# APMA 1650 – Homework 4

Due Friday, October 6th, 2023

1. You have $m$ red marbles which you will put uniformly at random into $n$ boxes. What is the expected number of boxes which are empty after all $m$ marbles have been distributed.

   Let $X$ be the number of empty boxes. Since there is no essential difference between the boxes (we can swap them around without affecting the expected value), since we are looking for an expected value, and since the expected value involves counting these boxes, we will use the method of indicator random variables together with linearity of expectation. Define the indicator random variables $X_i, i = 1, \ldots, n$, where

$$X_i = \begin{cases} 0 & \text{box } i \text{ is not empty} \\ 1 & \text{box } i \text{ is empty} \end{cases}$$

   Then since $X$ counts the number of empty boxes, $X = X_1 + \cdots + X_n$. By linearity of expectation:

$$\mathbb{E}(X) = \mathbb{E}(X_1 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)$$

   Now we look at one of the terms in the sum on the right. Let's evaluate $\mathbb{E}(X_i)$ for some box $i$. By the definition of the expected value of a discrete random variable,

$$\begin{aligned} \mathbb{E}(X_i) &= 1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0) \\ &= \mathbb{P}(X_i = 1) \end{aligned}$$

   What is $\mathbb{P}(X_i = 1)$? $X_i = 1$ only if *every* red marble is thrown into a *different* box from box $i$. The marbles are thrown into the $n$ boxes uniformly at random. For each marble, there is a $1/n$ probability that it lands in box $i$ and thus a $(n - 1/n)$ chance it does not land in box $i$ (i.e. lands in one of the other $n - 1$ boxes). Since there are $m$ marbles and they are thrown independently, the chance that no marbles land in box $i$ is:

$$\left(\frac{n-1}{n}\right)^m$$

   Thus we have:

$$\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = \left(\frac{n-1}{n}\right)^m$$

   Summing all these up, we get:

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) \\ &= \underbrace{\left(\frac{n-1}{n}\right)^m + \cdots + \left(\frac{n-1}{n}\right)^m}_{n \text{ terms in sum, since } n \text{ boxes}} \\ &= n\left(\frac{n-1}{n}\right)^m \end{aligned}$$

2. Two people are playing a game. They take turns rolling a standard, fair six-sided die. The game ends when one player rolls a 6. The player who rolls the 6 is the winner of the game. A "round" of the game is defined as a single die roll.

   (a) What is the probability that the player who goes first wins?

   Let $S$ be the event that a 6 is rolled, and $N$ the event that a 6 is not rolled. We can represent each game as a string of 0 or more Ns followed by a single S. For the first player to win, the game must last an *odd* number of rounds, so one of the following game strings must occur:

| String | Probabiltiy |
|---|---|
| S | $1/6$ |
| NNS | $(5/6)^2(1/6)$ |
| NNNNS | $(5/6)^4(1/6)$ |
| $\vdots$ | $\vdots$ |

   To find the probability that the first player wins, we sum the probabilities in the right hand column above:

$$\mathbb{P}(\text{Player 1 wins}) = (1/6) + (1/6)(5/6)^2 + (1/6)(5/6)^4 + (1/6)(5/6)^6 + \cdots$$

   This is a geometric series with first element $1/6$ and common ratio $(5/6)^2 = 25/36$. Thus the probability that the first player wins is the sum of this geometric series.

$$\mathbb{P}(\text{Player 1 wins}) = \frac{1/6}{1 - 25/36} = \frac{1/6}{11/36} = \frac{6}{11}$$

   This is slightly greater than $1/2$, so if you are playing this game, you want to go first.

   (b) What is the expected value and variance for the number of rounds in the game?

   Here we do not care who win, just how long the game lasts. Let $X$ be the number of rounds the game lasts. Since we are rolling a die repeatedly until we roll a 6, this is a geometric distribution, with "success" defined as rolling a 6. Since the probability of rolling a 6 is $p = 1/6$, then $X \sim$Geometric(1/6). We can look up the expected value and variance of the geometric distribution to get:

$$\mathbb{E}(X) = \frac{1}{p} = \frac{1}{1/6} = 6$$
$$Var(X) = \frac{1-p}{p^2} = \frac{5/6}{1/36} = 30$$

   (c) What is the probability that the game lasts at least 4 rounds?

Using our geometric random variable $X$, we want $\mathbb{P}(X \geq 4)$. We could compute $\mathbb{P}(X < 4) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3)$ and subtract from 1. Or we could note that the game lasting at least 4 rounds is equivalent to "failure" on the first 3 rounds, i.e. not rolling a 6 on the first 3 rounds. So

$$\mathbb{P}(X \geq 4) = \mathbb{P}(\text{"failure" on first 3 rounds}) = (1-p)^3 = (5/6)^3$$

(d) Given the game has lasted four rounds, what is the probability that the game lasts at least 8 rounds?

The probability that the game lasts at least 8 rounds given that it has lasted 4 is $\mathbb{P}(X \geq 8 | X \geq 4)$. Using the definition of conditional probability, and noting that if a game lasts at least 8 rounds it has also lasted at least 4 rounds:

$$\mathbb{P}(X \geq 8 | X \geq 4) = \frac{\mathbb{P}(X \geq 8 \cap X \geq 4)}{\mathbb{P}(X \geq 4)} = \frac{\mathbb{P}(X \geq 8)}{\mathbb{P}(X \geq 4)}$$

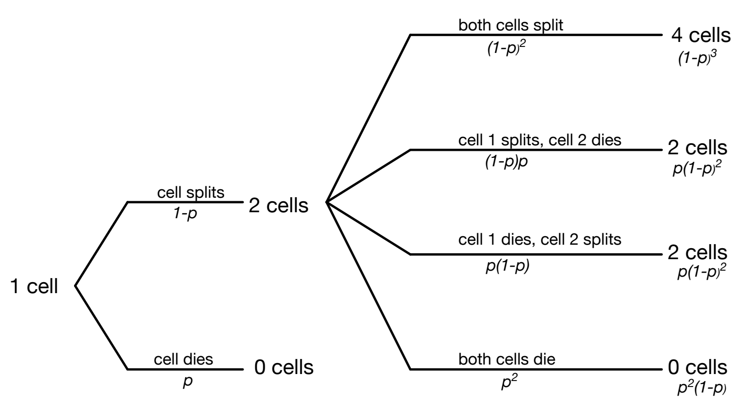
Using part (c), and extending it to the case where the game lasts at least 8 rounds, we have:

$$\mathbb{P}(X \geq 8 | X \geq 4) = \frac{(1-p)^7}{(1-p)^3} = (1-p)^4 = (5/6)^4$$

3. A single cell will die with probability $p$ or split into two cells with probability $1-p$, producing a second generation of cells. Each cell in the second generation (if there are any) will die or split into two with the same probabilities as the initial cell. You start with a single cell.

(a) Find the probability mass function (pmf) for the number of cells in the third generation.

To find the probability mass function, we can first draw a tree diagram to see what is happening.




Using the tree, we can write the pmf in a table. There are two branches in the tree which lead to 2 cells and two branches in the tree which lead to 0 cells. Don't forget to add them together! Let $Y$ be the number of cells in the 3rd generation. Then we have the following pmf for $Y$.

| $y$ | $p(y)$ |
|---|---|
| 0 | $p + p^2(1-p)$ |
| 2 | $2p(1-p)^2$ |
| 4 | $(1-p)^3$ |
| $\vdots$ | $\vdots$ |

(b) What is the expected value of the number of cells in the third generation?

Using the pmf table above and the formula for the expected value of a discrete random variable:

$$\begin{aligned}\mathbb{E}(Y) &= 0 \cdot (p + p^2(1-p)) + 2 \cdot 2p(1-p)^2 + 4 \cdot (1-p)^3 \\ &= 4p(1-p)^2 + 4(1-p)^3 \\ &= 4(1-p)^2(p + (1-p)) \\ &= 4(1-p)^2\end{aligned}$$

4. A particular flight on Peyamerican Airlines can only fit 200 people, but tickets were sold to 205 people. Suppose each ticket holder has a 0.05 probabiltiy of not showing up for the flight.

(a) What is the probability that the flight will be overbooked? Give an approximate numerical answer for this.

We can model this problem with a binomial distribution, since we have $n = 205$ customers, each of whom has (independently) a $p = 0.95$ probabiltiy of showing up. (Again, think of whether this is realistic; if a family travels together, then it is likely that they will either all show up or all not show up, so we might not have complete independence; however, in this case, it is good enough for our purposes). Defining "success" as showing up for the flight, let $X \sim$Binomial(205, 0.95). The probability that the flight will be overbooked is:

$$\begin{aligned}\mathbb{P}(X > 200) &= \mathbb{P}(X = 201) + \mathbb{P}(X = 202) + \mathbb{P}(X = 203) + \mathbb{P}(X = 204) + \mathbb{P}(X = 205) \\ &= \binom{205}{201}0.95^{201}0.05^4 + \binom{205}{202}0.95^{202}0.05^3 + \binom{205}{203}0.95^{203}0.05^2 \\ &\quad + \binom{205}{204}0.95^{204}0.05^1 + \binom{205}{205}0.95^{205}0.05^0 \\ &\approx 0.022\end{aligned}$$

So it is highly unlikely that the flight will be overbooked.

(b) What is the expected number of people that show up for the flight?

Using the formula for the expected value of a binomial random variable:

$$\mathbb{E}(X) = np = 205(0.95) = 194.75$$

(c) What is the variance of the number of people who show up for the flight?

Using the formula for the variance of a binomial random variable:

$$\mathbb{E}(X) = np(1-p) = 205(0.95)(0.05) = 9.7375$$

5. You reprise your role as the quality control manager for the Acme Widget Company. You have found that in every box of 100 widgets there is on average 1 defective widget.

(a) Model this problem with an appropriate probability distribution. What is the probability that a box of 100 widgets contains 2 or fewer defective widgets? Also give an approximate numerical answer for this.

Defining "success" as finding a defective widget, since we are looking for the number of successes out of $n = 100$ Bernoulli trials, we can model this with a Binomial random variable. The probabiltiy of finding a defective widget is $p = 1/100 = 0.01$. Let $X$ ~Binomial(100, 0.01). Then

$$\begin{aligned}\mathbb{P}(X \leq 2) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)\\ &= \binom{100}{0}0.01^0 0.99^{100} + \binom{100}{1}0.01^1 0.99^{99} + \binom{100}{2}0.01^2 0.99^{98}\\ &\approx 0.920627\end{aligned}$$

(b) **(optional)** Approximating this with a Poisson distribution, find the probability that you have 2 or fewer defective widgets.

We need a Poisson random variable with the same mean as the Binomial random variable in (a). Using the formula for the expected value of a Binomial random variable, $\mathbb{E}(X) = np = (100)(0.01) = 1$. Let $Y$ ~Poisson(1).

$$\begin{aligned}\mathbb{P}(Y \leq 2) &= \mathbb{P}(Y = 0) + \mathbb{P}(Y = 1) + \mathbb{P}(Y = 2)\\ &= \frac{e^{-1}1^0}{0!} + \frac{e^{-1}1^1}{1!} + \frac{e^{-1}1^2}{2!}\\ &\approx 0.919699\end{aligned}$$

(c) **(optional)** Evaluating both part (a) and part (b) numerically using Wolfram Alpha or your favorite software package, what is the relative error |True value - Approximate value| / True value for your Poisson approximation?

We computed the approximate values above. They look pretty close. How close are they?

$$\text{Relative error} = \frac{|0.920627 - 0.919699|}{0.919699} \approx 0.001$$

So the Poisson approximation is accurate within 0.1%, which is fantastic.