HW2 Solutions, for MATH441, STAT461, STAT561, due September 9th

1. You flip a coin until you get tails. Describe the sample space. How many points are in the sample space?

Solution.

The sample space consists of sequences of H and T where T occurs only at the last position. There are infinitely points in the sample space.

$$S = \{T, HT, HHT, HHHT, \dots\}$$

2. Two five-sided dice are rolled. The dice can be distinguished. Write down the sample space.

$$\begin{aligned} S = \{&(1,1),(1,2),(1,3),(1,4),(1,5),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),\\ &(5,1),(5,2),(5,3),(5,4),(5,5)\} \end{aligned}$$

3. Two five sided dice are rolled. Let A be the event that the sum is an odd number. (a) List the outcomes in A. (b) Determine $P(A)$.

Solution.
(a)

$$A = \{(1,2),(1,4),(2,1),(2,3),(2,5),(3,2),(3,4),(4,1),(4,3),(4,5),(5,2),(5,4)\}$$

(b) Since there are 12 points in A, and all outcomes are considered equally likely (we assume), $P(A) = \frac{12}{25}$.

4. In an ordinary deck of 52 cards, there are four aces (ace of spades, ace of diamonds, ace of clubs, ace of hearts). Suppose you are given 2 cards at random (without replacement) from a shuffled deck, and after your two cards are given, your friend is given two cards at random from the remaining 50 cards. Let A be the event that you have two aces. Let B the event that your friend has two aces. Compute (a) $P(A)$, (b) $P(B)$, (c) $P(AB)$, (d) $P(A \cup B)$. You can assume that by symmetry, $P(B) = P(A)$.

Solution.
(a) $P(A) = \frac{\binom{4}{2}}{\binom{52}{2}} = 0.004524887$

(b) $P(B) = P(A) = 0.004524887$
(c) $P(AB) = \frac{\binom{4}{4}}{\binom{52}{4}}$ or $P(AB) = \frac{1}{\binom{52}{4}}$ or $P(AB) = \frac{\binom{4}{2}}{\binom{52}{2}}\frac{1}{\binom{50}{2}} = .000003693785$

(d) $P(A \cup B) = P(A) + P(B) - P(AB) = 2 \cdot \frac{\binom{4}{2}}{\binom{52}{2}} - \frac{1}{\binom{52}{4}} = 0.00904608$

In this case, the probability of the intersection is so small that $P(A \cup B)$ is pretty close to $P(A) + P(B)$. Here $P(A) + P(B) = 0.009049774$.

5. Two six sided dice are rolled repeatedly until the sum is either 2 or 3. Find the probability that a 2 is rolled first. Hint: let E_i be the event that a 2 is rolled on the ith round and that neither a 2 nor a 3 has occurred on any previous round. Let E be the event of interest. Then argue that the event $E = \cup_{i=1}^{\infty} E_n$ and compute $P(E) = P\left(\cup_{i=1}^{\infty} E_n\right)$.

Solution.

The probability of rolling a 2 in one particular roll is $\frac{1}{36}$. The probability of rolling a 3 is $\frac{2}{36}$ because there are two ways to roll a 3: (1,2) and (2,1). The probability of neither a 2 nor a 3 is $1 - \frac{1}{36} - \frac{2}{36} = \frac{33}{36}$.

$$\begin{aligned} P(E_1) &= \frac{1}{36} \\ P(E_2) &= \frac{33}{36} \cdot \frac{1}{36} \\ P(E_3) &= \frac{33}{36} \cdot \frac{33}{36} \cdot \frac{1}{36} \\ P(E_n) &= \left(\frac{33}{36}\right)^{n-1} \cdot \frac{1}{36} \\ &= \left(\frac{11}{12}\right)^{n-1} \cdot \frac{1}{36} \end{aligned}$$

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} E_i\right) &= \sum_{i=1}^{\infty} \left(\frac{11}{12}\right)^{i-1} \cdot \frac{1}{36} \\ &= \frac{1}{36} \sum_{n=0}^{\infty} \left(\frac{11}{12}\right)^{n} \\ &= \frac{1}{36} \cdot \frac{1}{1 - 11/12} \\ &= \frac{12}{36} \\ &= \frac{1}{3} \end{aligned}$$

Here we used the fact that the sum is a geometric series. Therefore $P(E) = 1/3$ is the probability of rolling a 2 first.

Another way to solve this problem is to use conditional probabilities. Let A be the event that 2 is rolled on the first round. Let B the event that 3 is rolled on the first round. Let C be the event that neither occurs on the first round. Then using total probability,

$$\begin{aligned} P(E) &= P(E|A)P(A) + P(E|B)P(B) + P(E|C)P(C) \\ &= 1 \cdot \frac{1}{36} + 0 \cdot \frac{2}{36} + P(E) \cdot \frac{33}{36} \\ &= \frac{1}{36} + P(E) \cdot \frac{11}{12} \\ &\Rightarrow P(E)\left(1 - \frac{11}{12}\right) = \frac{1}{36} \\ &\Rightarrow P(E) = \frac{12}{36} = \frac{1}{3} \end{aligned}$$

This seems easier than the infinite series approach. The main trick is realizing that $P(E|C) = P(E)$ and the reason for this is that if no one wins in the first round, we essentially just start the game over, and the probability of rolling a 2 first starting on the second round is the same as rolling a 2 first starting on the first round.

6. Assume that the probability that a randomly selected person has a birthday in a given month is 1/12. In other words, the probability that a birthday occurs in January is 1/12, the probability that it occurs in February is 1/12, etc. What is the probability that in a group of 5 people, each person has a birthday in a distinct month? Hint: This is a variation of the famous birthday problem which asks the probability that a group of n people have distinct birthdays considering 365 possible birthdays.

Solution. It doesn't matter what month the first person's birthday is in. The probability that the second person has a birthday in the same month as the first person is 1/12, there for the probability that the second person has a birthday in a different month from the first person is $1-1/12 = 11/12$. The probability that the third person has a birthday in a month distinct from the first two given that the first two have distinct months is 10/12 because there are 10 months remaining that don't have birthdays. The probability we are interested in is therefore

$$\frac{11}{12} \times \frac{10}{12} \times \frac{9}{12} \times \frac{8}{12} \approx 0.38$$

Thus, the probability is less than 1/2 that five people will have birthdays in distinct months under this model.

7. In a certain population, 4% of people are color blind. A subset is to be randomly chosen. How large must the subset be if the probability of its containing at least one color blind person is at least 95%? (Consider the population to be large enough that you can consider it to be infinite, so the selection can be considered with replacement. In other words, don't worry about whether a random sample might sample the same person twice.)

Solution. Let A_i be the event that the ith person in the sample is color blind. Then $P(A_i) = .04$ for each i. The event that there is at least one color blind person in the sample is

$$A_1 \cup A_2 \cup \cdots \cup A_n$$

We are interested in the value of n for which $P(A_1 \cup A_2 \cup \cdots \cup A_n) \geq .95$. Using inclusion-exclusion is possible but difficult for large unions. It would be easier if we had intersections. What do we use to convert unions to intersions? DeMorgan's rule. Use $P(A) = 1 - P(A^c)$ to write the probability as

$$\begin{aligned}
&P(A_1 \cup A_2 \cup \cdots \cup A_n) \geq .95 \\
&\Rightarrow 1 - P[(A_1 \cup A_2 \cup \cdots \cup A_n)^c] \geq .95 \\
&\Rightarrow 1 - P[A_1^c A_2^c \cdots A_n^c] \geq .95 \\
&\Rightarrow 1 - P(A_1^c)P(A_2)^c \cdots P(A_n^c) \geq .95 \\
&\Rightarrow 1 - [P(A_1^c)]^n \geq .95 \\
&\Rightarrow 1 - (0.96)^n \geq .95 \\
&\Rightarrow 0.96^n \leq .05
\end{aligned}$$

To solve the last inequality for n, you can plug in different values of n until you find the smallest n for which the inequality is true, or you can solve directly using logarithms. I use log for the natural logarithm.

$$
\begin{aligned}
.96^n \leq .05 \Rightarrow n\log 0.96 &\leq \log .05 \\
\Rightarrow n &\geq \frac{\log .05}{\log .96} \\
\Rightarrow n &\geq 73.38 \\
\Rightarrow n &= 74 \text{ is sufficiently large}
\end{aligned}
$$

Because n is the number of individuals, we need n to be an integer, so we round up. The inequality was reversed in the second step because we are dividing both sides by a negative number, i.e., $\log .96 < 0$.

8. Consider flipping a coin twice where the only possibilities are heads and tails. Let A be the event that both tosses result in the same value (both heads or both tails). Let B be the event that the second toss is heads. Find $P(AB)$ and $P(A \cup B)$.

Solution. $P(AB) = P(\{HH\}) = 1/4$. $P(A\cup B) = P(A) + P(B) - P(AB) = 1/2 + 1/2 - 1/4 = 3/4$.

9. Suppose $P(A^c) = 1/5$ and $P(B) = 1/4$. Can A and B be mutually exclusive? Why or why not?

Solution. No, they cannot be mutually exclusive because $P(A) = 4/5$ and $P(A)+P(B) = 4/5+1/4 = 21/20 > 1$. If A and B were mutually exclusive, then $P(A\cup B) = P(A)+P(B)$, but since this is greater than 1, this is impossible.

10. Suppose 5 indistinguishable balls are placed at random into 5 distinguishable bins. Find the probability that exactly one bin is empty. Hint: First determine that probability that the first bin is empty.

Let A be the event that exactly one bin is empty. Let A_i be the event that bin i is empty and that no other bin is empty. We note that $A_1, A_2, \ldots, A_5$ are disjoint. Then

$$
P(A) = \left(\bigcup_{i=1}^{5} A_i\right) = \sum_{i=1}^{5} P(A_i) = 5 \cdot P(A_1)
$$

By symmetry, the probabilities $P(A_i)$ should all be equal (if balls are distributed randomly, it is just as probable that bin 2 is empty as that bin 1 is empty). So, we can just worry about the probability that bin 1 is the only empty bin.

In order for bin 1 to be empty and the other bins not empty, we need to distribute the 5 balls into the 4 bins in such a way that every bin other than bin 1 is non empty. This means that three bins get one ball, and one bin gets two balls (if any bin got three or more balls, then there would be at least two empty bins).

Let's pretend that the balls are distinguishable, and determine the probability of one such arrangement. Then we can count how many ways the balls could be relabled. If we divided by the number of distinguishable arrangements, then it won't have mattered that the balls are indistinguishable.

For bins 2 through 5, exactly one bin gets two balls, and the rest get one. Suppose bin 2 gets two balls, and bins 3–5 get 1 ball each. If the balls were distinguishable, then we could choose two out of the five to go to bin 2, this can be done in $\binom{5}{2}$ ways. Then there are 3 choices for bin 3, 2 remaining choices for bin 4 and 1 choice remaining for bin 5. Thus the total number of ways to distribute the 5 balls to bins 2–5 where bin 2 gets two balls is

$$
\binom{5}{2} \cdot 3 \cdot 2 = 60
$$

There are an equal number of ways to distribute the balls such that bin i gets 2 balls and the remainder 1 ball, with $i = 3, 4, 5$. Therefore there are 240 ways to distribute the balls to bins 2–5 such that none are empty. The number of ways to distribute the balls when they are considered distinguished is $5^5 = 3125$ because there are 5 choices for location of the first ball, 5 choices for the location of the second ball, etc. Thus the probability that bin 1 is the one and only empty bin is

$$P(A_1) = \frac{240}{3125}$$

and the probability that exactly one bin is empty is

$$P(A) = 5 \cdot \frac{240}{3125} = \frac{1200}{3125} = \frac{48}{125} = 0.384$$

This is a relatively complicated problem, and it can be helpful to program a simulation to check your answer. Here is a simulation in R on the next page. I need about 1,000,000 iterations (about 15–20 seconds on my computer) to get the answer to within about 3 decimal places. I did some lazy programming and just made a for loop, but this is not very efficient in R. Code in MATLAB could be similar to this, but again it is better to avoid for loops if you want something more efficient. The code would execute faster if we simulated all random variables at once, put them into matrices, and did column sums or row sums instead of using for loops. I am willing to wait 20 seconds for an answer to make the programming easier, and I find for loops more natural to think about.

```
J=100000 # number of iterations to simulate

count <- 0 # count the number of times event A occurs

#iterate the simulation J times
for(j in 1:J) {

  bin <- rep(0,5) # initialize all bins to be empty
  u <- floor(runif(5)*5)+1 # 5 random integers between 1 and 5 to
                      # determine bin for the 5 balls

  #put balls in the bins.  For example if u[1]==3, then the first ball is put in bin 3
  # and bin[3] is incremented by 1.  bin[i] counts how many balls are in bin i
  # but doesn't distinguish the balls -- it only counts them.
  for(i in 1:5) {
    bin[u[i]] <- bin[u[i]] + 1
  }

  num0 <- 0 # number of empty bins
  #check each bin.  If it is empty, increment number of empty bins
  for(i in 1:5) {
    if(bin[i] == 0) num0 <- num0 + 1
  }
  if(num0==1) count <- count+1

}

#The proportion of times that event A occurs is count/J
#this is an estimate of the probability of event A
print(count/J)
```