

Mat 354

Homework 7

Kenny Roffo

Due October 19, 2015

All calculations were done in R.

1. See HW 6 for background. A pharmaceutical company identifies numerous drugs similar to an already approved drug. For each of these derivative formulations a study is conducted: The drug is tapped for further development if at least 46 of the 50 people enrolled in a clinical trial of the drug experience no side effects. The probability of this occurring in any particular study is 0.03437 (a consequence of 0.184 / 0.816 side effect / no side effect rates for individual patients).

```
> 1-pbinom(45,50,0.816) [1] 0.03436944
```

The probability each study leads to further development is 0.03437, and all studies are independent (which amounts to using different subjects for each). The company will develop drug after drug until the find one that is tapped for further development; at that point they stop. The random variable here is the number of studies that are conducted.

- a) What is the probability they run more than 20 studies? 40 studies? 60 studies?

This is the probability that the first 20 (or 40 or 60) studies are unsuccessful. Since all studies are independent, this is just p^n where p is the probability that one study will fail, and n is the number of studies (20, 40 or 60). Therefore, we have

$$\begin{aligned} P(n > 20) &= (1 - 0.03437)^{20} \\ &= 0.496844 \end{aligned}$$

$$\begin{aligned} P(n > 40) &= (1 - 0.03437)^{40} \\ &= 0.2468539 \end{aligned}$$

$$\begin{aligned} P(n > 60) &= (1 - 0.03437)^{60} \\ &= 0.1226479 \end{aligned}$$

- b) What is the expected number of studies? The standard deviation?

This is a Geometric Distribution, so we use the formulas for the geometric distribution. The expected value is given by:

$$\begin{aligned} E(Y) &= 1/p \\ &= 1/(0.03437) \\ &= 29.09561 \end{aligned}$$

And the standard deviation:

$$\begin{aligned} SD(Y) &= \sqrt{\frac{1-p}{p^2}} \\ &= \frac{\sqrt{1-p}}{p} \\ &= \frac{\sqrt{1-0.03437}}{0.03437} \\ &= 28.59124 \end{aligned}$$

- c) Given that the company runs more than 40 studies, what is the probability they run more than 60 (which amounts to an additional at least 20 studies)?

Since the first 40 studies are completed, and unsuccessful, they don't matter anymore. Since individual events are independent, they cannot possibly change the likelihood of any individual event happening after the fact. Thus $P(n > 60|n > 40)$ is simply $P(n > 20)$. Therefore from part a we have

$$P(n > 60|n > 40) = 0.496844$$

2. (This can and should be solved without using results about summing geometric series. Use the distribution function F .) Y is a geometric random variable with $P(Y \geq 11) = 0.1$, determine the following.

a) The value of the Success probability p .

We know for a geometric distribution

$$F(y) = 1 - (1 - p)^y$$

Here we are given $P(Y \geq 11)$, which can also be written as $1 - P(Y \leq 10) = 1 - F(10) = 0.1$. Thus we can solve for p using the formula for F :

$$\begin{aligned} 1 - F(Y) &= 0.1 \\ \implies 1 - (1 - (1 - p)^{10}) &= 0.1 \\ \implies (1 - p)^{10} &= 0.1 \\ \implies 1 - p &= \sqrt[10]{0.1} \\ \implies p &= 1 - 0.7943282 \\ \implies p &= 0.2056718 \end{aligned}$$

b) $P(5 < Y < 15)$

This probability can be rewritten as

$$P(6 \leq Y \leq 14) = P(Y \leq 14) - P(Y \leq 5) = F(14) - F(5)$$

. Applying the distribution function for geometric distributions, we see:

$$\begin{aligned} P(5 < Y < 15) &= F(14) - F(5) \\ &= (1 - (1 - 0.2056718)^{14}) - (1 - (1 - 0.2056718)^5) \\ &= 0.276417 \end{aligned}$$

3. Odds for Odds Part 1

(This requires general results about summing geometric series.)

Assume $0 < p < 1$ and that Y has geometric distribution with probability p .

- a) Determine the probability of an odd outcome, $P(Y \text{ is odd})$. There are two ways to go about this. One (demonstrated in class) is direct. To handle it directly you need to apply the general result for summing geometric series. The other method is clever and avoids most of the work.

I am using the less clever approach here:

$$\begin{aligned}
 P(Y \text{ is odd}) &= P(Y = 1) + P(Y = 3) + P(Y = 5) + \dots \\
 &= (1-p)^0 p + (1-p)^2 p + (1-p)^4 p + \dots \\
 &= \sum_{k=0}^{\infty} (1-p)^{2k} p \\
 &= \frac{(1-p)^0 p - 0}{1 - (1-p)^2} \\
 &= \frac{p}{2p - p^2} \\
 &= \frac{1}{2-p}
 \end{aligned}$$

- b) Show that $P(Y \text{ is odd}) > 1/2$.

Since $0 < p < 1$, it must also be the case that $2-p < 2$. Since both sides of the inequality are positive, taking the reciprocal of both sides is legal, and requires us to reverse the inequality:

$$\frac{1}{2-p} > \frac{1}{2}$$

Since $P(Y \text{ is odd}) = \frac{1}{2-p}$, this means that $P(Y \text{ is odd}) > \frac{1}{2}$.

- c) Show what happens to $P(Y \text{ is odd})$ as p approaches 0 (from above), and as p approaches 1 (from below). That is, determine the following limits.

$$\lim_{p \downarrow 0} P(Y \text{ is odd})$$

$$\lim_{p \uparrow 1} P(Y \text{ is odd})$$

Approaching 0, we see

$$\begin{aligned}\lim_{p \downarrow 0} P(Y \text{ is odd}) &= \lim_{p \downarrow 0} \frac{1}{2-p} \\ &= \frac{1}{2-0} \\ &= \frac{1}{2}\end{aligned}$$

so as p goes to 0, the probability that the result will be odd gets closer to $1/2$.
Now for when p approaches 1...

$$\begin{aligned}\lim_{p \uparrow 1} P(Y \text{ is odd}) &= \lim_{p \uparrow 1} \frac{1}{2-p} \\ &= \frac{1}{2-1} \\ &= \frac{1}{1} \\ &= 1\end{aligned}$$

As it turns out, higher probabilities mean the result is more likely to be odd. This is almost obvious as a probability of 1 would mean the result is 1, an odd number, every single trial!

For d-f, you are welcome to suppose $p = 1/2$ (treating general p is really no more difficult).

- d) Determine the probability Y is divisible by 3. You will need to sum the appropriate series. (Did you derive the clever method in part a? If so: Why doesn't it work here?)

$$\begin{aligned}P(Y \text{ is divisible by } 3) &= P(3) + P(6) + P(9) + \dots \\ &= (1-p)^2p + (1-p)^5p + (1-p)^8p + \dots \\ &= \sum_{k=1}^{\infty} (1-p)^{3k-1}p \\ &= \frac{(1-p)^2p - 0}{1 - (1-p)^3} \\ &= \frac{(1-p)^2p}{p^3 - 3p^2 + 3p} \\ &= \frac{(1-p)^2}{p^2 - 3p + 3}\end{aligned}$$

Plugging in $p = 1/2$ we see

$$\begin{aligned}
 P(Y \text{ is divisible by } 3) &= \frac{(1 - (\frac{1}{2}))^2}{(\frac{1}{2})^2 - 3(\frac{1}{2}) + 3} \\
 &= 0.1428571
 \end{aligned}$$

e) Determine the probability Y is even, given that it is divisible by 3.

This is equal to the probability that $Y/3$ is even since the multiples of 3 alternate between even and odd in the same manner as the positive integers. Thus the probability that Y is even given that Y is divisible by 3 is the same as the probability that Y is even given Y is divisible by 1 (so given nothing really). Thus from part a, we have

$$\begin{aligned}
 P(Y \text{ is even}) &= 1 - P(Y \text{ is odd}) \\
 &= 1 - \frac{1}{2 - p} \\
 &= 1 - \frac{1}{2 - \frac{1}{2}} && \text{plugging in } 1/2 \text{ for } p \\
 &= \frac{1}{3}
 \end{aligned}$$

f) Are the events Y is even and Y is divisible by 3 independent? Why (not)?
 No. If you know that Y is even, then the probability that Y is divisible by 3 drops from

$$P(Y = 3) + P(Y = 6) + P(Y = 9) + \dots$$

to

$$P(Y = 6) + P(Y = 12) + P(Y = 18) + \dots$$

(These are not equal since the first contains the second, plus more positive numbers).