STAT2120 Written Assignment #1

Aiman Kassam

October 2023

Question 1

Rolling a die is a common example of an **equiprobable** sample space, meaning all outcomes are equally likely to occur. The sample space for an 8-sided die is:

$$S = \{1, 2, 3, 4, 5, 6, 7, 8\} \tag{1}$$

Consider an equiprobable sample space B, and a space $A \subseteq B$:

$$P(A) = \frac{\#\ \textit{of outcomes in } A}{\#\ \textit{of outcomes in } B} \tag{2}$$

Assume three 8-sided dice with equiprobable outcomes are rolled. One green die, one red die, and one blue die. In this scenario, how many different permutations are possible? Assume order matters and repeats are allowed. The order in which the numbers are rolled matters because the first, second, and third roll each correspond to different coloured dice. Repeats are allowed because there is no reason why two or more dice can not roll the same number (e.g. green die rolls 3, red die rolls 3, and blue die rolls 3 is a valid permutation). For each of the eight possible outcomes of the green die, there are eight possible outcomes for the red die, or $8 \times 8 = 64$ permutations. For each of those permutations of the green and red dice, there are 8 possible outcomes for the blue die, totalling $64 \times 8 = 512$ possible permutations (or simplified to $3^8 = 512$ possible permutations).

A) Probability of the green die rolling an even number:

Let:

$$G = \{\textit{all even-numbered outcomes}\} = \{2, 4, 6, 8\} \tag{3}$$

Each die has eight possible outcomes, but in this scenario, only the green die is of concern. To understand why the red and blue dice do not matter, consider a scenario in which the green die rolls the number 2. The number of possible permutations that can be made by the red and blue dice is $8 \times 8 = 64$. Each permutation created by this process would be an example of a permutation in which the green die rolled an even number. This process can be repeated for all even-numbered outcomes of the green die. Since there are four even-numbered outcomes, there are $4 \times 64 = 256$ permutations in which the green die rolls an even number. Recall that there are 512 total permutations and *P(A)* from Equation 2. Substituting our possible permutations for even-numbers on the green die in for A, and the total possible permutations in for B:

$$P(G) = \frac{\#\ \textit{of permutations where green is even}}{\#\ \textit{of total possible permutations}} = \frac{256}{512} = \frac{1}{2} \tag{4}$$

Alternatively, consider the equation assuming the green die alone was looked at:

$$P(G) = \frac{\#\ \textit{of outcomes in } G}{\#\ \textit{of outcomes in } S} = \frac{4}{8} = \frac{1}{2} \tag{5}$$

Since the green die is independent of the red and blue dice in this scenario, both equations yield the same probability. In conclusion, the probability of the green die rolling an even number is 50%.

B) Probability of the sum of outcomes on the red and green dice being equal to the blue die:

Recall Equation 2. Since rolling three dice is an equiprobable experiment, the probability of any event can be calculated using Equation 2. The only condition is that the number of valid outcomes for the event must be known. Since the blue die must be summed up, and not the red or green dice, the order matters. Consider the following table:

Blue Die	**Red Die**	**Green Die**
1	$\varnothing$	$\varnothing$
2	1	1
3	1	2
3	2	1
4	1	3
4	2	2
4	3	1
5	1	4
5	2	3
5	3	2
5	4	1
6	1	5
6	2	4
6	3	3
6	4	2
6	5	1
7	1	6
7	2	5
7	3	4
7	4	3
7	5	2
7	6	1
8	1	7
8	2	6
8	3	5
8	4	4
8	5	3
8	6	2
8	7	1

Table 1: *Permutations in which the sum of the red and green dice equals the blue die.*

It is not possible for two dice to sum to 1, since 0 is not on any side of a die. Counting up the permutations listed in the table, there are 28 possibilities. Let event $M = \{$*event of permutation where the sum of red and green dice equalling blue die*$\}$. Now that the number of valid outcomes are known, Equation 2 can be used:

$$P(M) = \frac{\#\ \textit{of permutations in } M}{\#\ \textit{of possible permutations}} = \frac{28}{512} = \frac{7}{128} \tag{6}$$

There is a $\frac{7}{128}$ chance of a permutation yielding the blue die as the sum of the red and green dice.

C) Probability of the sum of outcomes on any two dice equalling the third die:

This scenario resembles that in which the values of the red and green dice were to sum to the value of the blue die. The difference is that order no longer matters. Permutations are sets of numbers in which the order matters (e.g. 1234 and 4321 are NOT the same since the order is different). Combinations are sets of numbers in which the order does not matter (e.g. 1234 IS the same as 4321 since the same set of numbers are used). To start, consider an edited version of Table 1, but with all repeated combinations taken out:

Die #1	**Die #2**	**Die #3**
1	∅	∅
2	1	1
3	1	2
4	1	3
4	2	2
5	1	4
5	2	3
6	1	5
6	2	4
6	3	3
7	1	6
7	2	5
7	3	4
8	1	7
8	2	6
8	3	5
8	4	4

Table 2: Combinations in which two dice sum to equal the third die.

There are now 16 possible combinations in which the sum of two dice adds up to the third. Each combination yields a certain number of permutations. In the event that all three numbers are distinct (e.g. $1 + 2 = 3$), there are $3! = 6$ permutations. A factorial is used here because all three terms must be used. Every time one term is selected it is eliminated as an option, and leaves all possibilities except itself behind. In the event that two of the three numbers are repeats, the total number of permutations would be 3. Typically, three numbers can be arranged in six different ways, but if 2 of the three numbers are repeats, then half of the permutations formed will be identical to the other half. Of the 16 combinations, four have repeated numbers ($1 + 1 = 2$, $2 + 2 = 4$, etc.), and the remaining 12 have three unique numbers. Therefore, the total number of permutations yielded are:

$$(12 \times 6) + (4 \times 3) = 84 \text{ permutations} \tag{7}$$

Let $N = \{\textit{space of permutations where two dice add up to the third}\}$. Recalling Equation 2:

$$P(N) = \frac{\#\ \textit{of permutations in } N}{\#\ \textit{of possible permutations}} = \frac{84}{512} = \frac{21}{128} \tag{8}$$

Question 2

Take a hypothetical rare disease that affects 0.005% of the population. A medical test for this disease has an accuracy rate of 98%. If someone has the disease, there is a 98% chance they will test positive, and a 2% chance they will test negative. If someone does not have the disease, there is a 98% chance they will test negative, and a 2% chance they will test positive. Let E be the event that someone has the disease, and let F be the event that someone has a positive test result.

A) What is the probability of $P(F|\bar{E})$?

$P(F|\bar{E})$, or the probability of someone testing positive when they are known to not have the disease can also be described as the false-positive rate. Given that the accuracy of the test is 98%, 2% of people who do not have the disease will test positive. With these initial conditions, no calculation is required. The probability is:

$$P(F|\bar{E}) = 0.02 = 2\% \tag{9}$$

B) What is the probability that someone who tests positive for the disease actually has it?

The probability of someone who tests positive for the disease actually having it can be written as $P(E|F)$. Among 99.995% of the population who do not have the disease, 2% would test positive. Among the 0.005% of the population that have the disease, 98% would test positive. Calculating these percentages using the multiplication rule and adding them up yields the portion of the population that has positive test results:

$$P(F) = (0.02)(99.995\%) + (0.98)(0.005\%) = 2.0048\% \tag{10}$$

Within this 2.0048% contains 98% of the cases of people who have the disease, the remaining 2% being false-negatives. To calculate the portion of the 2.0048% who have the disease, divide the number of true positive cases by the number of all positive cases:

$$P(E|F) = \frac{(0.98)(0.005\%)}{2.0048\%} = 0.24\% \tag{11}$$

The calculated percentage above says that if a random person in the population were tested and had a positive result, their probability of having the disease updated from 0.005% to 0.24%. Positive test results gives very little certainty about if someone has the disease. This is because a 2% false-positive rate is deceivingly high, given that only 0.005% of the population actually has the disease. The number of total positive cases is overwhelmingly dominated by false-positive cases due to the relatively high false-positive rate coupled with the very small population of people with the disease.

The process above was a breakdown of what Baye's Theorem computes in a single step (shown with the variables used rather than an abstract case):

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)} \tag{12}$$

In the context of the $P(E|F)$, Baye's Theorem says:

$$P(E|F) = \frac{(\text{\% of positive results if someone has the disease})(\text{\% of people with disease})}{(\text{\% of total possible cases})} \tag{13}$$

Baye's Theorem computes updates a probability from its value before evidence is presented to its probability after the condition is presented.