1 Elevator Variance

Note 16

A building has n upper floors numbered $1,2,\dots,n$, plus a ground floor G. At the ground floor, m people get on the elevator together, and each person gets off at one of the n upper floors uniformly at random and independently of everyone else. What is the *variance* of the number of floors the elevator *does not* stop at?

Solution: Let N be the number of floors the elevator does not stop at. We can represent N as the sum of the indicator variables $I_1,\dots,I_n$, where $I_i = 1$ if no one gets off on floor i. Thus, we have

$$\mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

and from linearity of expectation,

$$\mathbb{E}[N] = \sum_{i=1}^{n} \mathbb{E}[I_i] = n\left(\frac{n-1}{n}\right)^m.$$

To find the variance, we cannot simply sum the variance of our indicator variables. However, since $\text{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2$ the only piece we don't already know is $\mathbb{E}[N^2]$. We can calculate this by again expanding N as a sum:

$$\mathbb{E}[N^2] = \mathbb{E}[(I_1 + \cdots + I_n)^2] = \mathbb{E}\left[\sum_{i,j} I_i I_j\right] = \sum_{i,j} \mathbb{E}[I_i I_j] = \sum_{i}^{n} \mathbb{E}[I_i^2] + \sum_{i \neq j} \mathbb{E}[I_i I_j].$$

The first term is simple to calculate: since I_i is an indicator, $I_i^2 = I_i$, so we have

$$\mathbb{E}[I_i^2] = \mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \left(\frac{n-1}{n}\right)^m,$$

meaning that

$$\sum_{i=1}^{n} \mathbb{E}\left[I_i^2\right] = n\left(\frac{n-1}{n}\right)^m.$$

From the definition of the variables I_i, we see that $I_i I_j = 1$ when both I_i and I_j are 1, which means no one gets off the elevator on floor i and floor j. This happens with probability

$$\mathbb{P}[I_i = I_j = 1] = \mathbb{P}[I_i = 1 \cap I_j = 1] = \left(\frac{n-2}{n}\right)^m.$$

Thus we now know

$$\sum_{i \neq j} \mathbb{E}[I_i I_j] = n(n-1)\left(\frac{n-2}{n}\right)^m,$$

and we can assemble everything we've done so far to see that

$$\mathrm{Var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = n\left(\frac{n-1}{n}\right)^m + n(n-1)\left(\frac{n-2}{n}\right)^m - n^2\left(\frac{n-1}{n}\right)^{2m}.$$

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

$$\begin{aligned}
\mathbb{E}[X_1] &= \frac{5}{10}\times 1 + \frac{5}{10}\times 0 = \frac{1}{2} \\
\mathbb{E}[X_2] &= \left(\frac{5}{10}\times\frac{6}{11} + \frac{5}{10}\times\frac{5}{11}\right)\times 1 + \left(\frac{5}{10}\times\frac{5}{11} + \frac{5}{10}\times\frac{6}{11}\right)\times 0 = \frac{1}{2} \\
\mathbb{E}[X_1X_2] &= \frac{5}{10}\times\frac{6}{11}\times 1 = \frac{30}{110}.
\end{aligned}$$

Therefore,

$$\mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \frac{30}{110} - \frac{1}{4} = \frac{1}{44}.$$

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.

3 Number Game

Note 20

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 $\mathbb{E}[V \mid S = s]$, 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 $\mathbb{E}[V \mid S = s] = \frac{s+100}{2}$.

(c) With the law of total expectation, we have that

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

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 $\mathbb{E}[V] = 75$.