marbles, and $\binom{4}{3}$ sets with only non-blue marbles, which gives us the same result.

$$\mathbb{P}[X=0] = \frac{\binom{4}{3}}{\binom{6}{3}} = \frac{1}{5}.$$

- We will continue to use counting. The number of sets of three marbles with exactly one blue marble amounts to the number of ways to choose 1 blue marble out of 2, and 2 non-blues out of 4.

$$\mathbb{P}[X=1] = \frac{\binom{2}{1}\binom{4}{2}}{\binom{6}{3}} = \frac{3}{5}$$

- Choose 2 blue marbles out of 2, and 1 non-blue out of 4.

$$\mathbb{P}[X=2] = \frac{\binom{2}{2}\binom{4}{1}}{\binom{6}{3}} = \frac{1}{5}.$$

(b) We check:

$$\mathbb{P}[X=0]+\mathbb{P}[X=1]+\mathbb{P}[X=2]=\frac{1+3+1}{5}=1$$

(c) From the definition, $\mathbb{E}[X]=\sum_{k=0}^{2} k\mathbb{P}[X=k]$, so

$$\mathbb{E}[X]=0\cdot\frac{1}{5}+1\cdot\frac{3}{5}+2\cdot\frac{1}{5}=1.$$

(d) We know that $\mathbb{E}[X_i]=\mathbb{P}[\text{marble } i \text{ is a blue marble}]+0\cdot\mathbb{P}[\text{marble } i \text{ is not a blue marble}]=\frac{2}{6}=\frac{1}{3}$, so

$$\mathbb{E}[X]=\mathbb{E}[X_1]+\mathbb{E}[X_2]+\mathbb{E}[X_3]=\frac{1}{3}+\frac{1}{3}+\frac{1}{3}=1.$$

Notice how much faster it was to compute the expectation using indicators!

(e) No, they are not independent. As an example:

$$\mathbb{P}[X_1=1]\mathbb{P}[X_2=1]=\frac{1}{3}\cdot\frac{1}{3}=\frac{1}{9}$$

However,

$$\mathbb{P}[X_1=1,X_2=1]=\mathbb{P}(\text{the first and second marbles are both blue})=\frac{2}{6}\cdot\frac{1}{5}=\frac{1}{15}.$$

Even though the indicators are not independent, this does not change our answer for part (d). Linearity of expectation *always* holds, which makes it an extremely powerful tool.

2 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

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

Then the expected total number of tickets you receive, by linearity of expectation, is

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

Note that $10\left(3 \cdot \frac{1}{3}\right)$ and $20\left(4 \cdot \frac{1}{5}\right)$ matches the expression directly gotten using the expected value of a binomial random variable.

(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

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

3 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 a ball lands in a nonempty 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 I_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

$$\mathbb{E}[I_1 + \ldots + I_n] = \mathbb{E}[I_1] + \ldots + \mathbb{E}[I_n] = n\left(\frac{n-1}{n}\right)^k$$

(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.

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