CS1450 - Homework 1: Counting, Sets, Basic Probability

Due Date: September 22, 2020 at 2:20 PM (No late submissions accepted)

NOTE 1: For this and all homework assignments, we expect a typed response (using LaTeX) within this document. If you have LaTeX questions, consult the course website, Piazza, or Google. You should scan and embed any diagrams into this document, if necessary.

NOTE 2: We will be using Gradescope to handle this and all future assignments. Please leave yourself time to upload and submit.

Question 1

(a) Consider a class of students who took Algebra and Biology exam yesterday. Let S be the set of all students, among them let $A \subseteq S$ be the subset of students who failed Algebra, and $B \subseteq S$ be the subset of them who failed Biology. Draw a Venn diagram of these sets. Is the following statement True or False? "$|A \cup B| = |A| + |B|$" ?(we denote the size of a set using $|.|$) If False, can you correct it?

(b) Assume that we have 50 students, 3 have failed Algebra, 5 have failed Biology and 1 has failed both. How many people have failed either Algebra or Biology? If we take a student uniformly at random from this class what is the probability that he or she has failed at least Algebra or Biology?

(c) For two events A and B in a probability space, prove that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Hint: The part of A that is not contained in B can be written as $A \setminus B$.

(d) In parts (b), (c), and (d) we will walk through a proof by induction. For an event A_1, we know that $P(A_1) \geq P(A_1)$. Show that for events A_1, A_2, $P(A_1) + P(A_2) \geq P(A_1 \cup A_2)$.

(e) Assume $\sum_{i=1}^{n} P(A_i) \geq P(\bigcup_{i=1}^{n} A_i)$. Show for $n+1$, $\sum_{i=1}^{n+1} P(A_i) \geq P(\bigcup_{i=1}^{n+1} A_i)$

(f) For any n events $A_1, A_2, ..., A_n$ in a probability space, conclude that

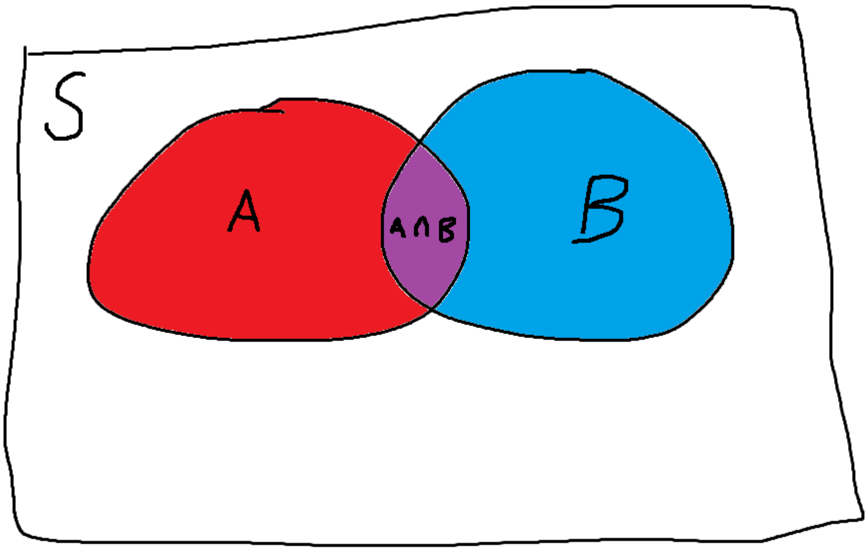
$$P(A_1) + P(A_2) + ... + P(A_n) \geq P(A_1 \cup A_2 \cup ... \cup A_n)$$

(g) Remember the class example of problems (a) and (b). Assume students took exams on Algebra, Biology, Computer Since, Data Science and Economics. Let $A \subset S$ be the subset of students who failed Algebra, $B \subseteq S$ the subset who failed Biology, $C \subseteq S$ the subset who failed Computer Science, $D \subseteq S$ the subset who failed Data Science and $E \subseteq S$ the subset who failed Economics. Assume $|S| = 50$, $|A| = 5$, $|B| = 3$, $|C| = 2$, $|D| = 8$ and $|E| = 3$. Alice is a student in this class that we have picked uniformly at random, find an upper bound on the probability that Alice has failed at least one course. Find a lower bound on the probability that Alice passed all the courses (Write both events using set notation).

Answer

(a) The statement above is false and should instead be:

$$|A \cup B| = |A| + |B| - |A \cap B|$$



(b) There are 7 students who have either failed either Algebra or Biology, as illustrated below. The probability of picking a student who has failed at least Algebra or Biology is 7 in 50, or 14%.

$$|A \cup B| = 3 + 5 - 1$$
$$|A \cup B| = 7$$

(c) I will first assume that the statement $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ is true.

(1) $P(A \cup B) = P(A \setminus B) + P(B \setminus A) + P(A \cap B)$: I got to this equation by the definition of union. The union of A and B contains all the elements in A that are not in B, the elements in B that are not in A, and the elements that are in both A and B. Assuming the conjecture is true, I can substitute $P(A) + P(B) - P(A \cap B)$ for $P(A \cup B)$

(2) $P(A)+P(B)-P(A\cap B) = P(A\setminus B)+P(B\setminus A)+P(A\cap B)$: I can use algebra to transform this equation.

(3) $P(A) - P(A \cap B) + P(B) - P(A \cap B) = P(A \setminus B) + P(B \setminus A)$: By definition, $P(A \setminus B) = P(A) - P(A \cap B)$ and the same can be said for $P(B \setminus A)$. Using this definition, I can substitute in $P(A \setminus B)$ and $P(B \setminus A)$ to demonstrate the validity of the equation.

(4) $P(A \setminus B) + P(B \setminus A) = P(A \setminus B) + P(B \setminus A)$

(d) I know that $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$ and I also know $P(A_1) \geq P(A_1)$ and $P(A_2) \geq P(A_2)$

(1) $P(A_1) \geq P(A_1)$: I can add $P(A_2)$ to both sides of the inequality

(2) $P(A_1) + P(A_2) \geq P(A_1) + P(A_2)$: From here, If I subtract $P(A_1 \cap A_2)$ from the right side of the inequality, it still remains true as $P(A_1 \cap A_2)$ can never be negative.

(3) $P(A_1) + P(A_2) \geq P(A_1) + P(A_2) - P(A_1 \cap A_2)$: Now I can substitute $P(A_1 \cup A_2)$ into the right side of the inequality

(4) $P(A_1) + P(A_2) \geq P(A_1 \cup A_2)$: This gives us the conclusion I was looking for.

(e) I can write $\sum_{i=1}^{n+1} P(A_i)$ as the sum of $\sum_{i=1}^{n} P(A_i)$ and $P(A_{n+1})$ as explained by algebraic properties. I can also write $P(\bigcup_{i=1}^{n+1} A_i)$ as the union of the two smaller groups $P(\bigcup_{i=1}^{n} A_i)$ and $P(A_{n+1})$. I can rewrite the initial assumption and add $P(A_{n+1})$ to both sides of the inequality. From here, I notice I can use substitution to obtain a relationship between $\sum_{i=1}^{n+1} P(A_i)$ and $P(\bigcup_{i=1}^{n+1} A_i)$

$$\sum_{i=1}^{n+1} P(A_i) = \sum_{i=1}^{n} P(A_i) + P(A_{n+1})$$
$$P(\bigcup_{i=1}^{n+1} A_i) = P(\bigcup_{i=1}^{n} A_i) \cup P(A_{n+1})$$
$$\sum_{i=1}^{n} P(A_i) \geq P(\bigcup_{i=1}^{n} A_i)$$
$$\sum_{i=1}^{n} P(A_i) + P(A_{n+1}) \geq P(\bigcup_{i=1}^{n} A_i) + P(A_{n+1})$$
$$\sum_{i=1}^{n+1} P(A_i) \geq P(\bigcup_{i=1}^{n+1} A_i)$$

(f)

(g) Upper bound = $\{|A|, |B|, |C|, |D|, |E|\} \ /|S| = 21/50 = 0.42$ where all of the events for failing a class are disjoint
Lower bound = $\{|D|\} \ /|S| = 8/50 = 0.16$ where the 8 people who failed Data Science were also the same people who failed all of the other classes.

Question 2

You have a standard deck of 52 playing cards, shuffled in a random order, and you draw cards one at a time from the deck.

(a) What is the probability of drawing five cards in consecutive increasing order (i.e. the value of the second card is one greater than the value of the first and so on), where jacks, queens, kings, and aces have values 11, 12, 13, and 14, respectively.

(b) What is the probability of drawing a hand that can be arranged to be in consecutive increasing order?

(c) Without looking at the actual numbers obtained above, which probability is higher between (a) and (b)? Can you provide an intuitive explanation as to why?

(d) Now, assume that after each card is drawn, you shuffle it back into the deck. Now what is the probability that the five cards you drew were in increasing consecutive order?

(e) Without looking at the actual numbers obtained above, which probability is higher between (a) and (d)? Can you provide an intuitive explanation on why?

(f) Given that you drew five cards in increasing consecutive order, what is the probability that all of the cards are red *or* you have exactly three diamond suits (the diamond suit is red)?

Answer

(a) For this situation, the probability of choosing a valid 1st card differs from choosing a valid 2nd, 3rd, 4th, or 5th card. 5 cards can only be drawn in consecutive increasing order if the first card is between 2 and 10 (inclusive). Therefore,

$$P(\text{drawing a valid first card}) = \frac{36}{52}$$

$$P(\text{drawing 5 cards in consecutive increasing order}) = (\frac{36}{52})(\frac{4}{51})(\frac{4}{50})(\frac{4}{49})(\frac{4}{48})$$

$$P(\text{drawing 5 cards in consecutive increasing order}) = 2.95x10^-5$$

(b) The number of combinations to achieve a hand of five cards in consecutive increasing order is represented by $(10)(4)^5 = 10,240$. The number of combinations to achieve a hand of just five cards is $\binom{52}{5} = 2,598,960$. To find the probability of achieving a hand of five cards in consecutive increasing order, divide the number of five card consecutive increasing order hands by the total amount of five card hands.

c.i.o = consecutive increasing order

$$P(\text{drawing 5 cards that can be arranged in c.i.o}) = \frac{(10)(4)^5}{\binom{52}{5}}$$

$$P(\text{drawing 5 cards that can be arranged in c.i.o}) = \frac{10,240}{2,598,960}$$

$$P(\text{drawing 5 cards that can be arranged in c.i.o}) = 0.00394$$

$$P(\text{drawing 5 cards that can be arranged in c.i.o}) \approx 0.4\%$$

(c) The probability described in (b) is higher. This is because (b) does not put a constraint on the order in which cards are picked. As a result, there are different permutations of the same cards that can lead to a hand in consecutive increasing order, thus increasing the probability.

(d) This situation is very similar to the situation described in (a). The difference here is that we are drawing with replacement. Therefore, the total number of cards between each draw remains the same.

$$P(\text{drawing 5 cards in consecutive increasing order}) = (\frac{36}{52})(\frac{4}{52})(\frac{4}{52})(\frac{4}{52})(\frac{4}{52})$$
$$P(\text{drawing 5 cards in consecutive increasing order}) = 2.42x10^-5$$

(e) The probability described in (a) is higher. In (a) we are drawing without replacement. This means that for each card we draw, we do not put it back in the deck. Each subsequent draw has a smaller pool of "incorrect" cards to choose from, which increases the probability of choosing a card we want. There is a higher probability of choosing a desirable card on the 2nd, 3rd, 4th, and 5th draws when we draw without replacement, therefore the probability of drawing 5 cards in consecutive increasing order is higher.

(f)

Question 3

In the following questions you are asked to plot some probabilities. Use Matlab or Python as your programming language. Each trial is an event when you generate the information you need randomly. For example if you are comparing birthdays of 50 people, use a random number generating function from Matlab or Python and generate 50 birthdays in each trial.

(a) Write a program that simulates the birthday problem by counting how many times two or more people have the same birthday in CS 1450 for $2 <= n <= 70$ students for 10000 trials. Assume no one has a birthday on a leap year.

(b) Plot these probabilities on a graph. If there are 25 people in the class, what is the probability that two people share a birthday?

(c) Write a program that counts how many times a student shares Eli's birthday, for $2 <= n <= 70$ people (n includes Eli) for 10000 trials.

(d) Plot these probabilities on a the same graph as in (b). If there are 25 people in the class, what is the probability that a person shares a birthday with Eli? Provide an intuitive explanation for why this answer is different than that of (b).

(e) On another planet (far, far away), a class of super-intelligent aliens is also taking CS 1450. There are 25 aliens in the class, but there are N days in their year and no leap years. Write a program that simulates the birthday problem by counting how many times two or more aliens share a birthday in CS 1450 for $365 < N < 500$ days in the year for 10000 trials.

(f) Plot these probabilities on a graph. What is the largest number of days in a year such that there is a more than 0.5 chance that at least two aliens share a birthday? NOTE: A rough estimate is okay for this part.

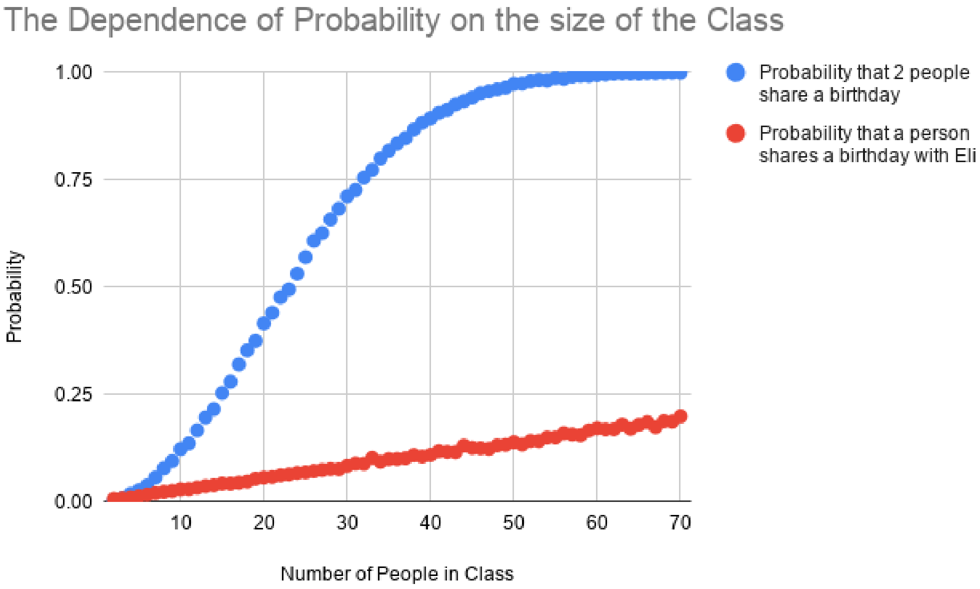
Answer

(a)

```
# birthday function
# Calculates the probability that two or more people
# have the same birthday for 2 <= n <= 70 students in 10,000 trials
# Writes these percentages to a file titled percentages.txt
def birthday():

file1 = open("percentages.txt", "w")
for i in range(2, 71):
    percentage = 0
    for j in range(10000):
        seenDays = []
        counter = 0
        randDay = 0
        while counter != i:
            randDay = randint(1, 366)
            if randDay in seenDays:
                percentage += 1
                break
            else:
                seenDays.append(randDay)
                counter += 1
    percentage = percentage / 10000.000
    toWrite = str(percentage) + "\n"
    file1.write(toWrite)
    print(percentage, "\n")
file1.close()
```

(b) The probability that two or more people share a birthday when n = 25 is 0.5697 or around 57%.



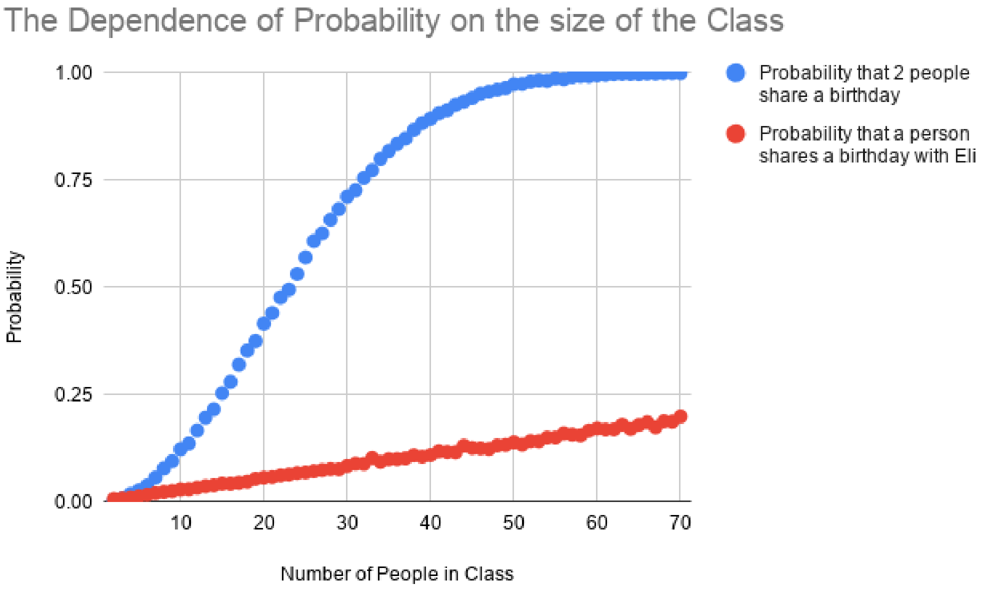
(c)

```
# eliBirthday function
# Chooses a random day to represent Professor Eli's birthday
# For class sizes if 2 <= n <= 70, finds the average amount
# of people
# Writes these probabilities to a file called eli.txt

def eliBirthday():
file1 = open("eli.txt", "w")
eliBirth = randint(1, 366)
for i in range(2, 71):
    percentage = 0
    for j in range(10000):
        counter = 0
        randDay = 0
        while counter != i:
            randDay = randint(1, 366)
            if randDay == eliBirth:
                percentage += 1
            else:
                counter += 1
    percentage = percentage / 10000.000
    toWrite = str(percentage) + "\n"
    file1.write(toWrite)
```

```
        print(percentage, "\n")
    file1.close()
```

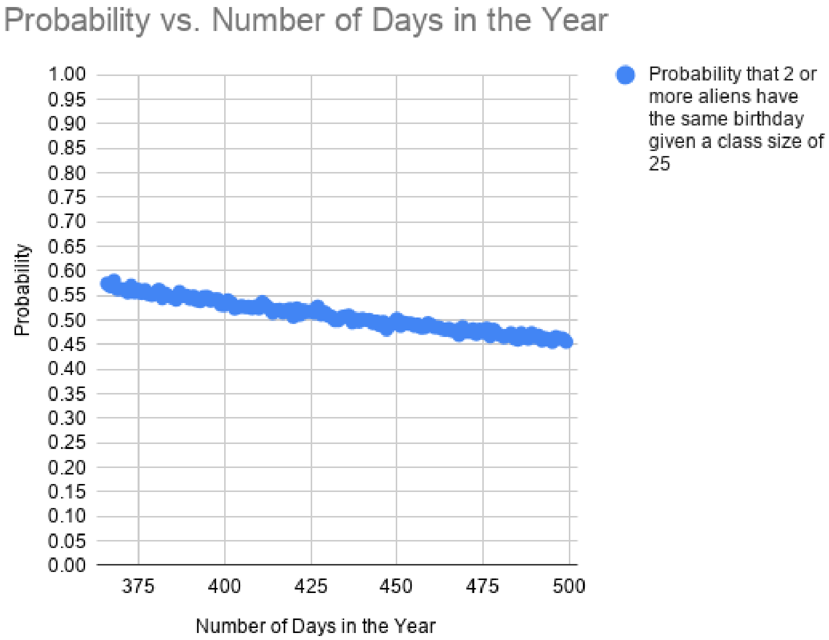
(d) If there are 25 people in the class, the probability that at least one of them shares a birthday with Professor Eli is 0.0636, or around 6%. This probability is much lower than that of (b). Intuitively, this makes sense because there are less dates to match with. In this scenario, when you go through all of the birthdays of the people in the class, the requirement is that they match with Eli's birthday. In (b) however, when you go through all the birthdays, the requirement is that the birthday you're looking at matches with a birthday you've already seen before. As you continue through the list of birthdays, the number of other birthdays to match with increases. The restriction to only match with one particular date causes this probability to be much lower.



(e)

```
# alienBirthday function
# Parameters:
#   - num, an int, represents the size of the class
# Calculates the probability that two or more people
# have the same birthday for 365 < N < 500 days in the year for
# 10,000 trials
def alienBirthday(num):
file1 = open("alien.txt", "w")
classSize = num
for i in range(366, 500):
    percentage = 0
    for j in range(10000):
        seenDays = []
        counter = 0
        randDay = 0
        while counter != classSize:
            randDay = randint(1, i)
            if randDay in seenDays:
                percentage += 1
                break
            else:
                seenDays.append(randDay)
                counter += 1
    percentage = percentage / 10000.000
    toWrite = str(percentage) + "\n"
    file1.write(toWrite)
    print(percentage, "\n")
file1.close()
```

(f) The largest number of days in a year such that there is a more than 0.5 chance at least two aliens share a birthday is around 437 days.


Probability vs. Number of Days in the Year
Probability
Number of Days in the Year
● Probability that 2 or more aliens have the same birthday given a class size of 25