Example 3d

$n = 120$: $n_1 = 36$, $n_2 = 40$, $n_3 = 44$. When the buses arrive, one of the 120 is chosen at random.

Define X as the number of students on bus that the student was chosen from. Find EX. Here, let's not be confused by the numbers. We have a discrete random variable X, and its range has 3 values: $R_X = \{36, 40, 44\}$. In this question, their probabilities are determined by their frequencies:

$$P(X = 36) = \frac{36}{120}$$

and likewise for the other two buses. Hence by definition:

$$EX = 36 \cdot \frac{36}{120} + 40 \cdot \frac{40}{120} + 44 \cdot \frac{44}{120}$$

$$EX = 40.2\bar{6}$$

This may come as a surprising result. If we look at the average number of kids per bus, that would give us 40. But by the way we defined X, taking the traditional mean as an estimator for our expectation would give us a biased result (it would be underestimating). This is an important concept in statistics. However, Stats2MB3 will deal with this, so for now, we will focus on probability.

Example 7b

A machine fails with $p = .1$, find the probability that in a sample of $n = 10$, at most 1 item is defective. Compare the methods of binomial and Poisson. *What is the difference?*

with binomial distribution

P(at least 1)=P(0 or 1):

$$= \binom{10}{0}.9^{10} + \binom{10}{1}.1 \cdot .9^9 = .7361$$

This method gives the exact answer.

with Poisson

Recall, $\lambda = np \rightarrow= 1$. So P(at most 1)

$$\approx \frac{e^{-1}1^0}{0!} + \frac{e^{-1}1^1}{1!} = e^{-1} + e^{-1} = 2e^{-1} = .7358$$

The Poisson approximates the binomial and does a good job when we have n large, and p small. The Poisson is actually the limit of the binomial, if we let $n \to \infty$. See page 136 for an explanation.

Example 8c

For $X \sim Geometric(p)$, find $VarX$.

$$VarX = E[(X - EX)^2] = E[X^2] - (EX)^2$$

Simple, but important exercise: Make sure you can prove the above line.

Hence, we need the first and second raw moments (EX and EX^2). Let's assume nothing is given. Firstly, let's write our distribution:

$$f_X(x) = p(1-p)^{x-1}$$

Another side note, there is not *one unique* "geometric distribution" . . . It all depends how it is defined/understood. For example, X can be defined as number of failures before a success, or just number of trials total to get a success (as is defined there above).

$$EX = \sum_{x=1}^{\infty} xp(1-p)^{x-1}$$
$$= p\sum_{x=1}^{\infty} x(1-p)^{x-1}$$

At this point, we look stuck, but you may notice that the above looks like a derivative of $(1-p)^x$ . . . See below, using first year calculus (maybe this will make it obvious why it's called the geometric distribution if it wasn't clear already):

$$\sum_{x=0}^{\infty} p^x = \frac{1}{1-p}, \text{ for } |p| < 1$$

$$\Rightarrow \frac{d}{dp}\sum_{x=0}^{\infty} p^x = \frac{d}{dp}\frac{1}{1-p}, \text{ for } |p| < 1$$

$$\sum_{x=1}^{\infty} xp^{x-1} = \frac{1}{(1-p)^2}, \text{ for } |p| < 1$$

The summation index can start from $x = 1$ (verify why). Let's use this above fact to complete our answer:

$$p\sum_{x=1}^{\infty} x(1-p)^{x-1} = p\frac{1}{(1-(1-p))^2}$$
$$EX = \frac{p}{p^2} = \frac{1}{p}$$

With the first raw moment in hand, let's solve for the second raw moment. The preferred method for doing this would be by method of *Moment Generating Function*, but we'll save that for another day. Now we'll solve for it using chapter 4's ideas:

$$E[X^2] = \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1}$$

$$= p\sum_{x=1}^{\infty} x^2(1-p)^{x-1}$$

There are different ways to prove this, but there is no easy way out. It'll require a some sort of clever idea. The text shows one of the straightforward proofs:

$$= p\left[\sum_{x=1}^{\infty}((x-1)+1)^2(1-p)^{x-1}\right]$$

$$= p\left[\sum_{x=1}^{\infty}(x-1)^2(1-p)^{x-1} + 2\sum_{x=1}^{\infty}(x-1)(1-p)^{x-1} + \sum_{x=1}^{\infty}(1-p)^{x-1}\right]$$

Let $x-1=y$, allowing us to change our summation index.

$$= \sum_{y=0}^{\infty} y^2 p(1-p)^y + 2\sum_{y=0}^{\infty} yp(1-p)^y + \sum_{y=0}^{\infty} p(1-p)^y$$

$$= \sum_{y=0}^{\infty} y^2 p(1-p)^y + 2\sum_{y=1}^{\infty} yp(1-p)^y + \frac{p}{1-(1-p)}$$

$$= \sum_{y=0}^{\infty} y^2 p(1-p)^y + 2\sum_{y=1}^{\infty} yp(1-p)^y + 1$$

$$= (1-p)\sum_{y=0}^{\infty} y^2 p(1-p)^{y-1} + 2(1-p)\sum_{y=1}^{\infty} yp(1-p)^{y-1} + 1$$

Recall in the discrete case:

$$E[g(X)] = \sum_{x=0}^{n} g(x)\cdot f_X(x) = \sum_{x=0}^{n} g(x)\cdot P(X=x)$$

Thus, we have an equation in terms of EX^2 and EX, and since we already know EX we can solve for our desired result:

$$E[X^2] = (1-p)E[X^2] + 2(1-p)E[X] + 1$$

$$pE[X^2] = 2(1-p)\frac{1}{p} + 1$$

$$pE[X^2] = \frac{2-2p+p}{p}$$

$$E[X^2] = \frac{2-p}{p^2}$$

All that's left to do is to plug in and solve for Variance:

$$Var[X] = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{1-p}{p^2}$$

Theoretical Exercise 4.10

For $X \sim Binomial(n, p)$, show:

$$E\left[\frac{1}{X+1}\right] = \frac{1-(1-p)^{n+1}}{(n+1)p}$$

We have $g(X) = \frac{1}{X+1}$, so let's start by plugging that in, along with our probability mass function into the definition of expectation:

$$\begin{aligned}
E\left[\frac{1}{X+1}\right] &= \sum_{x=0}^{n} \frac{1}{x+1}\binom{n}{x}p^x(1-p)^{n-x} \\
&= \sum_{x=0}^{n} \frac{1}{(x+1)}\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \\
\textit{combine the x+1 into the factorial} &= \sum_{x=0}^{n} \frac{n!}{(x+1)!(n-x)!}p^x(1-p)^{n-x} \\
\textit{let's make a new factorial} &= \sum_{x=0}^{n} \frac{(n+1)n!}{(n+1)(x+1)!(n-x)!}p^x(1-p)^{n-x} \\
\textit{n+1 is not indexed by x} &= \frac{1}{n+1}\sum_{x=0}^{n} \frac{(n+1)!}{(x+1)!(n+1-(x+1))!}p^x(1-p)^{n-x} \\
&= \frac{1}{n+1}\sum_{x=0}^{n} \binom{n+1}{x+1}p^x(1-p)^{n-x} \\
&= \frac{1}{p(n+1)}\sum_{x=0}^{n} \binom{n+1}{x+1}p^{x+1}(1-p)^{n+1-(x+1)}
\end{aligned}$$

We are almost done, we can see we have the required denominator, but why is

$$\sum_{x=0}^{n} \binom{n+1}{x+1}p^{x+1}(1-p)^{n-x} = 1-(1-p)^{n+1}$$

One way to solve this is to manipulate your summation index (the "choose x+1" is awkward).

Let $x+1 = y$, (or $x = y-1$), which means our indices on the sum must change accordingly:

$$x = 0 \to y = 0+1 = 1$$

$$x = n \to y = n+1$$

Let's make those substitutions now:

$$\sum_{x=0}^{n} \binom{n+1}{x+1}p^{x+1}(1-p)^{n-x} = \sum_{y=1}^{n+1} \binom{n+1}{y}p^{y}(1-p)^{n-(y-1)}$$

$$= \sum_{y=1}^{n+1} \binom{n+1}{y}p^{y}(1-p)^{n+1-y}$$

This is almost in the form of the binomial distribution: $Y \sim Binomial(n+1, p)$, the only issue, is y starts at 1. Hence if we force the sum to start at $y = 0$, we'll need to subtract that value from the overall expression:

$$= \left[\sum_{y=0}^{n+1} \binom{n+1}{y} p^y (1-p)^{n+1-y}\right] - (1-p)^{n+1}$$

Now the first part of the expression is a valid distribution function, and since it is summed across the whole range of Y, it equals 1. (This is a very useful "trick" that will serve you well – make sure you are comfortable with your distributions.)

$$= 1 - (1-p)^{n+1}$$

Hence we have everything we need, so let's sub it back into the equation up above:

$$\Rightarrow E\left[\frac{1}{X+1}\right] = \frac{1}{p(n+1)} \sum_{x=0}^{n} \binom{n+1}{x+1} p^{x+1} (1-p)^{n+1-(x+1)}$$

Hence for $X \sim Bin(n, p)$

$$E\left[\frac{1}{X+1}\right] = \frac{1-(1-p)^{n+1}}{p(n+1)}$$

■