Homework 6

Autumn 2023

KEY

2023-11-14

Instructions

- This homework is due in Gradescope on Wednesday Nov 15 by midnight PST. There is a 15 minute grace period and submissions made during this time will not be marked as late. Any work submitted past this period is considered late.
- Please answer the following questions in the order in which they are posed.
- Don't forget to (i) make a local copy this document for your work and to (ii) knit the document frequently to make sure there are no compilation errors.
- When you are done, download the PDF file as instructed in section and submit it in Gradescope.

Exercises

1. (Aphids) A large number of insects are expected to be attracted to a variety of rose plant. A commercial insecticide is advertised as being 99% effective. Suppose 2,000 aphids infest a rose garden where the insecticide has been applied and let the random variable X denote the number of surviving aphids.

a. What probability distribution might provide a reasonable model for the random variable X? Be sure to:
   - state the values for the parameters of the distribution
   - state any assumption you need to make

Note that since the insecticide is advertised as being 99% effective, then the probability that an insect will not survive the insecticide is 0.99. Therefore, the probability that an insect will survive is $1 - 0.99 = 0.01$.

If we make the assumptions that

- Whether an aphid survives the insecticide is independent of whether another aphid survives the insecticide.
- The probability that an aphid survives the insecticide, $\pi = 0.01$, is the same for every of the $n = 2,000$ aphids.

then, because X is the number of surviving aphids, $X \sim Binom(n = 2000, \pi = 0.01)$ might be a reasonable model for X.

b. Using your model in part a, calculate $P(X < 10)$, the probability that fewer than 10 aphids survive. Be sure to do your calculation in a code chunk and then report the answer (rounded to three decimal places) in a complete sentence using inline code.

Since X takes non-negative integer values, $P(X < 10) = P(X \leq 9)$.

```
binom_prob_less100 = pbinom(q=9, size=2000, p=0.01)
```

The probability that fewer than 10 aphids will survive, using a binomial model, is 0.005.

c. What other probability distribution might be computationally more convenient and would provide a good approximation for the probability in part b? Be sure to state the values for the parameters of the distribution.

By chapter 8.2 slide 10, we know that $X \sim Pois(\lambda = n\pi)$ is another probability distribution that might be computationally more convenient and would provide a good approximation to the probability in part b. Plugging in $n = 2000$ and $p = 0.01$, we have $X \sim Pois(\lambda = 20)$.

d. Repeat the calculation in part b. using your model from part c. Be sure to do your calculation in a code chunk and then report the answer (rounded to three decimal places) in a complete sentence using inline code.

```
pois_prob_less100 = ppois(q = 9, lambda = 20)
```

The probability that fewer than 10 aphids will survive, using a poisson model, is 0.005.

2. (Oysters) An oyster contains a pearl with probability π. You need a pearl for a tiara you are making, and keep opening oysters until you find one with a pearl. (poor oysters….)

a. Let the random variable X denote the number of oysters you throw away before you find one with a pearl. Write the PMF of X.

Assume each oyster you open is independent of another oyster. Then, since the probability that an oyster contains a pearl is π, $X \sim Geom(\pi)$. So the PMF of X is

$$P(X = x) = (1 - \pi)^x \pi, \qquad x = 0, 1, 2, \ldots$$

b. We want the probability that you have to throw away 3 or more oysters to be no larger than 0.05. To be specific, we want to find π so that $P(X \geq 3) \leq 0.05$. Find the minimum value of π that ensures this probability. (In this part, you are solving this manually. Show your steps, do any calculations in a code chunk and report the value of π with inline code)

Using the formula from Ch 8.1 slide 8, $P(X \geq 3) = (1 - \pi)^3$. Therefore, to find the minimum value of π, we do the following calculations

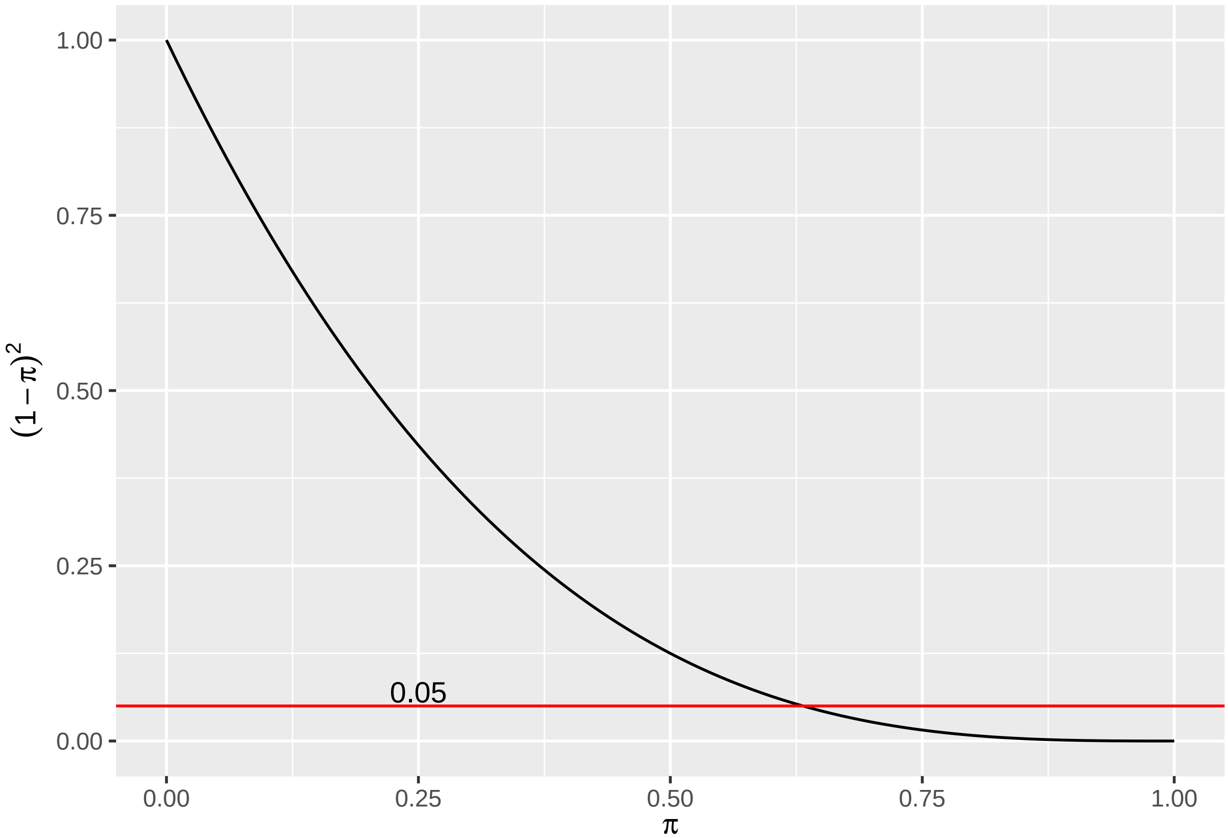
$$
\begin{aligned}
(1-\pi)^3 &\le 0.05 \\
1-\pi &\le (0.05)^{\frac{1}{3}} \\
-\pi &\le (0.05)^{\frac{1}{3}} - 1 \\
\pi &\ge 1-(0.05)^{\frac{1}{3}}
\end{aligned}
$$

```
min_pi = 1 - 0.05^(1/3)
```

Therefore, the minimum value of π that ensures $P(X \geq 3) \leq 0.05$ is 0.6316.

c. Repeat part b. using the `uniroot` function to obtain the value of π by numerical methods. Be sure to show your code in a code chunk and print the result. (See problem 3 from `Problem6.Rmd`)

```
ggplot() +
  geom_function( fun = function(x){(1-x)^3}, xlim = c(0,1)) +
  geom_hline(yintercept=0.05, color="red") +
  annotate(geom="text", x=0.25, y=0.07, label="0.05") +
  labs(x=expression(pi), y = expression((1-pi)^2))
```



From the graph above, we can see that the function $(1-\pi)^3$ is a decreasing function of π. Therefore, in order to find the smallest value of π that ensures

$$(1-\pi)^3 \leq 0.05$$

we need to find the π that satisfies the equation exactly. In other words, we need to find the root of the equation

$$(1-\pi)^3 - 0.05 = 0$$

```
#Looks like we need a value between 0.5 and 0.75
#Get piMin using uniroot function

piMin <- uniroot(f = function(x){(1-x)^3 - 0.05}, lower = 0, upper = 1)$root

paste("The minimum value for the probability pi is", round(piMin,4))
```

```
## [1] "The minimum value for the probability pi is 0.6316"
```

3. (Burnout) In a large factory building, where the fluorescent lights are kept on day and night, the lights burn out according to a Poisson process at a rate of $\lambda = 6$ per day. (Assume that lights are replaced as soon as they burn out for simplicity. Also a day refers to a 24 hour period.)

   For each part below, be sure to state the random variable and its distribution (if not already stated earlier), do calculations (if any) in a code chunk and report numerical answers in a complete sentence using inline code. (Rounding decimals to four digits is always a good idea)

a. Find the probability that there are more than two burnouts between noon and 1 PM tomorrow.

First, we note that lights burn out at a rate of $\lambda = 6$ per day implies that lights burn out at a rate of $\lambda_{\text{hour}} = 6/24 = 1/4$ per hour. Thus, let X be the random variable denoting the number of lights that will burn out between noon and 1PM tomorrow. We have that $X \sim \text{Poisson}(\lambda = 1/4)$ and we solve for $P(X > 2)$:

```
ans.3a <- ppois(2, lambda = 1/4, lower.tail=FALSE)
```

The probability that there are more than two burnouts between noon and 1PM tomorrow is 0.0022.

b. Find the probability that the next two burnouts will be at least 3 hours apart.

To solve this problem, we assume we start counting right after the first light burns out. Thus, let X denote the number of burnouts that occur during a three-hour period. Then $X \sim \text{Poisson}(\lambda = 3 \times (1/4))$ and we want to calculate $P(X = 0)$:

```
ans.3b <- dpois(x = 0, lambda = 3/4)
```

The probability that the next two burnouts will be at least three hours apart is 0.4724.

c. On a certain day, you count the number of burnouts between 8 AM and 8 PM. Let the random variable X denote the number of burnouts that occur during this 12 hour time period. How many burnouts should you expect? With what standard deviation? (Please give your answer in a complete sentence, don't just write two numbers)

Since we are looking at a 12-hour period, $X \sim \text{Poisson}(\lambda = 6/2 = 3)$. This also implies that $\mathbb{E}[X] = \text{Var}(X) = \lambda = 3$. Thus, we get the standard deviation $\text{sd}(X) = \sqrt{\text{Var}(X)} = \sqrt{3}$.

d. On a certain day, you count the number of burnouts between 8 AM and 8 PM. Let the random variable X denote the number of burnouts that occur during this 12 hour time period. What does Chebychev's inequality say about $P(1 < X < 5)$? Be sure to show your work.

We can use the expected value and standard deviation for X that we calculated in (c). Thus, for Chebyshev's inequality we have $\mu = 3$ and $\sigma = \sqrt{3}$. Now we calculate:

$$\begin{aligned}
P(1 < X < 5) &= P(-2 < X - 3 < 2) \\
&= P(|X - 3| < 2) \\
&= 1 - P(|X - 3| \geq 2) \\
&= 1 - P\Big(|X - 3| \geq \underbrace{\frac{2}{\sqrt{3}}}_{k} \underbrace{\sqrt{3}}_{\sigma}\Big) \qquad \geq 1 - \frac{1}{(2/\sqrt{3})^2} \\
&= 1 - \frac{3}{4} \\
&= \frac{1}{4}.
\end{aligned}$$

Thus, Chebyshev's inequality tell us that $P(1 < X < 5) \geq 1/4$.

e. Calculate the probability from part d. using the exact distribution of X.

```
ans.3e <- sum(dpois(2:4, lambda=3))
```

The exact probability from part (d) using the distribution of X is $P(1 < X < 5) = 0.6161$.

4. (Flooding river) A river floods every year. Suppose that the low-water mark is set at 1 and the high-water mark X has CDF

$$\begin{aligned}
F(x) &= P(X \leq x) \\
&= \begin{cases} 0 & x < 1 \\ 1 - 1/x^2 & 1 \leq x < \infty. \end{cases}
\end{aligned}$$

a. Find a PDF, $f(x)$ for X.

Hint: see example 9.3 from Chapter 9.

Assume that $x \geq 1$, then we calculate the following for the PDF of X:

$$\begin{aligned}
f(x) &= \frac{d}{dx} F(x) \\
&= \frac{d}{dx}\left(1 - \frac{1}{x^2}\right) \\
&= \frac{2}{x^3}
\end{aligned}$$

Thus, we obtain the PDF:

$$f(x) = \begin{cases} 0 & x < 1 \\ 2/x^3 & 1 \leq x < \infty \end{cases}$$

b. If the low-water mark is set to 0 and we use a unit of measurement that is $\frac{1}{10}$ of that given previously, the high-water mark becomes $Y = 10\,(X - 1)$. Find $P(Y \geq 1)$.

Hint: Y is just a transformation of X. Rewrite the event $Y \geq 1$ in terms of X and find the probability of this event.

We have the following calculation for $P(Y \geq 1)$:

$$\begin{aligned} P(Y \geq 1) &= P(10(X-1) \geq 1) \\ &= P(X - 1 \geq 1/10) \\ &= P(X \geq 11/10) \\ &= 1 - P(X < 11/10) \\ &= 1 - P(X \leq 11/10) && \text{(continuous distribution)} \\ &= 1 - \left(1 - \frac{1}{(11/10)^2}\right) \\ &= 0.8264 \end{aligned}$$