1 Student Life

Note 19

In an attempt to avoid having to do laundry often, Marcus comes up with a system. Every night, he designates one of his shirts as his dirtiest shirt. In the morning, he randomly picks one of his shirts to wear. If he picked the dirtiest one, he puts it in a dirty pile at the end of the day (a shirt in the dirty pile is not used again until it is cleaned).

When Marcus puts his last shirt into the dirty pile, he finally does his laundry, and again designates one of his shirts as his dirtiest shirt (laundry isn't perfect) before going to bed. This process then repeats.

(a) If Marcus has n shirts, what is the expected number of days that transpire between laundry events? Your answer should be a function of n involving no summations.

(b) Say he gets even lazier, and instead of organizing his shirts in his dresser every night, he throws his shirts randomly onto one of n different locations in his room (one shirt per location), designates one of his shirts as his dirtiest shirt, and one location as the dirtiest location.

In the morning, if he happens to pick the dirtiest shirt, *and* the dirtiest shirt was in the dirtiest location, then he puts the shirt into the dirty pile at the end of the day and does not throw any future shirts into that location and also does not consider it as a candidate for future dirtiest locations (it is too dirty).

What is the expected number of days that transpire between laundry events now? Again, your answer should be a function of n involving no summations.

Solution:

(a) The number of days that it takes for him to throw a shirt into the dirty pile can be represented as a geometric RV. For the first shirt, this is the geometric RV with $p = 1/n$. We can see this by noticing that every day up to the day he picks the dirtiest shirt, the probability of getting the dirtiest shirt remains $1/n$.

We'll call X_i the number of days that go until he throws the ith shirt into the dirty pile. Since on the ith shirt, there are $n-i+1$ shirts left, we get that $X_i \sim \text{Geometric}(1/(n-i+1))$. The number of days until he does his laundry is a sum of these variables. Therefore, we can get the following result:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (n-i+1) = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

(b) For this part we can use a similar approach but the probability for X_i becomes $1/(n-i+1)^2$. This is because the dirtiest shirt falls into the dirtiest spot with probability $1/(n-i+1)$ and we pick it after that with probability $1/(n-i+1)$, so the probability of picking the dirtiest shirt from the dirtiest spot for the ith shirt is $1/(n-i+1)^2$. Using the same approach, we get the following sum:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} (n-i+1)^2 = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

2 Such High Expectations

Note 19

Suppose X and Y are independently drawn from a Geometric distribution with parameter p.

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

(b) Compute $\mathbb{E}[\max(X,Y)]$.

Solution:

(a) By independence,

$$\mathbb{P}[\min(X,Y) \geq t] = \mathbb{P}[X \geq t]\mathbb{P}[Y \geq t] = (1-p)^{2(t-1)}.$$

By Tail Sum,

$$\mathbb{E}[\min(X,Y)] = \sum_{t=1}^{\infty} \mathbb{P}[\min(X,Y) \geq t] = \sum_{t=1}^{\infty} (1-p)^{2(t-1)} = \frac{1}{1-(1-p)^2}.$$

(b) We see that

$$\begin{aligned}
\mathbb{P}[\max(X,Y) \geq t] = 1 - \mathbb{P}[\max(X,Y) < t] &= 1 - \mathbb{P}[X < t]\mathbb{P}[Y < t] \\
&= 1 - (1 - \mathbb{P}[X \geq t])(1 - \mathbb{P}[Y \geq t]) \\
&= 1 - \left(1 - (1-p)^{t-1}\right)\left(1 - (1-p)^{t-1}\right) \\
&= 1 - \left(1 - 2(1-p)^{t-1} + (1-p)^{2(t-1)}\right) \\
&= 2(1-p)^{t-1} - (1-p)^{2(t-1)}.
\end{aligned}$$

Using the result from part (a),

$$\begin{aligned}
\mathbb{E}[\max(X,Y)] &= \sum_{t=1}^{\infty} \mathbb{P}[\max(X,Y) \geq t] \\
&= \sum_{t=1}^{\infty} 2(1-p)^{t-1} - (1-p)^{2(t-1)} \\
&= \sum_{t=1}^{\infty} 2(1-p)^{t-1} - \sum_{t=1}^{\infty} (1-p)^{2(t-1)} \\
&= \frac{2}{p} - \frac{1}{1-(1-p)^2}.
\end{aligned}$$

Alternate Solution: An extremely elegant one-liner with linearity:

$$\mathbb{E}[\max(X,Y)] = \mathbb{E}[X+Y-\min(X,Y)] = \mathbb{E}[X]+\mathbb{E}[Y]-\mathbb{E}[\min(X,Y)] = \frac{2}{p} - \frac{1}{1-(1-p)^2}.$$

3 Dice Variance

Note 16

(a) Let X be a random variable representing the outcome of the roll of one fair 6-sided die. What is $\mathrm{Var}(X)$?

(b) Let Z be a random variable representing the average of n rolls of a fair 6-sided die. What is $\mathrm{Var}(Z)$?

Solution:

(a) Recall that $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$. We can compute each of the individual terms using the definition of expectation:

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} \\
\mathbb{E}\left[X^2\right] &= \frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right) \\
&= \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6}
\end{aligned}$$

Now, we plug back into the variance expression:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \\
&= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}
\end{aligned}$$

(b) Because each die roll is independent of the others, we can utilize the fact that for independent random variables X and Y, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. Let X_i be a random variable representing the outcome of the ith dice roll. We now have:

$$\begin{aligned}
\mathrm{Var}(Z) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}(X_i) && (X_i\text{'s are independent}) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \frac{35}{12} && (\text{from (a)}) \\
&= \left(\frac{1}{n}\right)^2 \cdot n \cdot \frac{35}{12} = \frac{35}{12n}
\end{aligned}$$

4 Covariance

Note 16

(a) We have a bag of 5 red and 5 blue balls. We take two balls uniformly at random from the bag without replacement. Let X_1 and X_2 be indicator random variables for the events of the first and second ball being red, respectively. What is $\mathrm{cov}(X_1, X_2)$? Recall that $\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$.

(b) Now, we have two bags A and B, with 5 red and 5 blue balls each. Draw a ball uniformly at random from A, record its color, and then place it in B. Then draw a ball uniformly at random from B and record its color. Let X_1 and X_2 be indicator random variables for the events of the first and second draws being red, respectively. What is $\mathrm{cov}(X_1, X_2)$?

Solution:

(a) We can use the formula $\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2]$.

$$\begin{aligned}
\mathbb{E}[X_1] &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2}, \\
\mathbb{E}[X_2] &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2}, \\
\mathbb{E}[X_1X_2] &= \frac{5}{10} \cdot \frac{4}{9} \times 1 + \left(1 - \frac{5}{10} \cdot \frac{4}{9}\right) \times 0 = \frac{2}{9}.
\end{aligned}$$

Therefore,

$$\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\,\mathbb{E}[X_2] = \frac{2}{9} - \frac{1}{2} \times \frac{1}{2} = -\frac{1}{36}.$$

(b) Again, we use the formula $\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$.

$$\begin{aligned}
\mathbb{E}[X_1] &= \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2} \\
\mathbb{E}[X_2] &= \left(\frac{5}{10} \times \frac{6}{11} + \frac{5}{10} \times \frac{5}{11}\right) \times 1 + \left(\frac{5}{10} \times \frac{5}{11} + \frac{5}{10} \times \frac{6}{11}\right) \times 0 = \frac{1}{2} \\
\mathbb{E}[X_1 X_2] &= \frac{5}{10} \times \frac{6}{11} \times 1 = \frac{30}{110}.
\end{aligned}$$

Therefore,

$$\mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{30}{110} - \frac{1}{4} = \frac{1}{44}.$$

Note that in part (a), if one event happened, the other would be less likely to happen, and thus the covariance was negative. Similarly, in part (b), if one event happened, the other would be more likely to happen, and thus the covariance was positive.