

# Homework # 4

Elliott Pryor

9/18/2020

## Problem 1

### Statement

A person is deciding whether to go on hike A or hike B this weekend. Let  $t$  be the number of hours in the hike,  $Y_1$  be the number of snack stops needed for hike A, and  $Y_2$  be the number of snack stops for hike B. Let,  $E(Y_1) = 0.75t$ ,  $Var(Y_1) = 0.75t$ ,  $E(Y_2) = 0.50t$ , and  $Var(Y_2) = 0.50t$ . The cost of going on hike A is  $C_A(t) = 5t + 10Y_1^2$ , and the cost of going on hike B is  $C_B(t) = 3t + 10Y_2^2$ . Which hike minimizes the expected cost if the hike A takes 2 hours, and B takes 10 hrs

---

### Solution

We know that  $Var(Y) = E(Y^2) - E(Y)^2$  so  $E(Y^2) = Var(Y) + E(Y)^2$ .

So the expected cost  $E(C_A(t)) = 5t + 10(E(Y_1^2)) = 5t + 10(Var(Y) + E(Y)^2) = 5*2 + 10(0.75*2 + (0.75*2)^2) = 47.5$

$E(C_B(t)) = 3t + 10(E(Y_2^2)) = 3t + 10(Var(Y) + E(Y)^2) = 3*10 + 10(0.5*10 + (0.5*10)^2) = 330$

Hike A is expected to be significantly cheaper than hike B. So hike A minimizes the expected cost.

## Problem 2

### Statement

Suppose that  $Y$  is a binomial random variable based on  $n$  trials with success probability  $p$  and consider  $Y^* = n - Y$

1. Argue that for  $y^* = 0, 1, \dots, n$

$$P(Y^* = y^*) = P(n - Y = y^*) = P(Y = n - y^*)$$

2. Use the result from (1) to show that

$$P(Y^* = y^*) = \binom{n}{n - y^*} p^{n - y^*} q^{y^*} = \binom{n}{y^*} q^{y^*} p^{n - y^*}$$

3. The result in (2) implies that  $Y^*$  has a binomial distribution based on  $n$  trials and “success” probability  $p^* = p = 1 - p$ . Why is this result “obvious”?
- 

### Solution

1. We are looking for the probability that the event did not happen. We can arrive at the given formula with a simple substitution. It is given that  $Y^* = n - Y$ . So we substitute this in to get:  
 $Y^* = y^* \iff n - Y = y^*$

We can then re-arrange this (move  $n$  to other side and divide by  $-1$ ) to get  $n - Y = y^* \iff Y = n - y^*$

So the three statements are equivalent to each other, and their probabilities must therefore also be the same. As required we have:

$$P(Y^* = y^*) = P(n - Y = y^*) = P(Y = n - y^*)$$

This result makes intuitive sense.  $y^*$  is the number of failures. So  $n - y^*$  would be the number of successes of a given trial. So the probability that we have no failures is the probability that every one is a success  $P(Y^* = 0) = P(Y = n)$  for example.

2. We know that  $f_Y(y) = P(Y = y) = \binom{n}{y} p^y q^{n-y}$ . So using the third equation  $P(Y = n - y^*)$  we arrive at  $P(Y = n - y^*) = \binom{n}{n-y^*} p^{n-y^*} q^{n-(n-y^*)} = \binom{n}{n-y^*} p^{n-y^*} q^{y^*}$ . The next relation is easy to see if we expand  $\binom{n}{n-y^*} = \frac{n!}{(n-(n-y^*))!(n-y^*)!} = \frac{n!}{(y^*)!(n-y^*)!} = \binom{n}{y^*}$

So by using this expansion we arrive at:

$$P(Y^* = y^*) = \binom{n}{n-y^*} p^{n-y^*} q^{y^*} = \binom{n}{y^*} p^{n-y^*} q^{y^*}$$

3. This is “obvious” because we are considering a ‘success’ to be a failure. So we essentially are just redefining what a success means in our space. Since the new probability of a success (which was the probability of failure) is  $p' = 1 - p$ .

## Problem 3

### Statement

If  $X$  is a Geometric random variable, show  $P(X = n + k | x > n) = P(X = k)$ . Let  $A$  be the event of a success on trial  $n+k$ , and  $B$  be the event that there are at least  $n$  failures before the first success. Use the definition of conditional probability and what you know about infinite geometric series to show the equality. Provide an explanation for why this is called the memory less property of the geometric distribution. HINT: Recall, if  $A \subset B \implies P(A \cap B) = P(A)$

---

### Solution

We know that  $P(Q|R) = \frac{P(Q \cap R)}{P(R)}$  by the definition of conditional probability. Translating the given events into mathematical notation yields  $A = n + k$  so  $P(A) = P(n + k)$  and  $B = x > n$  so  $P(B) = P(x > n)$ . So we end up with  $P(X = n + k | x > n) = P(A|B) = \frac{P(A \cap B)}{P(B)}$ . If we assume that  $k \neq 0$  then  $n + k \in \{x \in X | x > n\}$ . So  $A \subset B$ . So  $P(X = n + k | x > n) = \frac{P(A)}{P(B)}$

So we need to calculate the probabilities of each event. If an event is in  $B$ , then the first  $n$  options were fails. So the probability that an event is in  $B$  is just the probability of getting  $n$  fails at the beginning as we do not care what happens after that point. So:

$$P(B) = q^n$$

Then the probability that an event is in  $A$  occurs only if the first  $n + k - 1$  events are fails and the  $n + k$ -th event is a success. So:

$$P(A) = q^{n+k-1} p$$

So our total probability is

$$P(X = n + k | x > n) = \frac{q^{n+k-1}p}{q^n} = \frac{q^n q^{k-1}p}{q^n} = q^{k-1}p = P(X = k)$$

The last step comes from the *pmf* of a Geometric random variable derived in class. This can easily be re-derived, since  $P(X = k)$  means that the first success occurs at  $k$  so the first  $k - 1$  successes are fails. So the probability of this is  $q * q * q... * q * p = q^{k-1}p$ .

This is memory-less because the first  $n$  trials do not count. It simply forgets about them. Since it is given that all of the first  $n$  trials are fails, the geometric distribution can forget about them and imagine that they don't exist. Since the distribution does not depend on  $n$  it is considered memoryless. So it does not in any way depend on the previous trials ( $n$  flips of coin, etc). If it had memory, then the more times you flip it the more likely it will be to turn up heads, and  $p$  would not be constant. Clearly this isn't the case in the geometric distribution.

## Problem 4

### Statement

First, run the example R code provided, then answer the following question. The Bernoulli sample space is the set of all infinite sequences of zeros and ones corresponding to the tosses of a potentially biased coin with probability of heads equal to  $p$ . Let the random variable,  $W$ , be the waiting time until the first head occurs (i.e., the number of tosses until the first head occurs.) Last HW you wrote a function to generate realizations of  $W$  assuming  $p = 0.1$ , and an example of a function that does this is given in the example code for this HWw\_fun. Now,

1. Derive the pmf for  $W$ ,  $P(W = w)$  for  $w = 0, 1, 2, \dots$ . This is a re-parameterization of one of the distributions we have discussed in class. What distribution is this?
2. Create a graph comparing the theoretical distribution of  $W$  to your simulated version. If necessary, play around with increasing your number of realizations to get the plots to match. Approximately how many realizations do you need to accurately approximate the probability distribution?
3. Compute the average value for  $W$  from your simulation using R
4. Calculate the expected value under the true probability distribution (i.e., from the distribution sheet you are creating!). How does it compare to the average calculated via simulation?

### Solution

1. Verbally,  $P(W = w)$  is the probability the first success after  $w$  flips. So  $w = 0$  the first flip is a success,  $w = 1$  is the first is a fail and the second is a success,  $w = 3$  is the first two are fails and the third is a success. So for the case where  $w = k$  the probability of this happening is the probability that the first  $k$  are fails and the  $k + 1$  is a success. So  $P(W = k) = q^k p = (1 - p)^k p$ . This is just the geometric distribution.

$$f_W(w) = \begin{cases} (1 - p)^w p & \text{if } w \in \mathbb{N} \cup \{0\} \\ 0 & \text{otherwise} \end{cases}$$

2. Below is the function copied from the example code. The number of simulations required clearly depends very heavily on how close to the theoretical curve you want the simulations to be. I found that 5000 simulations typically produces results very close to the theoretical ones.

```
rg_fun <-function() {# generate a sequence of 100 independent Bernoulli trials
  samp <-rbinom(n = 100, size = 1, prob = 0.1)
  g <-which(samp==1)[1] - 1
  return(g)}
```

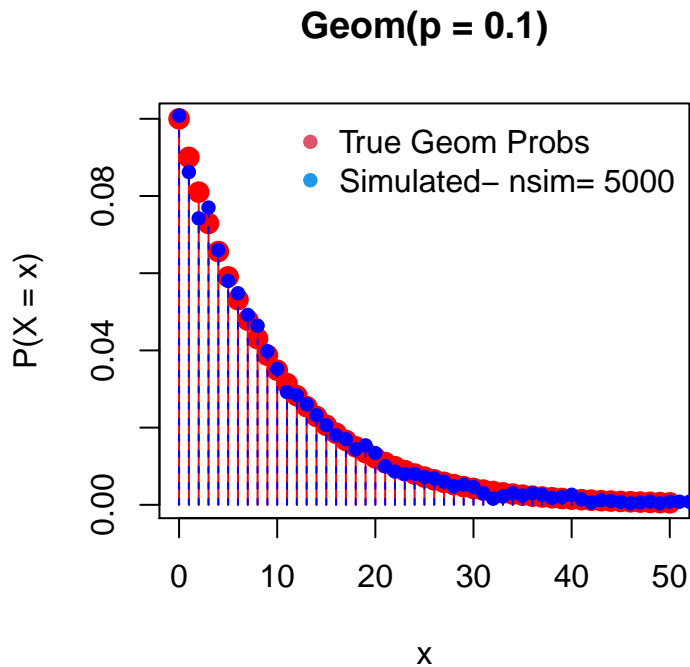
```

sim_num <- 5000
x_sim <- replicate(n = sim_num, expr = rg_fun())

# make relative frequencies (empirical probs)
x_rel_freq <- table(x_sim)/sim_num
# make a plot of the pmf of  $X \sim \text{geom}(0.1)$  for
#  $x = 0, 1, 2, \dots, 25$ 
x <- 0:50
plot(x, dgeom(x = x, prob = 0.1),
     type = "h",
     col = "red",
     main = "Geom(p = 0.1)",
     xlab = "x",
     ylab = "P(X = x)")
# add points at probability masses in  $S_X$ 
points(x, dgeom(x = x, prob = 0.1),
       col = "red",
       pch = 16,
       cex = 1.5)
# add relative frequencies from your simulation
# lines
points(x = as.numeric(names(x_rel_freq)),
       y = x_rel_freq,
       col = "blue",
       type = "h",
       lty = 2)

# points
points(x = as.numeric(names(x_rel_freq)),
       y = x_rel_freq,
       col = "blue",
       lty = 2,
       pch = 16)
# add legend
legend("topright",
      legend=c("True Geom Probs", paste0("Simulated- nsim= ", sim_num)),
      col = c(2,4), pch = 16, bty = "n")

```



3. We take the mean of all the simulated values.

```
mean(x_sim)
```

```
## [1] 9.0604
```

4. We know from in class that  $E(Y) = 1/p$  for  $Y \sim \text{geom}(p)$  so in our case the expected value is:  $1/p = 1/0.1 = 10$ . This is reasonably close to the value from the simulation, but is slightly higher. This would make sense as the theoretical value considers the infinite sum, so it necessarily considers  $x$  values beyond the simulated distribution. I.e. let  $x'$  be the highest observed  $x$  value in the simulated distribution, then for the simulated distribution  $P(y > x) = 0$  but we know this is not the case. So the theoretical expectation is influenced by these values bringing the expected value to a slightly larger number than the average of the simulated values.

## Problem 5 - Extra Credit

### Statement

Please share any questions or concerns you have with me about how class has been going now that we are through the first unit of material. You may share concerns about the workload, the use of class time, questions about the organization of content on the course website, or anything else that comes to mind that you feel might help you succeed in this course, and explain why. If you have no concerns please share something that is working well for you in this class so far, and explain why.

---

### Solution

I think the class is going well overall. Sometimes during the live zoom lectures it can feel rather boring and a bit of a waste. I just don't feel like I learn much material with some of the examples. Like with the rat experiment, it was OK, but did not learn much. I like when you work through problems in class as they are typically much harder and more like homework or potentially test problems. It helps me see different ways

you can approach the problem and problem-solving techniques. It also helps me understand a bit more of what you look for in the solutions.

The group time can be good, but typically only a few people participate. Many times group members haven't watched the lectures and don't have the mathematical understanding to do the problems. Being able to choose our own groups could be better as I could be with other friends who I enjoy working with and know their level of mastery, so we can feed more off of each other to learn. I also get that this is probably not possible, due to the issues around the constant groups that we had at the beginning. But as is can help me reinforce the concepts by explaining and helps me meet new people which is always good :)

I would prefer more days where we are primarily in groups working through a large number of problems. Like when we did the Week2 practice problems and this week the problems in the word document. This shows me what areas I am lacking and helps me see a wide variety of problems. When we only have 1 or 2 problems an hour long zoom meeting is overkill for me, so I end up checking out and not paying much attention. When we are working in a group for extended time we can end up getting a system together and being significantly more productive. I think this is more conducive for learning and feeding off each other's knowledge. Also I appreciate it when you go over some of them with the class, but I think I could look back at the ones that troubled me specifically if you just posted solutions. I like reconvening at the end and taking a poll for the hardest one, but spending another lecture or two covering the problems is a bit much.

It just feels like a lot of different things: zoom, videos, and HW every week. I just get bogged down a bit. But to be fair, the workload is very manageable for a 400 level course. I think you have very reasonable expectations of what we can do and considerate of time.