Homework 2

Autumn 2023

KEY

Instructions

- This homework is due in Gradescope on Wednesday Oct 18 by midnight PST.
- Please answer the following questions in the order in which they are posed.
- Don't forget to 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. Suppose 12 coins are tossed and the outcome (head or tail) is recorded for each.

a. The sample space S corresponding to this "experiment" consists of all possible sequences of heads and tails that result from tossing 12 coins. How many elements are in S? Calculate this number in a code chunk and also explain your answer very briefly.

   Hint: Refer to Braille alphabet example 3.2. This problem is similar but with a 12 dot matrix.

   For each toss, there are 2 possible elements. So by the multiplication principle there are 2^{12} possible elements in S.

```
num_S <- 2^{12}
num_S
```

```
## [1] 4096
```

b. Let E denote the event that 7 of the 12 coins land on heads. How many elements are in E? Explain your answer very briefly.

   Hint: Referring to the Braille alphabet example again, suppose we now want to count all the letters we could form with 3 raised dots. All you need to do is decide which three of the six dots will be raised. How many ways can you make this decision?

   We are now trying to decide which of the 7 tosses will be heads. This corresponds to choosing an undordered subset - or combination - of 7 numbers from the numbers 1, 2, ..., 12. Therefore the event E has $\binom{12}{7}$.

```
num_E <- choose(12,7)
num_E
```

```
## [1] 792
```

c. Calculate $P(E)$ assuming all the elements in S are equally likely. Report your final answer in a sentence using inline code.

By the equally likely rule,

$$P(E) = |E|/|S|.$$

This is equal to 0.1933594.

2. To estimate the number N of goldfish in a pond, $r = 25$ fish were caught, tagged and released. Later, a second sample of $n = 20$ fish were caught and 5 fish in this sample were noted to be tagged.

a. How many possible samples of size $n = 20$ can be formed from the N fish in the pond? (Leave your answer in terms of a binomial coefficient - you cannot calculate it because you don't know N)

There are $\binom{N}{20}$ possible samples.

b. The event E contains all the samples which have 5 tagged and 15 untagged fish. How many elements are in the event E? (Leave your answer in terms of N)

We need to make two decisions in series: which 5 tagged fish and which 15 untagged fish will be in our sample. There are 25 tagged fish and $N - 25$ untagged fish in the pond.

By the multiplication principle, then there are

$$\binom{N-25}{15} \times \binom{25}{5}$$

samples which have 5 tagged and 15 untagged fish.

c. Assuming each possible sample is equally likely, give an expression for $P(E)$. (Leave your answer in terms of N)

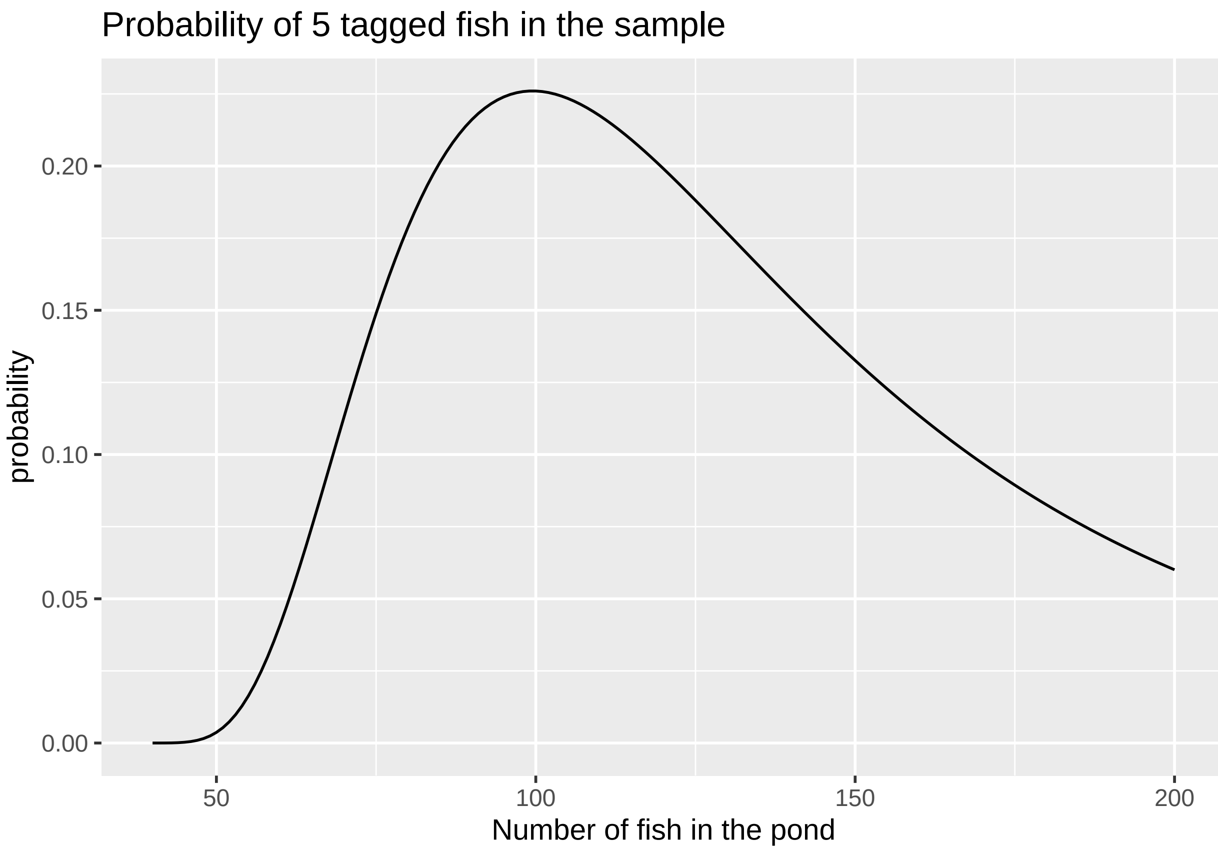
By the equally likely rule:

$$P(E) = \frac{\binom{N-25}{15} \times \binom{25}{5}}{\binom{N}{20}}.$$

d. In this part, we will examine visually how $P(E)$ varies as a function of N. Fill in the blanks in the R code provided and run it in R to create the plot. You should remove the `eval = FALSE` chunk option before knitting. (Note: you do not need to know **tidyverse** or **ggplot** functions to answer this question)

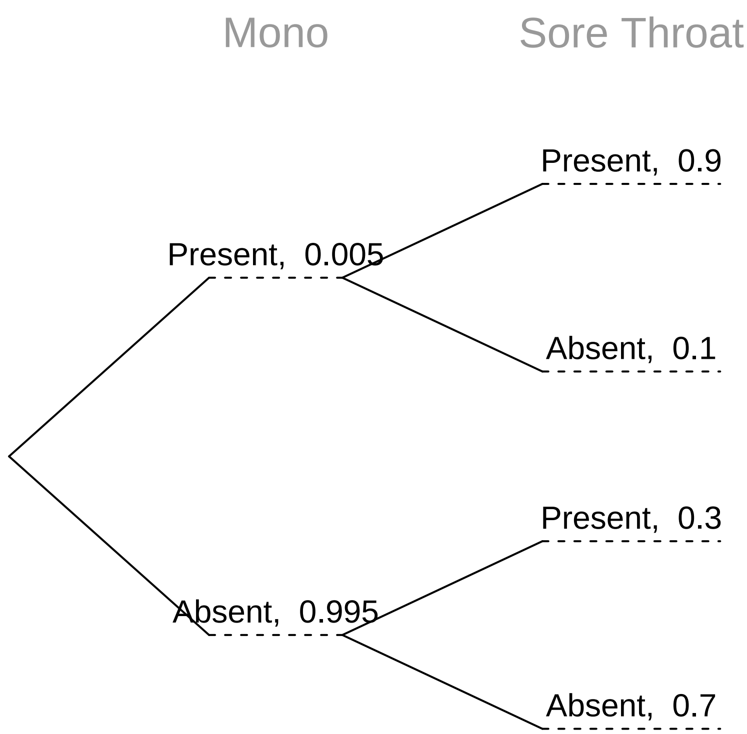
```
fishes <- tibble( #data frame
     N = 40:200,     #possible values for N: 40,41, ...,200
    prob = choose(25,5)*choose(N-25,15)/choose(N,20) )  #write expression for P(E) in terms of N

ggplot(data = fishes,
       mapping = aes(x = N,     #x and y variables
                     y = prob) ) +
      geom_line() +           #type of plot
      labs( title = "Probability of 5 tagged fish in the sample" ,              #labels for plot
                  x = "Number of fish in the pond",
                  y = "probability")
```



3. Among all students seeking treatment at Hall Health, 0.5% are eventually diagnosed as having mononucleosis (event A). Of those who do have mono, 90% complain of a sore throat (event B). But 30% of those not having mono also have sore throats.

a. Make a tree diagram of the probabilities relating presenting with a sore throat and a diagnosis of mononucleosis. (Don't forget to include the **openintro** package in the setup chunk)



b. If a student comes to Hall Health and says that she has a sore throat, what is the probability that she has mono? Be sure to show your steps carefully.

As stated in the problem, A is the event that a randomly selected student has mono and B that they have a sore throat. In this part, we want $P(A|B)$.

By definition of conditional probability:

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)}, \\ &= \frac{P(B|A) \times P(A)}{P(B)} = \frac{0.9 \times 0.005}{P(B)}, \end{aligned}$$

where we have used the definition of the conditional probability

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

to replace the expression in the numerator. Also from the information given in the problem:

$$P(B|A) = 0.9, P(A) = 0.005.$$

In order to find $P(B)$ for the denominator, we write B as the union of two disjoint sets:

$$B = (B \cap A) \cup (B \cap A^c).$$

By axiom 3 then:

$$\begin{aligned} P(B) &= P(B \cap A) + P(B \cap A^c), \\ &= P(B|A) \times P(A) + P(B|A^c) \times P(A^c), \\ &= 0.9 \times 0.005 + 0.3 \times (1 - 0.005). \end{aligned}$$

Putting this all together:

$$P(A|B) = \frac{0.9 \times 0.005}{0.9 \times 0.005 + 0.3 \times (1 - 0.005)}.$$

We can calculate $P(A|B)$ in a code chunk as shown below.

```
prob_b_and_a <- 0.9*0.005
prob_b_and_ac <- 0.3*(1-0.005)
prob_b <- prob_b_and_a + prob_b_and_ac
prob_a_given_b <- prob_b_and_a/prob_b
prob_a_given_b
```

```
## [1] 0.01485149
```

4. If A and B are independent events, then the following pairs are also independent.

a. A and B^c

b. A^c and B

c. A^c and B^c.

Prove result c. only. Show your work carefully with justification.

Two events E and F are independent if

$$P(E \cap F) = P(E) \times P(F).$$

To prove that A^c and B^c are independent, we begin with DeMorgan's Law which states that

$$\begin{aligned}
P(A^c \cap B^c) &= P((A \cup B)^c), \\
&= 1 - P(A \cup B), \\
&= 1 - [P(A) + P(B) - P(A \cap B)]\,, \text{ Inclusion-exclusion} \\
&= 1 - P(A) - P(B) + P(A) \times P(B), \quad \text{independence} \\
&= P(A^c) - P(B)\,(1 - P(A)) \\
&= P(A^c) - P(B) \times P(A^c) \\
&= P(A^c)\,(1 - P(B)) \\
&= P(A^c) \times P(B^c).
\end{aligned}$$

Hence A^c and B^c are independent events.