1 Expectation and Variance

This problem will give you some practice calculating expectations and variances of random variables. Suppose that the random variable X takes on 3 values, $10, 25, 70$. Suppose $\mathbb{P}[X = 10] = 0.5$, $\mathbb{P}[X = 25] = 0.2$, and $\mathbb{P}[X = 70] = 0.3$.

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

(b) What is $\mathbb{E}[X^2]$?

(c) What is $\text{var}(X)$?

Solution:

(a) $10 \cdot 0.5 + 25 \cdot 0.2 + 70 \cdot 0.3 = 31$.

(b) $10^2 \cdot 0.5 + 25^2 \cdot 0.2 + 70^2 \cdot 0.3 = 1645$

(c) $\mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1645 - 31^2 = 684$

2 Diversify Your Hand

You are dealt 13 cards from a standard 52 card deck. Let X be the number of distinct values in your hand. For instance, the hand (A, A, A, 2, 3, 4, 4, 5, 7, 9, 10, J, J) has 9 distinct values.

(a) Calculate $E[X]$.

(b) Calculate $Var[X]$.

Solution:

(a) Let X_i be the indicator of the ith value appearing in your hand. Then, $X = X_1 + X_2 + \ldots + X_13$ (Let 13 correspond to K, 12 correspond to Q, 11 correspond to J). By linearity of expectation then, $E[X] = \sum_{i=1}^{13} E[X_i]$. We can calculate $\mathbb{P}[X_i = 1]$ by taking the complement, $1 - Pr[X_i = 0]$, or 1 minus the probability that the card does not appear in your hand. This is $1 - \frac{\binom{48}{13}}{\binom{52}{13}}$. Then, $E[X] = 13\mathbb{P}[X_1 = 1] = 13(1 - \frac{\binom{48}{13}}{\binom{52}{13}})$.

(b) To calculate variance, since the indicators are not independent, we have to use the formula $E[X^2] = \sum_{i=j} E[X_i^2] + \sum_{i \neq j} E[X_i X_j]$.

$\sum_{i=j} E[X_i^2] = \sum_{i=j} E[X_i] = 13(1 - \frac{\binom{48}{13}}{\binom{52}{13}})$

To calculate $\mathbb{P}[X_i X_j = 1]$, we note that $\mathbb{P}[X_i X_j = 1] = 1 - \mathbb{P}[X_i = 0] - \mathbb{P}[X_j = 0] + \mathbb{P}[X_i = 0, X_j = 0]$.

$\sum_{i \neq j} E[X_i X_j] = 13 \cdot 12 \mathbb{P}[X_i X_j = 1] = 13 \cdot 12(1 - \mathbb{P}[X_i = 0] - \mathbb{P}[X_j = 0] + \mathbb{P}[X_i = 0, X_j = 0])$
$= 156(1 - 2\frac{\binom{48}{13}}{\binom{52}{13}} + \frac{\binom{44}{13}}{\binom{52}{13}})$

Putting it all together, we have $Var[X] = E[X^2] - E[X]^2 = 13(1 - \frac{\binom{48}{13}}{\binom{52}{13}}) + 156(1 - 2\frac{\binom{48}{13}}{\binom{52}{13}} + \frac{\binom{44}{13}}{\binom{52}{13}}) - (13(1 - \frac{\binom{48}{13}}{\binom{52}{13}}))^2$

3 Rolling Dice

(a) If we roll a fair 6-sided die, what is the expected number of times we have to roll before we roll a 6? What is the variance?

(b) Suppose we have two independent, fair n-sided dice labeled Die 1 and Die 2. If we roll the two dice until the value on Die 1 is smaller than the value on Die 2, what is the expected number of times that we roll? What is the variance?

(c) Let $n = 6$, so we are back to fair 6-sided die. Suppose we roll Die 1 until a 6 comes up, and we roll Die 2 until a 6 comes up. Let X be a random variable representing the number of times Die 1 is rolled before getting a 6, and let Y be the corresponding random variable for Die 2. Compute $\mathbb{P}[\min(X,Y) = n]$ and $\mathbb{P}[X + Y = n]$, where n is an integer.

Solution:

(a) Since rolling a die until obtaining a 6 involves independent rolls with a constant probability of success per roll, the expected number of times we roll follows a geometric distribution.

This question seeks to review basic formulas for the geometric distribution. The probability of rolling a 6 is $1/6$. Recall that the expectation is the inverse of the probability, and that the variance is $(1-p)/p^2$. Thus the expectation is $1/(1/6) = 6$ rolls. Thus the variance is $(1-p)/p^2 = (1-1/6)/(1/6)^2 = 30$ rolls^2.

(b) If we roll the two dice, three outcomes are possible: the two dice show the same number, Die 1 is greater than Die 2, or Die 2 is greater than Die 1. The last two events occur with the same likelihood and the first event occurs with chance $n/n^2 = 1/n$, since there are n^2 possible rolls and n different numbers for which there could to be duplicates. Thus the number of ways that Die 1 is smaller than Die 2 on a given roll is $(n^2 - n)/2$, so the probability that this occurs on a given roll is $(n^2 - n)/(2n^2) = 1/2 - 1/(2n)$.

The expected number of times we roll is therefore geometrically distributed with

$$p = \frac{1}{2} - \frac{1}{2n}.$$

Plugging this into the formulas for expectation and variance yields the answer.

(c) To compute $\mathbb{P}[\min(X,Y) = n]$, we notice that $\min(X,Y) = n$ corresponds to $\{X = n \text{ and } Y \geq n\}$ or $\{X \geq n \text{ and } Y = n\}$. Using the inclusion-exclusion principle, we have:

$$\begin{aligned}
\mathbb{P}[\min(X,Y) = n] &= \mathbb{P}[\{X = n\} \cap \{Y \geq n\}] + \mathbb{P}[\{X \geq n\} \cap \{Y = n\}] - \mathbb{P}[\{X = n\} \cap \{Y = n\}] \\
&= \mathbb{P}[\{X = n\}] \cdot \mathbb{P}[\{Y \geq n\}] + \mathbb{P}[\{X \geq n\}] \cdot \mathbb{P}[\{Y = n\}] \\
&\quad - \mathbb{P}[\{X = n\}] \cdot \mathbb{P}[\{Y = n\}] \quad \text{(independence)}
\end{aligned}$$

To compute $\mathbb{P}[X \geq n]$ (which is also $\mathbb{P}[Y = n]$, by symmetry), we have:

$$\begin{aligned}
\mathbb{P}[X \geq n] &= \sum_{i \geq n} \mathbb{P}[X = i] \\
&= \sum_{i \geq n} \left[\left(\frac{5}{6}\right)^{i-1} \cdot \left(\frac{1}{6}\right) \right] \\
&= \frac{\left[\left(\frac{5}{6}\right)^{n-1} \cdot \left(\frac{1}{6}\right) \right]}{\left(1 - \frac{5}{6}\right)} \quad \text{(sum of infinite geometric series)} \\
&= \left(\frac{5}{6}\right)^{n-1}
\end{aligned}$$

We can now plug this value into our probability above.

$$\begin{aligned}
\mathbb{P}[\min(X,Y) = n] &= \mathbb{P}[\{X = n\}] \cdot \mathbb{P}[\{Y \geq n\}] + \mathbb{P}[\{X \geq n\}] \cdot \mathbb{P}[\{Y = n\}] \\
&\quad - \mathbb{P}[\{X = n\}] \cdot \mathbb{P}[\{Y = n\}] \\
&= \left(\frac{5}{6}\right)^{n-1} \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^{n-1} + \left(\frac{5}{6}\right)^{n-1} \cdot \left(\frac{5}{6}\right)^{n-1} \cdot \left(\frac{1}{6}\right) \\
&\quad - \left(\frac{5}{6}\right)^{n-1} \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^{n-1} \cdot \left(\frac{1}{6}\right) \\
&= 2 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^{2(n-1)} - \frac{1}{36} \cdot \left(\frac{5}{6}\right)^{2(n-1)} \\
&= \left(\frac{5}{6}\right)^{2(n-1)} \cdot \left(\frac{2}{6} - \frac{1}{36}\right) \\
&= \left(\frac{25}{36}\right)^{(n-1)} \cdot \frac{11}{36}
\end{aligned}$$

This looks suspiciously like a probability from a geometric distribution...surprising or not?

This brings us to an alternative approach for computing $\mathbb{P}[\min(X,Y) = n]$, we can define a new random variable $Z = \min(X,Y)$. $Z = n$ corresponds to the event that either Die 1 or Die 2 is a

6 at their n-th rolls, *and* neither Die 1 nor Die 2 is a 6 in any previous roll. Note that the events $\{$Die 1 or Die 2 are 6$\}$ and $\{$Die 1 and Die 2 are not 6$\}$ are complements, so we can correspond them to "success" and "failure" events. Thus, Z is also a geometric random variable, whose success probability is $\mathbb{P}[\text{Die 1 is 6, or Die 2 is 6}] = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36}$. Using the distribution of a geometric random variable, we have $\mathbb{P}[Z = n] = (\frac{25}{36})^{n-1}(\frac{11}{36})$.

To compute $\mathbb{P}[X+Y] = n$, we notice that there are $(n-1)$ combinations of $X = i$ and $Y = n-i$ such that $X+Y = n$. Each combination of $X = i$ and $Y = n-i$ correspond to disjoint events. Therefore:

$$\begin{aligned}
\mathbb{P}[X+Y=n] &= \sum_{i=1}^{n-1} \mathbb{P}[\{X=i\} \cap \{Y=n-i\}] \\
&= \sum_{i=1}^{n-1} \mathbb{P}[\{X=i\}] \cdot \mathbb{P}[\{Y=n-i\}] \quad \text{(independence)} \\
&= \sum_{i=1}^{n-1} \left[\left(\frac{5}{6}\right)^{i-1} \cdot \left(\frac{1}{6}\right)\right] \left[\left(\frac{5}{6}\right)^{n-i-1} \cdot \left(\frac{1}{6}\right)\right] \\
&= \sum_{i=1}^{n-1} \left[\left(\frac{5}{6}\right)^{n-2} \cdot \left(\frac{1}{6}\right)^2\right] \\
&= \left(\frac{5}{6}\right)^{n-2} \cdot \left(\frac{1}{6}\right)^2 \cdot \sum_{i=1}^{n-1} 1 \\
&= \left(\frac{5}{6}\right)^{n-2} \cdot \left(\frac{1}{6}\right)^2 \cdot (n-1)
\end{aligned}$$

4 Fishy Computations

Use the Poisson distribution to answer these questions:

(a) Suppose that on average, a fisherman catches 20 salmon per week. What is the probability that he will catch exactly 7 salmon this week?

(b) Suppose that on average, you go to Fisherman's Wharf twice a year. What is the probability that you will go at most once in 2018?

(c) Suppose that in March, on average, there are 5.7 boats that sail in Laguna Beach per day. What is the probability there will be *at least* 3 boats sailing throughout the *next two days* in Laguna?

Solution:

(a) $X \sim \text{Poiss}(20)$.

$$\mathbb{P}[X=7] = \frac{20^7}{7!}e^{-20} \approx 5.23 \cdot 10^{-4}.$$

(b) $X \sim \text{Poiss}(2)$.

$$\mathbb{P}[X \le 1] = \frac{2^0}{0!}e^{-2} + \frac{2^1}{1!}e^{-2} \approx 0.41.$$

(c) Let Y be the number of boats that sail in the next two days. We can approximate Y as a Poisson distribution $Y \sim \text{Poiss}(\lambda = 11.4)$, where λ is the average number of boats that sail over two days. Now, we compute

$$\begin{aligned}
\mathbb{P}[Y \ge 3] &= 1 - \mathbb{P}[Y < 3] \\
&= 1 - \mathbb{P}[Y = 0 \cup Y = 1 \cup Y = 2] \\
&= 1 - (\mathbb{P}[Y = 0] + \mathbb{P}[Y = 1] + \mathbb{P}[Y = 2]) \\
&= 1 - \left(\frac{11.4^0}{0!}e^{-11.4} + \frac{11.4^1}{1!}e^{-11.4} + \frac{11.4^2}{2!}e^{-11.4}\right) \\
&\approx 0.999.
\end{aligned}$$

We can show what we did above formally with the following claim: the sum of two independent Poisson random variables is Poisson. We won't prove this, but from the above, you should intuitively know why this is true. Now, we can model sailing boats on day i as a Poisson distribution $X_i \sim \text{Poiss}(\lambda = 5.7)$. Now, let X_1 be the number of sailing boats on the next day, and X_2 be the number of sailing boats on the day after next. We are interested in $Y = X_1 + X_2$. Thus, we know $Y \sim \text{Poiss}(\lambda = 5.7 + 5.7 = 11.4)$.