Homework 4

Autumn 2023

WRITE YOUR NAME HERE

2023-10-26

Instructions

- This homework is due in Gradescope on Wednesday Nov 1 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. (Therapy) In the past, a person afflicted with a certain neurological disease had a 30% chance of complete recovery. A radically different therapy has been tested on 10 patients, 7 of whom recovered. Let the random variable X denote the number (in a sample of 10) who recover using the new therapy.

a. What is the distribution of X assuming the new therapy is no better than the old one? State the name of the distribution and also the values of its parameters.

b. Make a probability histogram of the distribution in part a. Add a vertical line at $x = 7$ to the histogram using the `geom_vline` layer. (*Hint* Type `? geom_vline` in the Console for help.)

c. How *extreme* is a value of 7 under the presumed distribution in part a? The one-sided P-value is $P(X \geq 7)$. The smaller this probability, the more unusual the value is to have been observed. Calculate this number and report your answer (rounded to 4 decimals) in a complete sentence using inline code.

d. Do you believe the presumption that the new therapy is no better than the old one? Why or why not?

2. (Pooled Testing) Suppose that fifty people are to be given a blood test to see who has a certain disease. The obvious laboratory procedure is to examine each person's blood individually, meaning that fifty tests would eventually be run. An alternative strategy is to divide each person's blood sample into two parts (say), A and B. All of the A's would then be mixed together and treated as one sample. If that "pooled" sample proved to be negative for the disease, all fifty individuals must necessarily be free

of the infection, and no further testing would need to be done. If the pooled sample gave a positive reading, of course, all fifty B samples would have to be analyzed separately.

Let the random variable X denote the number of tests which will need to be performed. Also let π denote the probability that a randomly selected person is infected with the disease.

a. Write the PMF of X in a tabular format. You may assume independence of outcomes among people. (*Hint: X only has 2 values: 1, 51*).

Table 1: PMF of X: number of tests

x	1	51
$f(x)$		

b. Give an expression for $E[X]$. Show your steps beginning with the definition of an expected value.

c. Make a plot of $E[X]$ versus π. Does the graph make sense intuitively? Explain briefly. (You can use the `geom_function` layer as we did to graph the variance of a Bernoulli random variable in Example 7.7)

3. (Chebychev) Suppose $X \sim Binom(n, \frac{1}{2})$.

   a. What is the mean μ and standard deviation σ of X? Just cite results and use them. (See Chapters 7.1 and 7.2 for these results)

   b. Use Chebyshev's inequality to find the smallest n in order for

   $$P\left(0.9 \times \frac{n}{2} < X < 1.1 \times \frac{n}{2}\right)$$

   to be at least 90%.

   Hint: Show that the event $0.9 \times \frac{n}{2} < X < 1.1 \times \frac{n}{2}$ can be rewritten as $|X - \frac{n}{2}| < 0.1 \times \sqrt{n} \times \sigma$. Then apply Chebyshev's inequality.)

4. (Golfers) Two golfers are playing sudden death to decide a tournament. The first one wins a hole with probability p, the second one wins with probability q, and holes are tied with probability r. Holes are independent, and the game stops the first time someone wins a hole. What is the probability that the first player wins? Show your work step by step and cite any rules/results you use.

   Hint: Let X be the number of games it takes for player A to win. Then X can be 1 (win), 2 (tie, win), 3(tie, tie, win), and so on. The probability that player A wins is $P(X = 1 \cup X = 2 \cup \dots)$