Homework 03

Christina Lee

9/24/2021

1.

a. What's the chance of getting a sequential pair on two rolls of a die (eg, a 3 then a 4 counts, but a 4 then a 3 does not). (Hint: you can calculate this manually if you like, by counting up the sample space and finding the fraction of that sample space that consists of ordered pairs.)

To answer this, I listed out all sample space = 36 possible outcomes, and the total number of sequential pairs are (1,2)(2,3)(3,4)(4,5)(5,6) = 5 pairs. Therefore, the answer to this question is $P(seqpair) = 5/36$

(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)
(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)
(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)
(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)
(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)
(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)

b. Given a dartboard with a inner circle that is 2/3 of the total area, and a bulls-eye that is 5% of the total area (and entirely within the inner circle): if you are throwing a random dart (that is guaranteed to hit somewhere on the board, but everywhere inside is equally likely), what is the chance of hitting the bulls-eye conditional on knowing your dart is somewhere inside the inner circle?

To answer this, we know that the inner circle P(A) = 2/3 and P(B) = 0.05 and P(A|B) = 1. All we have to do is to put what is known into the equation which gives us the answer of 0.075% chance of hitting the bulls-eye.

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)} = \frac{1\,*\,0.05}{\frac{2}{3}} = 0.075$$

c. You take a test for a scary disease, and get a positive result. The disease is quite rare – 1 in 1000 in the general population. The test has a sensitivity of 95%, and a false positive rate of only 5%. What is the chance you have the disease?

To answer this, put what we know into Bayes' equation:
P(disease) = 0.001
P(+|disease) = 0.95
P(+|No disease) = 0.05
P(No disease) = 1 - 0.001 = 0.999

Therefore, given that you tested positive for this disease, there is only approximately 1.9% chance of you actually having the disease.

$$P(disease|+) = \frac{P(disease)\,P(+|disease)}{P(disease)\,P(+|disease)\;+\;P(No\,disease)\,P(+|No\,disease)}$$

$$P(disease|+) = \frac{(0.001)(0.95)}{(0.001)(0.95) + (0.999)(0.05)} = \frac{0.00095}{0.00095 + 0.04995} = \frac{0.00095}{0.0509} = 0.019$$

d. What is the chance you have the disease if everything remains the same, but the disease is even rarer, 1 in 10,000?

To answer this, put what we know into Bayes' equation:
P(disease) = 0.0001
P(+|disease) = 0.95
P(+|No disease) = 0.05
P(No disease) = 1 - 0.0001 = 0.9999

Therefore, given that you tested positive for this disease, there is only approximately 0.19% chance of you actually having the disease.

$$P(disease|+) = \frac{P(disease)\ P(+|disease)}{P(disease)\ P(+|disease)\ +\ P(No\, disease)\ P(+|No\, disease)}$$

$$P(disease|+) = \frac{(0.0001)(0.95)}{(0.0001)(0.95) + (0.9999)(0.05)} = \frac{0.000095}{0.000095 + 0.049995} = \frac{0.000095}{0.05009} = 0.0019$$

e. What does this tell you about the dangers of tests for rare diseases?

This tells us that the rarer the disease is –there is a less chance that you actually have the disease even though you tested positive for it. Such that, the accuracy rate for tests decreases as the disease gets rarer.

2.

a. You have a 20-side die. Using sample, roll it 10,000 times and count the number of rolls that are 10 or less.

To answer this, I created a die with 1:20 sides and named it die1. I used sample() function to sample die1, rolled it 10,000 times, and added the replace argument. In order to count the number of rolls that are 10 or less, I placed the sample into a variable called "rollthedie" and used the sum() to count rolls with numbers that are <= 10 by adding the argument (rollthedie <= 10), which it will generate a series of TRUE = number is <= 10 and FALSE = number is > 10. The sum() will then add up number of TRUEs that appeared which essentially tells me the number of rolls that are <= 10.

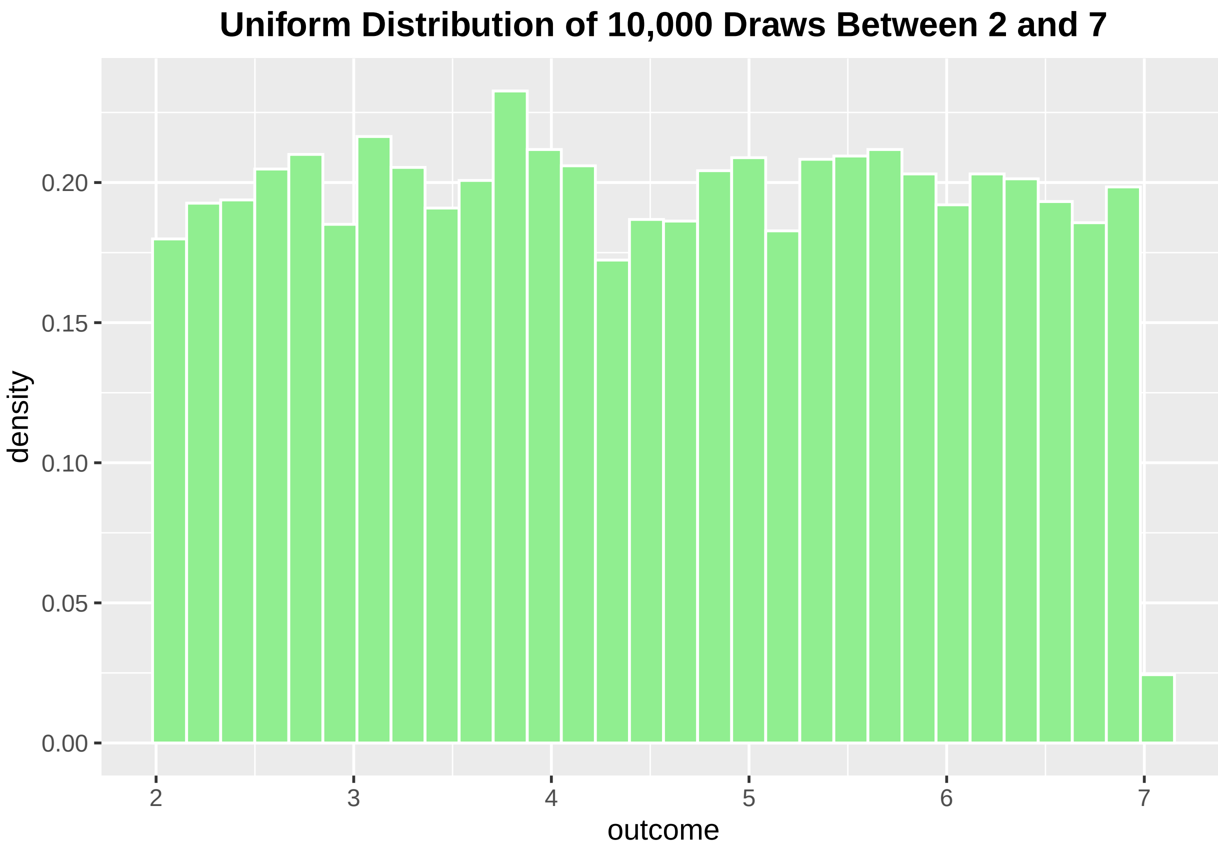
```
die1 = c(1:20)

rollthedie <- sample(die1,10000,replace=TRUE)
sum(rollthedie<=10)
```

```
## [1] 5032
```

b. Generate a histogram using ggplot of 10,000 draws from a uniform distribution between 2 and 7.

To answer this, I used the runif() function since we want a random uniform distribution and added in the first argument (generate 10000 draws), and the second and third argument (lower and upper bound, 2 and 7). I then named it "randunifs" and plotted the histogram via ggplot().

```
library(ggplot2)
randunifs <- runif(10000, 2,7)
ggplot(data=data.frame(randunifs),aes(x=randunifs)) +
  geom_histogram(aes(y=..density..),col="white",fill="lightgreen", bins=30) +
  ylab("density") + xlab("outcome") +
  ggtitle("Uniform Distribution of 10,000 Draws Between 2 and 7") +
  theme(plot.title = element_text(hjust = 0.5, face="bold"))
```



c. Try to write down the equation for this probability density function.

$$f(x) = \int_{0\ for\ x<2\ or\ x>7}^{\frac{1}{b-a}\ for\ 2\leq x\leq 7}$$

d. What is the probability that a draw from this distribution will be between 1.5 and 3.2?

To answer this, I used the punif() function since I want to know the probability from this uniform distribution. Next, the probability between 1.5 and 3.2 is basically P(X<=3.2) - P(X>=1.5) so, all I have to do now is set up the arguments for each punif() functions (the min and max is the same as the one above since we are using the same uniform distribution as 2b. The answer is 0.24% of chance that a draw from this distribution will be between 1.5 and 3.2.

```
punif(3.2, min=2,max=7) - punif(1.5, min=2, max=7)
```

```
## [1] 0.24
```

3.

a. Using R's cdf for the binomial, what is the probability of getting 500 or fewer "20"s when rolling your 20-sided die 10,000 times. Looking back at 2a, how many of your rolls were actually 20s?

To answer this, since we are, again, asking for probability for the binomial, I used pbinom() function to calculate. I filled in the parameter for this function according to the question where 500 is q, 10000 is size, and 1/20 is the prob, which gives the answer of 0.51% chance of getting 500 or fewer "20"s when rolling my 20-sided die 10000 times.

```
pbinom(500,10000,1/20)
```

```
## [1] 0.511895
```

Using the same method as 2a, I changed the argument from <= 10 to == 20 since I am asking for the rolls that were = 20.

```
die1 = c(1:20)

rollthedie <- sample(die1,10000,replace=TRUE)

sum(rollthedie == 20)
```

```
## [1] 506
```

b. Using rbinom, roll a 100-sided die 100 times and report the total number of 7s you get.

To answer this, I used rbinom() to generate required number of random values of given probability from a given sample. To do that, I added in some arguments in the function: I only want 1 observation, 100 trial per observation, and the probability is 1/100 which is 0.01. I renamed the function and called it "hundie" for me to output the outcome and the output that I will get is the total number of 7s that I get in the 100 rolls.

```
hundie <- rbinom(1,100,0.01)
hundie
```

```
## [1] 0
```

This is another way to answer this question. Again, I used rbinom() to generate the required number of random values of given probability from a given sample, however, I added the following arguments in the function: this time, I want 100 observations, 1 trial per observation, and the probability is still 1/100 which is 0.01. I basically took this question and thought of it as a binary experiment. Although the probability stays the same, but instead of a dice, I thought of the question as flipping a two-sided coin, which I generated 100 flips of one coin, and the outcome can only be Heads or Tails. Heads (0) = all numbers that doesn't equal to 7 and Tails (1) = 7. The probability of getting Heads is 0.99 and 0.01 for Tails. Therefore, ribnom() will generate 100 numbers that are either 0 = fail or 1 = pass. I named the rbinom() function to "hundie2" and passed it into the sum() to count the number of 1s which essentially gives me the total number of 7s that was generated in the 100 rolls.

```
hundie2 <- rbinom(100,1,0.01)
sum(hundie2 == 1)
```

```
## [1] 0
```

c. You are a klutz, and the average number of times you drop your pencil in a day is 1. Using the poisson functions in R, what's the chance of dropping your pencil two or more times in a day? (Hint: calculate the chance of dropping it one or fewer times, and then take 1 minus that.)

To answer this, I used ppois() function to calculate the probability of dropping pencil two or more times by first calculating the chance of dropping it one or fewer times (which is my first argument =1) and with a mean of 1 (my second argument). Then I took 1 minus the ppois() function and named it variable "pencildrop" to output the answer.

```
pencildrop <- 1-ppois(1,1)
pencildrop
```

```
## [1] 0.2642411
```

d. Because he is lazy, your teacher has assigned grades for an exam at random, and to help hide his deception he has given the fake grades a normal distribution with a mean of 70 and a standard deviation of 10. What is the chance your exam got a score of 85 or above? What is the chance you got a score between 50 and 60?

To answer this, I used pnorm() function to find the probability of getting a score of 85 or above, where 85 is x, 70 is mean, 10 is sd, and lower.tail =FALSE because I am interesting in knowing the probability of value x=85 or larger which are located on the upper tail of the distribution.

```r
pnorm(85,70,10, lower.tail=FALSE)
```

```
## [1] 0.0668072
```

To answer this, I used pnorm() function to find the probability of getting a score between 50 and 60 by taking the probability of 60 and subtract it by probability of 50.

```r
pnorm(60,70,10, lower.tail=TRUE) - pnorm(50,70,10, lower.tail=TRUE)
```

```
## [1] 0.1359051
```