Group #5 @1pm
Mimi Morales, Yiyi Zhu, Vi Pham, Brian Fong May 16^{th}, 2019

Lecture Worksheet 18

Task 1

An agent sells life insurance policies to five equally aged, healthy people. According to recent data, the probability of a person living in these conditions for 30 years or more is $\frac{2}{3}$. Calculate the probability that after 30 years.

- All five people are still living

  We can approach the randomness as a binomial distribution. If we denote X as the number of people still living, the probability is given by:

$$P(X=5) = \binom{5}{5}\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^0 \approx 13.2\%$$

- At least three people are still living

  The probability is given by:

$$\begin{aligned} P(X \geq 3) &= \sum_{i=3}^{5}\binom{5}{i}\left(\frac{2}{3}\right)^i\left(\frac{1}{3}\right)^{5-i} \\ &= \binom{5}{3}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 + \binom{5}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^1 + \binom{5}{5}\left(\frac{2}{3}\right)^5\left(\frac{1}{3}\right)^0 \\ &\approx 79\% \end{aligned}$$

- Exactly two people are still living

  The probability is given by:

$$\begin{aligned} P(X=2) &= \binom{5}{2}\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3 \\ &\approx 16.5\% \end{aligned}$$

Task 2

Let X, Y be random variables with $X\ B(n,p)$ and $Y\ B(n, 1-p)$

- Show that $P(X \le i) = P(Y \le n-i-1)$

  This equation should not hold actually. If we view p as the probability of success, then X denotes the number of successes whereas Y would then denote the number of failures.

  Because we conduct n total experiments, we must have the number of successes and failures add up to n by definition.

  $$X + Y = n$$

  However, looking at the given equation, we find that the maximum number of successes and failures, which are i and $n-i-1$, respectively, do not add up to n.

  $$max(X) + max(Y) = (i) + (n-i-1) = n-1 \neq n$$

  To patch this up, we provide the following logical argument.

  The probability of success for Y is $1-p$, implying that the probability of failure is p. This is the same as the probability of success for X. Thus, successes for X are analogous to the failures for Y; their distributions are identical.

  Using the fact that these two events are equivalent:

  $$\textit{At most } i \textit{ successes} \iff \textit{At least } n-i \textit{ failures}$$

  The LHS is simply represented by $X \le i$, and the RHS can be represented as $Y \ge n-i$. Since the probability distributions are identical, we can equate the following statement.

  $$P(X \le i) = P(Y \ge n-i)$$

- If you proved the identity by computing the two values and verifying that they coincide, give a logical explanation for its validity. If you gave a logical argument, carry out the calculations to verify the identity.

  To verify the identity mathematically, we can input the arguments into the probability mass function for X and Y.

For X, it is straightforward:

$$P(X \leq i) = \sum_{k=0}^{i} \binom{n}{k} \, p^k \, (1-p)^{n-k}$$

For Y, we use a change of variables to obtain an expression identical to the one for X:

$$\begin{align}
P(Y \geq n-i) &= \sum_{l=n-i}^{n} \binom{n}{l} \, (1-p)^l \, p^{n-l} \tag{1} \\
&= \sum_{k=i}^{0} \binom{n}{n-k} \, (1-p)^{n-k} \, p^k \tag{2} \\
&= \sum_{k=0}^{i} \binom{n}{k} \, p^k \, (1-p)^{n-k} \tag{3} \\
&= P(X \leq i) \tag{4}
\end{align}$$

Above we make the following steps:

1. *Plug into the formula*
2. *Change of variables: $l = n-k$*
3. *Since l started at $n-i$ / end at n,*
   then k must start at i / end at 0
4. *Since $\binom{n}{n-k} = \binom{n}{k}$, and the summation starting from i to 0 is the same as starting from 0 to i*

Task 3

Each game you play is a win with probability $p \in (0,1)$. You plan on playing 5 games, but if you win the 5th game, then you will keep on playing until you lose.

- Compute the expected number of games you play.

Let X denote the number of games we play. The expected number of games can be denoted as $E[X]$.

We can approach this problem by segmenting the gameplay into two parts, the first 4 games and the following game(s), denoted by X_A and x_B. Thus, the expected number of games can be given by:

$$E[X] = E[X_A] + E[X_B]$$

Since it is given that we plan on playing 5 games, we will play the first 4 games so then expected value for the first segment is:

$$E[X_A] = 4$$

For the second segment, the first game (5th in general) is definitely played, however the probability of playing the next game (6th in general) depends on whether we win or lose. Thus the probability of playing the 6th game is p. This pattern continues, the probability of playing the next game is always dependent on whether we win or lose the current game. Moreover, since the probability of winning is consistent for each game, we can treat the games as independent, allowing us to multiply the probabilities.

$$\begin{aligned} E[X_B] &= p(x_5) + p(x_6) + p(x_7) + ... \\ &= 1 + p + p^2 + ... \\ &= \sum_{k=0}^{\infty} p^k \\ &= \frac{1}{1-p} \end{aligned}$$

NOTE: The expected value is given by $E[X] = \sum x_i \cdot p(x_i)$, where x_i is the value for the i-th game and $p(x_i)$ is the probability of winning the i-th game. However, the "value" of *the i-th game is just 1 for all the games because $E[X]$ is the expected values of games that we play, which depends on whether or not we win, so thus the expression for $E[X_B]$ resembles a sum of probabilities because the values for each game is just 1.*

Therefore, the expected number of games we play is given by:

$$E[X] = E[X_A] + E[X_B] = 4 + \frac{1}{1-p}$$

- Compute the expected number of games you lose.

Let Y denote the number of games we lose. The probability of losing a game is $1-p$. We are guaranteed at least one loss for that is what determines when we stop playing games / conducting the experiment. We can also approach this problem by splitting the experiment into two segments such that:

$$E[Y] = E[Y_A] + E[Y_B]$$

The first segment is the first 4 games, in which we may or may not win. The expected value for these is given by:

$$\begin{aligned} E[Y_A] &= \sum_{i=0}^{4} 1 \cdot (1-p) \\ &= 4 \cdot (1-p) \end{aligned}$$

For the second segment, the 5th game and so on. We find that there is only 1 failure that exists in this segment, and it has to exist because we assume that the experiment ends and thus we have lost a game.

$$E[Y_B] = 1$$

Therefore the expected value of the number of games we lose is given by:

$$E[Y] = E[Y_A] + E[Y_B] = 4 \cdot (1-p) + 1$$