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 Joint Distributions

(a) Give an example of discrete random variables X and Y with the property that $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$. You should specify the joint distribution of X and Y.

(b) Give an example of discrete random variables X and Y that (i) are *not independent* and (ii) have the property that $\mathbb{E}[XY] = 0$, $\mathbb{E}[X] = 0$, and $\mathbb{E}[Y] = 0$. Again you should specify the joint distribution of X and Y.

Solution:

(a) Let $P(X=1) = \frac{1}{2}$, $P(X=-1) = \frac{1}{2}$, and $Y \equiv X$. Then $\mathbb{E}[X] = 1\mathbb{P}[X=1] + (-1)\mathbb{P}[X=-1] = 0$, and $\mathbb{E}[Y] = \mathbb{E}[X]$. Similarly, since $X = Y$, $\mathbb{E}[XY] = \mathbb{E}[X^2] = 1$ and $\mathbb{E}[X]\mathbb{E}[Y] = 0$.

(b) One example is given by $P(X=-1, Y=\frac{1}{3}) = P(X=1, Y=\frac{1}{3}) = P(X=0, Y=-\frac{2}{3}) = \frac{1}{3}$.

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 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.

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