DSC 40A - Homework 6

Due: Sunday, May 22, 2022 at 11:59pm PDT

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Homeworks are due to Gradescope by 11:59pm PDT on Sunday.

Homework will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should **always explain and justify** your conclusions, using sound reasoning. Your goal should be to convince the reader of your assertions. If a question does not require explanation, it will be explicitly stated.

Homeworks should be written up and turned in by each student individually. You may talk to other students in the class about the problems and discuss solution strategies, but you should not share any written communication and you should not check answers with classmates. You can tell someone how to do a homework problem, but you cannot show them how to do it.

This policy also means that you **should not post or answer homework-related questions on Piazza**, which is a written medium. This includes private posts to instructors. Instead, when you need help with a homework question, talk to a classmate or an instructor in their office hours.

For each problem you submit, you should **cite your sources** by including a list of names of other students with whom you discussed the problem. Instructors do not need to be cited. The point value of each problem or sub-problem is indicated by the number of avocados shown.

Problem 1. Odd One Out

Consider a game played with $n \geq 3$ people that goes as follows: each person flips a fair coin. If exactly one person's coin comes up different from all the others, then that person is the winner. If there is no winner, then play the game again and again, until a winner is finally determined.

a) Find the probability that a winner is determined the first time the game is played.

b) Fix an integer $k \geq 1$. Find the probability that a winner is determined on the kth time the game is played.

Problem 2. Coin Tricks

a) Suppose that after some practice flipping coins, your probability of landing a coin Heads up is 2/3. What is the probability you land the coin Heads up at least once in 3 flips?

b) Assume again that your probability of landing a coin Heads up is 2/3. If you flip 7 coins, what is the probability that exactly 4 of them land Heads up?

c) Now suppose that you enter a coin-flipping competition where you have to do certain maneuvers after throwing the coin and before it lands. If you can complete a full turn of your body before the coin lands, and the coin lands Heads up, you earn 1 point. If you can do a somersault, and the coin lands Heads up, you earn 2 points.

You've practiced extensively and know that your probability of landing a coin Heads up after a full turn is 4/7 and your probability of landing a coin Heads up after a somersault is 3/5.

In the competition, you get to attempt each maneuver (full turn, somersault) 3 times. What is the probability that you earn exactly 4 points in the competition?

Problem 3. Probability Theory

Let S be a sample space, and let A, E_1, E_2, E_3 be events in that sample space. Suppose that $E_1 \cap E_2$, $E_1 \cap E_3$, and $E_2 \cap E_3$ are all empty. Given the following probabilities, find $P(E_2|A)$:

$$\begin{array}{ll} P(E_1) = 1/6 & P(A|E_1) = 1/9 \\ P(E_2) = 1/3 & P(A|E_2) = 1/7 \\ P(E_3) = 1/2 & P(A|E_3) = 1/5 \end{array}$$

Problem 4. House of Cards (continued)

Let's look back at Problem 4 in our Homework 5, where we use a standard deck of card containing 52 cards. There are 13 cards in each of 4 suits (hearts ♡, spades ♠, diamonds ♢, and clubs ♣.) Within a suit, the 13 cards each have a different rank. In ascending order, these ranks are 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King, Ace.

You are playing a four-player card game using two regular decks of cards. Each player will be dealt 26 cards. Two identical cards, with the same rank and same suit, are called a pair. A pair can only be outranked (beaten) by a pair of higher ranked cards within the same suit.

a) Let's reexamine Problem 4(f) in Homework 5. Suppose both decks of cards are first shuffled together and then dealt out to the four players. What is the probability that you get a pair of Kings of Hearts?

Redo this problem using $P(n, k)$ notation and interpret the solution in terms of ordered choices, or permutations. Next, rewrite the expression using $C(n, k)$ notation and interpret the solution in terms of unordered choices, or combinations.

b) Suppose both decks of cards are first shuffled together and then dealt out to the four players. You are dealt two Kings of Hearts. What is the probability that some other player has a pair of Aces of Hearts?

c) Suppose both decks of cards are first shuffled together and then dealt out to the four players. You are dealt two Kings of Hearts and you do not have any Aces of Hearts. What is the probability that some other player has a pair of Aces of Hearts?

d) Suppose both decks of cards are first shuffled together and then dealt out to the four players. What is the probability that you get either a pair of Kings of Hearts or a pair of Aces of Hearts?

Problem 5. Why does more data help (continued)?

In the last homework, we observed that with more data, the distribution of the data's mean is more concentrated. In this homework, let's explore this some more using combinatorics.

Let's think this way: we can separate the interval of $[\theta - 2, \theta + 2]$ into 30 bins and label them as $1, 2, \ldots, 30$. When a data point falls into bin i, we'll say that the data point has value i. Since our data are generated uniformly at random, they have the same probability of falling into any of these 30 bins, or the same probability of having any value from 1 to 30. Now the question is: what's the chance of the average taking certain value?

For instance, suppose we have just two data points x_1 and x_2, which each take on any value in $1, 2, \ldots, 30$. We would like to know the distribution of their mean, $(x_1 + x_2)/2$. Note that the smallest value for the sum, $x_1 + x_2$, is 2, which happens only when x_1 and x_2 both have value 1. Similarly, the largest value for $x_1 + x_2$ is 60, corresponding to when x_1 and x_2 both have value 30. This means $x_1 + x_2$ can take on any value in $2, 3, \ldots, 60$, so their mean $(x_1 + x_2)/2$ can take on any value in $1, 2, \ldots, 30$.

However, the probabilities of the mean $(x_1+x_2)/2$ taking on any value in $1, 2, \ldots, 30$ are not the same. This happens because the probabilities of the sum $x_1 + x_2$ taking on any value in $2, 3, \ldots, 60$ are not the same. As we've seen above, there is only one arrangement (x_1 and x_2 both having value 1) which leads to a sum of 2. But there are many arrangements that lead to a sum of 30, for example.

For two data points x_1, x_2 that can each take on a value in $1, 2, \ldots, 30$, let's define a function $f_2(k)$ to represent the number of arrangements in which $x_1 + x_2$ takes value k. We've shown that for k outside of the range $[2, 60]$, we have $f_2(k) = 0$. We've also established that $f_2(2) = 1$ and $f_2(60) = 1$.

a) Find $f_2(30)$. In other words, how many arrangements are there if we want $x_1 + x_2$ to take value 30?

b) For an arbitrary integer $2 \leq k \leq 60$, find $f_2(k)$.

Hint: Your answer should be a piecewise function of k.

c) Use your answer to part (b) to draw by hand a histogram of the distribution of the mean $(x_1 + x_2)/2$, using 30 bins. Don't worry about the scale of the vertical axis, just make sure the bins are the correct height relative to one another.

Compare this histogram to the plot you generated in Problem 5(b) in Homework 5 for $n = 2$. Are the histograms exactly the same? Why or why not?

d) What if we look at 3 data points, instead of 2? In that case, we are looking at the distribution of $(x_1 + x_2 + x_3)/3$, where $x_1 + x_2 + x_3$ takes a value from $3, 4, \ldots, 89, 90$. Define a function $f_3(k)$ to represent the number of arrangements in which $x_1 + x_2 + x_3$ takes value k.

Express $f_3(k)$ in terms of $f_2(k)$. Your expression should contain a summation symbol $\sum$, and there is no need to expand and simplify it.

e) For an arbitrary n, define a function $f_n(k)$ to represent the number of arrangements in which $\displaystyle\sum_{i=1}^{n} x_i$ takes value k.

Express $f_{n+1}(k)$ in terms of $f_n(k)$. Your expression should contain a summation symbol $\sum$, and there is no need to expand and simplify it.

f) From the above question, we have the number of arrangements for which the total $\displaystyle\sum_{i=1}^{n} x_i$ takes value k. What is the probability $p_n(k)$ that $\displaystyle\sum_{i=1}^{n} x_i$ takes on value k?

g) Fill in the `???` part of the following code in Python to plot $p_n(k)$ for $n = 2, 3, 4, 5$ on the same figure. Compare this graph to the plots you generated in Problem 1 of Groupwork 6. After doing this problem, summarize what you know about why the mean estimator is more concentrated with more data.

For the code below, note that the range of the x axis is always $[\theta - 2, \theta + 2]$, but the bins are finer as number of data points n increases. Hence to put pictures for different n on the same scale, we use the function `rescaled_prob_func` to combine the n consecutive bins together. You don't need to worry about how this part of the code works.

```
import numpy as np
import matplotlib.pyplot as plt

def arrangement_2(k):
```

```
# returns the number of arrangements of 2 elements that
# sum to k (your answer to (b))
    if ???:
        return ???
    elif ???:
        return ???
    else:
        return 0

def arrangement_func(n):
# returns an array where the element at index k is the number of arrangements
# of n elements that sum to k (your answer to (e))
    arrangement = np.zeros((n*30+1, n+1))
    for i in range(31):
    # calculate the number of arrangements of 1 element that sums to i
        arrangement[i,1] = 1
    for i in range(61):
    # calculate the number of arrangements of 2 elements that sum to i
        arrangement[i,2] = arrangement_2(i)
    for j in range(2, n):
        for x_1 in range(1,31):
            for k in range(x_1, (j+1)*30+1):
                # calculate the number of arrangements of j+1 elements
                # that sum to k ( use your summand in (e))
                arrangement[k,j+1] = arrangement[k,j+1] + ???
    return arrangement[:,n]

def prob_func(n):
# returns an array where the element at index k is the probability
# of an arrangement of n elements having a sum of k
    arr = arrangement_func(n)
    prob = ???              # use your answer to (f)
    return(prob)

def rescaled_prob_func(n): # the rescaled probability (do not change)
    prob = prob_func(n)
    rescaled_prob = np.zeros(30)
    for i in range(30):
        for j in range(n):
            rescaled_prob[i] = rescaled_prob[i] + prob[i*n+j]
    return rescaled_prob

plt.figure()                # plots (do not change)
xaxis = np.linspace(-2.0,2.0,num=30)
for n in range(2,6):
    rescaled_prob = rescaled_prob_func(n)
    plt.plot(xaxis, rescaled_prob)
plt.show()
```