Worksheet 13 (Solutions)

1. Let X and Y be independent random variables with the following probability mass functions:

$$p_X(x) = \begin{cases} 0.2, & \text{if } x = 1 \\ 0.5, & \text{if } x = 2 \\ 0.3, & \text{if } x = 3 \end{cases} \quad p_Y(y) = \begin{cases} 0.7, & \text{if } y = 1 \\ 0.2, & \text{if } y = 2 \\ 0.1, & \text{if } y = 3 \end{cases}$$

For X, find the expected value, the variance and the moment generating function. Sketch the cdf of X. What are the expected value of Y and the expected value of $X + Y$?

Solution:

The expected value is:

$$\begin{aligned} \mathbb{E}X &= 0.2 \cdot 1 + 0.5 \cdot 2 + 0.3 \cdot 3 \\ &= 2.1 \end{aligned}$$

To get the variance, we first get the second raw moment:

$$\begin{aligned} \mathbb{E}X^2 &= 0.2 \cdot 1^2 + 0.5 \cdot 2^2 + 0.3 \cdot 3^2 \\ &= 4.9 \end{aligned}$$

Which gives the variance as:

$$\begin{aligned} Var(X) &= \mathbb{E}X^2 - [\mathbb{E}X]^2 \\ &= 4.9 - 2.1^2 = 0.49 \end{aligned}$$

The moment generating function is the expected value of e^{tX}. It does not have a fancy form here; it is just a sum of three elements:

$$\begin{aligned} \mathbb{E}[e^{tX}] &= \sum_x \mathbb{P}[X = x] \cdot e^{tx} \\ &= 0.2 \cdot e^t + 0.5 \cdot e^{2t} + 0.3 \cdot e^{3t} \end{aligned}$$

The CDF is just a step function with jumps at 1, 2, and 3, where the jump sizes are equal to the probabilities. This gives step values of 0 on $(-\infty, 1)$, 0.2 on $[1, 2)$, 0.7 on $[2, 3)$ and 1 for $[3, \infty)$.

The expected value of Y is:

$$\begin{aligned} \mathbb{E}Y &= 0.7 \cdot 1 + 0.2 \cdot 2 + 0.1 \cdot 3 \\ &= 1.4 \end{aligned}$$

And so, the expected value of $X + Y$ is:

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 2.1 + 1.4 = 3.5$$

2. Let X and Y be defined as in the previous question. Let $Z = max(X, Y)$. Write the pmf and sketch the cdf of Z.[1]

Solution:

The pmf is given by:

$$\begin{aligned}
\mathbb{P}[Z=1] &= \mathbb{P}[(X=1)\cap(Y=1)] = 0.2\cdot 0.7 = 0.14\\
\mathbb{P}[Z=2] &= \mathbb{P}[(X=1)\cap(Y=2)] + \mathbb{P}[(X=2)\cap(Y=1)] + \mathbb{P}[(X=2)\cap(Y=2)]\\
&= 0.2\cdot 0.2 + 0.5\cdot 0.7 + 0.5\cdot 0.2 = 0.49\\
\mathbb{P}[Z=3] &= 1 - \mathbb{P}[Z=1] - \mathbb{P}[Z=2] = 1 - 0.14 - 0.49 = 0.37
\end{aligned}$$

The cdf again is just a step function with jumps at 1, 2, and 3 where the size of the jump is given by the probability.

3. Let $X_1, X_2, \overset{\text{i.i.d.}}{\sim} Bernoulli(p)$ be an infinite sequence of independent random variables. We say that Y follows the *silly geometric* distribution if Y counts the number of X_i's that are zero before the first X_i that is a one. What are the pmf, expected value, and variance of Y?[2]

Solution: This is just a different parameterization of the geometric distribution, where the silly geometric is one less than the normal geometric (the normal one counts the last 1 in the tally). The pmf is the same but with y in place of $x-1$ from the value in the table. You could also re-derive it from the counting rules (what the probability of y zeros followed by a one).

$$\mathbb{P}[Y=y] = (1-p)^y p$$

In terms of the random variable, it will have an expected value 1 less from the normal geometric and the same variance (linear shifts do not affect the variance). Formally, if $T \sim Geom(p)$, then we could write Y in terms of T as $Y = T - 1$. Then:

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}[T-1] = \frac{1}{p} - 1 = \frac{1-p}{p}\\
Var[Y] &= Var[T-1] = \frac{1-p}{p^2}
\end{aligned}$$

Where the values for the expected value and variance of Y come from the table of distributions.

[1] In general, you need to work out all nine possible combinations of X and Y. A shortcut is to realize that you don't need to figure out the mass at 3 directly since it is one minus the mass at 1 and 2.

[2] For this and two following questions, the trick is to write the variable of interest in terms of a known distribution.

4. Let $X_1, \ldots, X_n$ a sequence of independent random variables where each X_i has a probability equal to $1/2$ of being $+1$ and probability of $1/2$ of being -1. Let W be the sum of the X_i's. What are the expected value and variance of W?

Solution: There are number of different approaches here. Probably the easiest is to find the expected value and variance and X_i (they are all the same) using the approach from question one. We have:

$$\begin{aligned} \mathbb{E}X_i &= (1) \cdot 0.5 + (-1) \cdot 0.5 = 0 \\ \mathbb{E}X_i^2 &= (1)^2 \cdot 0.5 + (-1)^2 \cdot 0.5 = 1 \\ Var(X_i) &= \mathbb{E}X_i^2 - [\mathbb{E}X_i]^2 = 1 - 0 = 1 \end{aligned}$$

Then, we see:

$$\begin{aligned} \mathbb{E}W &= \mathbb{E}\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \mathbb{E}X_i = 0 \\ Var(W) &= Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} (1) = n \end{aligned}$$

5. Let $X \sim Bin(n,p)$ and $Y = X/n$. Find the expected value and variance of Y. What are the limits of these two quantities as $n \to \infty$?

Solution: The expected value is just:

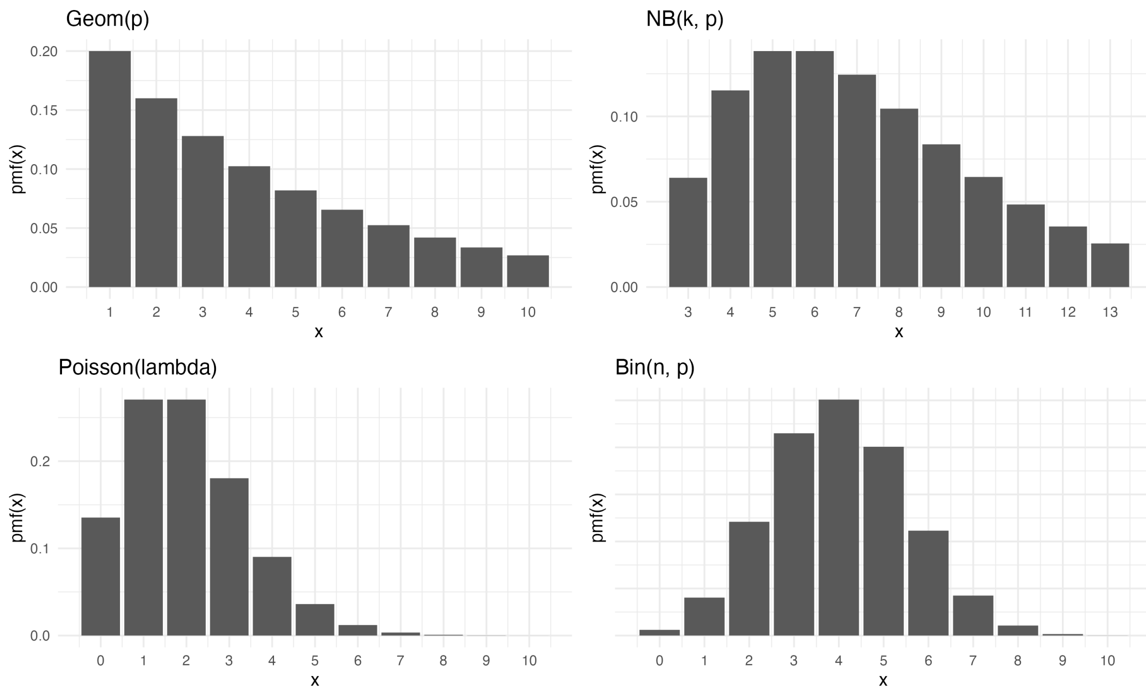
$$\begin{aligned} \mathbb{E}Y &= \mathbb{E}(X/n) \\ &= \frac{1}{n} \cdot \mathbb{E}X \\ &= \frac{np}{n} \\ &= p \end{aligned}$$

And, using the previous result from (1), the variance is just:

$$\begin{aligned} Var(Y) &= Var(X/n) \\ &= \frac{1}{n^2} Var(X) \\ &= \frac{np \cdot (1-p)}{n^2} \\ &= \frac{p \cdot (1-p)}{n} \end{aligned}$$

In the limit, the expected value goes to p and the variance goes to zero. So, with a large sample size, the number of successes limits to p, which makes quite a lot of sense.

6. Four plots of probability mass functions are given on the following page. The Binomial shows the whole pmf; the others truncate the values for larger x. Estimate as best as possible the unknown parameters for each of the four distributions. Hint: You can use the range of the data to figure out one of the parameters for the two-parameter distributions. The mode is a good way to estimate the value of p for the Binomial. For the others, pick one value of the pmf and solve given the pmf formula.



Solution: For the geometric, we just use the first value. The pmf of 1 is always p; here the pmf is 0.2 so therefore $p = 0.2$.

The smallest possible value for the negative binomial is k, so we know that $k = 3$ in this case. The smallest value of the negative binomial is always p^k. Here the smallest value is about 0.065 and therefore $p \approx (0.065)^{1/3} \approx 0.402$. I actually used a value of 0.4, but anything with this kind of logic is fine.

Exact same logic with the Poisson. The value at zero is always $e^{-\lambda}$. It looks like this is about .14 and therefore $\lambda \approx -\log(.14) \approx 1.96$. The actual value is 2; anything remotely close that uses this approach is fine.

We cannot use the same trick from the others because it this one has no scale on the y-axis. The binomial distribution has a positive pmf on the integers 0 through n. Given that the plot goes from 0 to 10, we can deduce that $n = 10$. The binomial also will have its mode (most frequent value) at np (it's mean) rounded to the nearest integer. Since the tallest peak is at 4, this means that p must be between .35 and .45. Anything in that range is fine; a guess of 0.4 would be most natural (and is the one that I used).