Probability

1. Before we begin calculating probabilities, it is worthwhile to think about what probabilities mean in practice. Discuss the following in your group:
   (a) What does it mean to say a real-world event has a particular probability of occurring? (For example, what does 25% mean in the claim "There is a 25% chance of rain tomorrow"?)
      i. Does your answer change if you are describing an uncertain phenomenon that can't be repeated? (For example, in the statement "Candidate A has a 25% chance of winning the 2022 election.")
      ii. Does your answer change if the phenomenon has a cost and reward? (For example, suppose a casino game costs $1 to play and the operator states it has a 25% chance of awarding $4)
   (b) Suppose someone claims that an event has a particular probability. How could we verify or falsify this claim?
   (c) What is the benefit of quantifying uncertainty? What are some negative consequences?

The Birthday Problem

The classic birthday problem states:

There are k people in a room. Assume that each person's birthday is equally likely to be any of the 365 days of the year (excluding Feb. 29), and that people's birthdays are independent. What is the probability that one pair of people in the group have the same birthday?

Using the formulas for sampling with and sampling without replacement, we can show that the probability of at least one birthday match is

$$P(\text{at least 1 birthday match}) = 1 - \frac{365 \cdot 364 \cdots (365-k+1)}{365^k} = 1 - \frac{365!}{(365-k)!365^k}$$

2. Use similar reasoning to determine the probabibilities of a match in the following 4 cases. Assume in each case that all numbers are equally likely, that student answers are independent of one another, and that our class has 25 students). Report your answer both as a fraction and as a decimal approximation.
   (a) Each student reports their birthday.
   (b) Each student reports their favorite integer between 1 and 100.
   (c) Each student reports the last three digits of their phone number.
   (d) Each student randomly chooses a seat in a stadium with 25,000 seats.
3. In the previous problem, we assumed that all numbers were equally likely, and that each student's answer was independent of another (that is, knowing one student's answer gives no information about another student's answer).
   (a) In each of the 4 cases, how reasonable are these assumptions?
   (b) If these assumptions are incorrect, do you think the probability of a match will be greater or less than the value you calculated? (Think about simple as well as extreme cases to help build your intuition)
4. (*) A survey is being conducted in a city with 1 million residents. It would be far too expensive to survey all of the residents, so a random sample of size 1000 is chosen. The survey is conducted by choosing people one at a time, *with* replacement and with equal probabiltities.
   (a) Explain how sampling with vs. without replacement here relates to the birthday problem.
   (b) Find the probability that at least one person will get chosen more than once.

(*) Indicates problems that will also be collected for homework.