CS 70 Spring 2022 — Discrete Mathematics and Probability Theory — Satish Rao and Koushik Sen — DIS 12A

1 LLSE

We have two bags of balls. The fractions of red balls and blue balls in bag A are $2/3$ and $1/3$ respectively. The fractions of red balls and blue balls in bag B are $1/2$ and $1/2$ respectively. Someone gives you one of the bags (unmarked) uniformly at random. You then draw 6 balls from that same bag with replacement. Let X_i be the indicator random variable that ball i is red. Now, let us define $X = \sum_{1 \le i \le 3} X_i$ and $Y = \sum_{4 \le i \le 6} X_i$.

(a) Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(b) Compute $\text{Var}(X)$.

(c) Compute $\text{cov}(X,Y)$. (*Hint*: Recall that covariance is bilinear.)

(d) Now, we are going to try and predict Y from a value of X. Compute $L(Y \mid X)$, the best linear estimator of Y given X. (*Hint*: Recall that

$$L(Y \mid X) = \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\text{Var}(X)}\big(X - \mathbb{E}[X]\big).$$

)

Solution: Although the indicator random variables are not independent, we can still apply linearity of expectation. By symmetry, we also know that each indicator follows the same distribution.

(a)

$$\mathbb{E}[X] = \mathbb{E}[Y] = 3 \cdot \mathbb{E}[X_1] = 3 \cdot \mathbb{P}[X_1 = 1] = 3 \cdot \left(\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{2}\right) = \frac{7}{4}.$$

(b)

$$\begin{aligned}
\text{Var}(X) &= \text{cov}\Big(\sum_{1 \le i \le 3} X_i, \sum_{1 \le j \le 3} X_j\Big) \\
&= 3 \cdot \text{Var}(X_1) + 6 \cdot \text{cov}(X_1, X_2) \\
&= 3\big(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2\big) + 6 \cdot \frac{1}{144} \\
&= 3\left[\frac{7}{12} - \left(\frac{7}{12}\right)^2\right] + 6 \cdot \frac{1}{144} = \frac{111}{144}.
\end{aligned}$$

(c)

$$\begin{aligned}\operatorname{cov}(X,Y) &= \operatorname{cov}\Big(\sum_{1\le i\le 3} X_i, \sum_{4\le j\le 6} X_j\Big)\\ &= 9\cdot \operatorname{cov}(X_1,X_4)\\ &= 9\cdot \big(\mathbb{E}[X_1X_4]-\mathbb{E}[X_1]\cdot\mathbb{E}[X_4]\big)\\ &= 9\cdot\left(\mathbb{P}[X_1=1,X_4=1]-\mathbb{P}[X_1=1]^2\right)\\ &= 9\cdot\left(\left[\frac{1}{2}\cdot\left(\frac{2}{3}\right)^2+\frac{1}{2}\cdot\left(\frac{1}{2}\right)^2\right]-\left[\frac{1}{2}\cdot\left(\frac{2}{3}\right)+\frac{1}{2}\cdot\left(\frac{1}{2}\right)\right]^2\right)=\frac{9}{144}.\end{aligned}$$

(d)

$$L(Y\mid X)=\frac{7}{4}+\frac{9}{111}\left(X-\frac{7}{4}\right)=\frac{3}{37}X+\frac{119}{74}.$$

2 Number Game

Sinho and Vrettos are playing a game where they each choose an integer uniformly at random from $[0,100]$, then whoever has the larger number wins (in the event of a tie, they replay). However, Vrettos doesn't like losing, so he's rigged his random number generator such that it instead picks randomly from the integers between Sinho's number and 100. Let S be Sinho's number and V be Vrettos' number.

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

(b) What is $\mathbb{E}[V|S=s]$, where s is any constant such that $0\le s\le 100$?

(c) What is $\mathbb{E}[V]$?

Solution:

(a) S is a (discrete) uniform random variable between 0 and 100, so its expectation is $\frac{0+100}{2}=50$.

(b) If $S=s$, we know that V will be uniformly distributed between s and 100. Similar to the previous part, this gives us that $\mathbb{E}[V|S=s]=\frac{s+100}{2}$.

(c) We have that

$$\begin{aligned}\mathbb{E}[V] &= \sum_{s=0}^{100}\mathbb{E}[V|S=s]\cdot\mathbb{P}[S=s]\\ &= \sum_{s=0}^{100}\frac{s+100}{2}\cdot\frac{1}{101}\\ &= \frac{1}{202}\left(\sum_{s=0}^{100}s+\sum_{s=0}^{100}100\right)\end{aligned}$$

The first summation comes out to $\frac{100(100+1)}{2} = 50 \cdot 101$; the second summation is just adding 100 to itself 101 times, so it comes out to $100 \cdot 101$. Plugging these values in, we get $\mathbb{E}[V] = 75$.

3 Number of Ones

In this problem, we will revisit dice-rolling, except with conditional expectation.

(a) If we roll a die until we see a 6, how many ones should we expect to see?

(b) If we roll a die until we see a number greater than 3, how many ones should we expect to see?

(Hint: for both of the above subparts, the Law of Total Expectation may be helpful)

Solution:

(a) Let Y be the number of ones we see. Let X be the number of rolls we take until we get a 6.

Let us first compute $\mathbb{E}[Y \mid X]$. We know that in each of our $k-1$ rolls before the kth, we necessarily roll a number in $\{1, 2, 3, 4, 5\}$. Thus, we have a $1/5$ chance of getting a one, meaning

$$\mathbb{E}[Y \mid X = k] = \frac{1}{5}(k-1)$$

so

$$\mathbb{E}[Y \mid X] = \frac{1}{5}(X-1).$$

If this is confusing, write Y as a sum of indicator variables.

$$Y = Y_1 + Y_2 + \cdots + Y_k$$

where Y_i is 1 if we see a one on the ith roll. This means

$$\mathbb{E}[Y \mid X = k] = \mathbb{E}[Y_1 \mid X = k] + \mathbb{E}[Y_2 \mid X = k] + \cdots + \mathbb{E}[Y_k \mid X = k].$$

We know for a fact that on the kth roll, we roll a 6, thus $\mathbb{E}[Y_k] = 0$. Thus, we actually consider

$$\begin{aligned}\mathbb{E}[Y_1 \mid X = k] + \mathbb{E}[Y_2 \mid X = k] + \cdots + \mathbb{E}[Y_{k-1} \mid X = k] &= (k-1)\,\mathbb{E}[Y_1 \mid X = k] \\ &= (k-1)\mathbb{P}[Y_1 = 1 \mid X = k] \\ &= (k-1)\frac{1}{5}.\end{aligned}$$

Using the Law of Total Expectation, we know that

$$\begin{aligned}\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] &= \mathbb{E}\left[\frac{1}{5}(X-1)\right] \\ &= \frac{1}{5}\,\mathbb{E}[X-1] \\ &= \frac{1}{5}(\mathbb{E}[X]-1).\end{aligned}$$

Since, $X \sim \text{Geometric}(1/6)$, the expected number of rolls until we roll a 6 is $\mathbb{E}[X] = 6$.

$$\frac{1}{5}(\mathbb{E}[X] - 1) = \frac{1}{5}(6 - 1) = 1.$$

(b) We use the same logic as the first part, except now each of the first $k-1$ rolls can only be 1, 2, or 3, so

$$\mathbb{E}[Y \mid X = k] = \frac{1}{3}(k - 1).$$

Then

$$\begin{aligned}\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}\left[\frac{1}{3}(X - 1)\right] \\ = \frac{1}{3}(\mathbb{E}[X] - 1).\end{aligned}$$

Since $X \sim \text{Geometric}(1/2)$, we know that the expected number of rolls until we roll a number greater than 3 is $\mathbb{E}[X] = 2$. This makes $\mathbb{E}[Y] = 1/3$.

4 Marbles in a Bag

We have r red marbles, b blue marbles, and g green marbles in the same bag. If we sample marbles with replacement until we get 3 red marbles (not necessarily consecutively), how many blue marbles should we expect to see? (*Hint*: It might be useful to use Law of Total Expectation, $E(Y) = E(E(Y|X))$.)

Solution:

Let Y be the number of blue marbles we see. Let X be the samples we take until we get 3 red marbles.

Let us first compute $\mathbb{E}[Y \mid X]$. Let Y_i be 1 if we see a blue marble on the ith sample and $Y = \sum_{i=1}^{k} Y_i$. This means

$$\begin{aligned}\mathbb{E}[Y \mid X = k] &= \mathbb{E}\left[\sum_{i=1}^{k} Y_i \,\middle|\, X = k\right] \\ &= \sum_{i=1}^{k} \mathbb{E}[Y_i \mid X = k].\end{aligned}$$

However, three Y_i (call them Y_a, Y_b, Y_c) have $\mathbb{E}[Y_i] = 0$, since there are necessarily 3 red marbles. This means the other $k-3$ marbles are necessarily blue or green.

$$\begin{aligned}\sum_{i \neq a,b,c} \mathbb{E}[Y_i \mid X = k] &= \sum_{i \neq a,b,c} \mathbb{P}[Y_i = 1 \mid X = k] \\ &= \sum_{i \neq a,b,c} \frac{b}{b+g} \\ &= (k-3)\frac{b}{b+g}.\end{aligned}$$

This means

$$\mathbb{E}[Y \mid X] = (X-3)\frac{b}{b+g}.$$

Using the Law of Total Expectation, we know that

$$\begin{aligned}
\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}\left[\frac{b}{b+g}(X-3)\right] \\
= \frac{b}{b+g}\,\mathbb{E}[X-3] \\
= \frac{b}{b+g}(\mathbb{E}[X]-3).
\end{aligned}$$

We notice that

$$X = X_1 + X_2 + X_3,$$

where each X_i represents the number of marbles seen between drawing the $(i-1)$th and ith red marble. We know that the absolute number of marbles seen between 2 consecutive red marbles is geometric, since we want to find the number of draws until the first red marble.

$$X_i \sim \text{Geometric}\left(\frac{r}{r+b+g}\right)$$

Since X_1, X_2, X_3 are identically distributed, we know that

$$\mathbb{E}[X] = 3\,\mathbb{E}[X_i] = 3(r+g+b)/r$$

.

$$\mathbb{E}[Y] = \frac{b}{b+g}\left(3\frac{r+g+b}{r} - 3\right) = \frac{3b}{r}.$$

Alternate Solution:

We know that the absolute number of marbles N seen between 2 consecutive red marbles is geometric, since we want to find the number of draws until the first red marble. And given the number of marbles, N, between 2 consecutive red marbles, the number of blue marbles among these is distributed binomially.

Therefore, each X_i is drawn from a binomial distribution, where the number of trials is distributed geometrically.

We exclude the very last marble in the binomial distribution, because we know it must be red (and therefore cannot be blue). And the probability for the binomial is $b/(b+g)$ because we know that in between 2 consecutive red balls, we can only have blue or green balls. So,

$$X_i \sim \text{Binomial}\left(\frac{b}{b+g}, N-1\right), \qquad \text{where} \qquad N \sim \text{Geometric}\left(\frac{r}{r+b+g}\right).$$

And, applying the law of conditional expectation, we have

$$\begin{aligned}
\mathbb{E}(X_i) &= \mathbb{E}(\mathbb{E}(X_i \mid N)) \\
&= \mathbb{E}\left((N-1)\frac{b}{b+g}\right) \\
&= \frac{b}{b+g}\,\mathbb{E}(N-1) \\
&= \frac{b}{b+g}\left(\frac{r+b+g}{r} - 1\right) \\
&= \frac{b}{b+g}\left(\frac{b+g}{r}\right) \\
&= \frac{b}{r}.
\end{aligned}$$

We know that each of the X_i's is identically distributed, so

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) = 3 \cdot \mathbb{E}(X_1) = \frac{3b}{r}.$$