Lecture 4 Worksheet

Chrysafis Vogiatzis

Every worksheet will work as follows.

1. You will be asked to form a group with other students in the class: you can make this as big or as small as you'd like, but groups of 4-5 work best.
2. Read through the worksheet, discussing any questions with the other members of your group.
   - You can call me at any time for help!
   - I will also be interrupting you for general guidance and announcements at random points during the class time.
3. Answer each question (preferably in the order provided) to the best of your knowledge.
4. While collaboration between students is highly encouraged and expected, each student has to submit their own version.
5. You will have 24 hours (see gradescope) to submit your work.

Activity 1: The law of total probability

Problem 1: Australian Open

In a game of tennis [1] two players compete for the next point. The person serving has (historically) an advantage and wins the next point with probability 70%. This probability changes to 50% (so both players are equally probable to win the next point), if the person serving loses the point [2]. When the person serving wins the point again, the probability returns to 70%.

What is the probability the player serving wins two points in a row? What is the probability the player serving wins the second point (no matter what happens in the first point)? What is the probability the other player wins the second point (no matter what happens in the first point)?

Answer to Problem 1.

Server wins both points:

Server wins second point:

Opponent wins second point:

[1] The U.S. Open is taking place these days; hence the motivation for this question!

[2] You may assume that this is because tennis is also a sport of momentum and morale.

Problem 2

A company needs to make a decision on whether to open a new facility or not. Their chief economic strategist has recommended one of three options:

1. Invest in a new facility.
2. Play it safe and invest in renovations in existing facilities.
3. Cut down on all expenses.

The market next year can be bullish, bearish, or stagnant [3]. The economic strategist has identified the following probabilities of success for the company depending on the economy:

	Market		
Option	Bullish	Bearish	Stagnant
Option 1	80%	5%	30%
Option 2	50%	50%	80%
Option 3	10%	95%	40%

[3] A bullish market is recognized by an increase in prices, due to the market participants having optimistic views of the economy; a bearish market on the other hand sees a decline in prices, as market participants become pessimistic of the economy. Finally, a stagnant economy sees neither increases or declines in prices and prices stay in a constant level.

What should the company choose to do assuming that next year's market is expected to be bullish with probability 30%, bearish with probability 30%, and stagnant with probability 40%? To answer the question, get the probability of success for each of three options and pick the one with the highest.

Answer to Problem 2.

Option 1: $P(success) =$

Option 2: $P(success) =$

Option 3: $P(success) =$

Activity 2: Using Bayes' theorem

In this activity, we will bring smaller and bigger applications of the Bayes' theorem. We already saw an interesting one that has to do with the Mantoux test (any diagnostic tool in general). Here, we will see examples from fair grading to detecting spam messages and blood supply logistics.

Problem 3: Fair grader

A quiz in IE 300 is scheduled to have just one multiple choice question with 5 different answers. A student decides to only study half the material for the quiz: hence with probability 50% they will know the answer to the question, and with probability 50% they will guess an answer at random [4]. A student receives a perfect score in the quiz! What is the probability they guessed the answer?

[4] Recall: this implies that all 5 answers are equally probable events!

Answer to Problem 3.

Problem 4: SPAM?

Gmail has observed that the word "inheritance" appears in 20% of all spam emails. It appears at only 0.1% of all known non-spam email communications. Roughly, the estimate right now is that 45% of all email communications are spam. You received an email entitled **Inheritance**: what is the probability it is spam?

Answer to Problem 4.

Problem 5: A need for blood

In Greece, blood type distribution is as follows: 44.4% type O, 37.9% type A, 13% type B, and 4.7% type AB [5]. However for people born in the 1960s or earlier, typing was not correctly done. We now know that:

[5] To be clear: blood types are mutually exclusive and collectively exhaustive.

- for a person with blood type A, they would be (correctly) found as A with probability 85%;
- for a person with blood type O, they would be (incorrectly) found as A with probability 5%;
- for a person with blood type B, they would be (incorrectly) found as A with probability 15%;
- for a person with blood type AB, they would be (incorrectly) found as A with probability 25%;

A person born in the period before the 1960s is brought to a hospital and needs blood. They are listed as having type A blood: what is the probability they actually need type A blood?

Answer to Problem 5.

Activity 3: Deriving Bayes' theorem

In Bayes' theorem, we assume the existence of n states or hypotheses $S_i, i = 1, \ldots, n$ that are either reinforced or weakened through the appearance of m test outcomes or evidence $O_j, j = 1, \ldots, m$.

Bayes' theorem also assumes the availability of prior information in the form of probabilities for each state ($P(S_i), i = 1, \ldots, n$) and the availability of historical information in the form of probabilities for each outcome depending on the state ($P(O_j|S_i), j = 1, \ldots, m, i = 1, \ldots, n$).

Finally, we are concerned with deriving what $P(S_i|O_j)$ or what is the probability that state/hypothesis S_i is true given the existence of a test outcome or evidence O_j.

With these in mind, answer the following questions.

Problem 6

Write $P(S_i|O_j)$ using the conditional probability formula. [6]

[6] How can we calculate a conditional probability? Check the previous lecture notes..

Answer to Problem 6.

Problem 7

You should have that $P(S_i|O_j)$ can be written as a fraction of two probabilities. Focus on the probability in the numerator, $P(S_i \cap O_j)$. In a sentence, explain why $P(S_i \cap O_j) = P(O_j \cap S_i)$.

Answer to Problem 7.

Problem 8

Use the multiplication rule to write $P(O_j \cap S_i)$. [7]

Answer to Problem 8.

[7] Recall that the multiplication rule for probabilities states that

$$P(A \cap B) = P(B) \cdot P(A|B).$$

Problem 9

Go back to your answer in Problem 6, however now focus on the denominator, $P(O_j)$. Use the total probability law to write $P(O_j)$ as a function of $P(S_i), i = 1, \ldots, n$ and $P(O_j|S_i), i = 1, \ldots, n$.

Answer to Problem 9.

Problem 10

Combine your answer to Problem 8 and Problem 9 to show what $P(S_i|O_j)$ is according to Bayes' theorem.

Answer to Problem 10.

As a reminder, Bayes' theorem states the following: given n mutually exclusive and collectively exhaustive events S_i with probabilities $P(S_i)$, and m outcomes O_j with probabilities $P(O_j|S_i)$, then:

$$\boxed{P(S_i|O_j) = \frac{P(S_i) \cdot P(O_j|S_i)}{\sum\limits_{i=1}^{n} P(S_i) \cdot P(O_j|S_i)}.}$$

Activity 4: Flipping a flipped classroom

Problem 11

Create one exercise using the Bayes' theorem. To help you:

1. Find an interesting application where you need to know something and you may run a test to figure it out.
2. Look online to see if you can find probabilities that appear realistic.
3. Pose the question of searching for a probability.
4. Solve the question. [8]

[8] Bonus points :) if the result is counter-intuitive!

Answer to Problem 11.