DA202-O Introduction to Data Science
Problem Set 2

Due on: 08 September 2021

---

1. We are given three coins. The first coin is a fair coin painted blue on the heads side and white on the tails side. The other two coins are biased so that the probability of heads is p. They are painted blue on the tails side and red on the heads side. Two of the three coins are to be selected at random and flipped. Describe the outcomes in the sample space. It was experimentally determined that the probability that the sides that land face up are the same color is $\frac{29}{96}$. What are the possible values of p? [5]
   key concepts: [Sample Space 50%, Events 50%]

2. An experiment consists of picking a student from the set of all students registered for DA202 this semester. Assume that all students are equally likely to be picked. [5]
   (a) Consider the two events:
   A: the student has Three Years of Industry experience (TYI)
   B: the student has a degree in Electrical Engineering (EE)
   For any student picked, if the probability that he/she has had neither TYI nor EE is 0.3, and the probability that he/she has missed at least one of the two is 0.8, what is the probability that he/she has had exactly one of the two?
   (b) Let C denote the event that the student is registered in DA231 this semester. Let events A and B and their probabilities be as in part (a). If students who had at most one of TYI and EE did not register for DA231 this semester,
      i. What is P($A^c \cap B^c \cap C^c$)?
      ii. What is the probability that the student picked is not registered for DA231 and has had exactly one of TYI or EE?

   key concepts: [Sample Space 50%, Events 50%]

3. A researcher, Y, tests the effectiveness of Covishield vaccine against Covid-19 infection. Based on evidence, it is hypothesised that for any large group of patients infected with Covid-19, 10% have an incurable form of the disease and respond to no vaccine. Unfortunately, there is no test to find out whether a patient has a curable or incurable form of the disease. So, both types of patients are indistinguishable at the outset. It is also known (from literature and expert opinion) that 50% of patients in a large group of patients recover without any vaccine. In a blind study (a type of clinical study where the patient doesn't know if they received a treatment or placebo, but the researcher knows. Such trials are also called as A/B testing or Randomized Control Trials (RCT)) over a very large group of patients, Y places one half of the patients on Covishield and the other half on a placebo. Patients who do not recover die. [5]
   (a) Y finds out that 80% of the patients on Covishield recover. What is the percentage of patients with a curable form of Covid-19 that are cured by Covishield?
   (b) What fraction of patients with a curable form of Covid-19 die in the placebo group?

(c) An unscrupulous researcher, W, finds a test to determine with certainty whether a patient has a curable or incurable form of Covid-19, but keeps his findings secret.

W wants to test drug V and wants to show that it is more effective than Covishield. W runs a blind study over a very large group of patients, placing one half of the patients on drug V and the other half on a placebo. W does not get to pick the members of the study. However, W gets to pick which patients go in which group and can run tests on the patients before performing his selection.

W's research assistants then run the study. The assistants are honest and the data for the group on drug V and the placebo group are recorded accurately. The study shows that 85% of the group on V recovers from Covid-19 and W declares victory.

With your knowledge of W's dastardly ways, what can you say about the percentage of patients with a curable form of Covid-19 that are cured by V?

(d) How can Y uncover W's deception without running another study?

key concepts: [Events 30%, Conditional Probability 70%]

4. Gopal and Ram each rotates a wheel of fortune and scores a number in the interval [0, 1] randomly independent of each other. We assume a uniform probability law under which the probability of an event is proportional to its area. Consider the following events: [5]

A : The magnitude of the difference of their scores is greater than 1/4.
B : At least one of their scores is greater than 1/4.
C : Both score equally.
D : Gopal's score is greater than 1/4.

Find the probabilities P(B), P(C), and P($A \cap D$).

key concepts: [Independence 50%, Uniform probability law 50%]

5. Oscar has lost his dog in either forest A (with a priori probability 0.4) or in forest B (with a priori probability 0.6). On any given day, if the dog is in A and Oscar spends a day searching for it in A, the conditional probability that he will find the dog that day is 0.25. Similarly, if the dog is in B and Oscar spends a day looking for it there, the conditional probability that he will find the dog that day is 0.15. The dog cannot go from one forest to the other. Oscar can search only in the daytime, and he can travel from one forest to the other only at night. [5]

(a) In which forest should Oscar look to maximize the probability he finds his dog on the first day of the search?

(b) Given that Oscar looked in A on the first day but didn't find his dog, what is the probability that the dog is in A?

(c) If Oscar flips a fair coin to determine where to look on the first day and finds the dog on the first day, what is the probability that he looked in A?

(d) If the dog is alive and not found by the Nth day of the search, it will die that evening with probability $\frac{N}{N+2}$. Oscar has decided to look in A for the first two days. What is the probability that he will find a live dog for the first time on the second day?

key concepts: [Conditional probability 33.3%, Bayes rule 33.3%, Total probability theorem 33.3%]

6. There are four doors, one and only one of which conceals a prize. You pick a door and the game-show host opens one of the remaining doors that does not conceal a prize. You are given the choice of sticking to your original decision or switching to either of the other two unopened doors. [5]

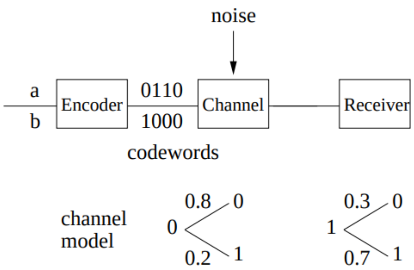
(a) Let the door you originally picked be door 1. What is the probability that the prize is behind door 4?

(b) Suppose you picked 1, and the host opened door 2. What is the conditional probability that the prize is behind door 4? What is the conditional probability that the prize is behind door 1?

Whether you switch or not, the host opens another door that does not have the prize behind it (so now two doors have been opened, one of which could be your initial choice if you switched). You are again offered the option to switch. Should you switch or keep your choice in the two cases that you are given a choice, if your objective is to maximize the probability of winning?

key concepts: [Conditional Probability 100%]

7. A transmitter wishes to send one of two alternative messages, "a" or "b". Because of the way these messages originate, it is known that "b" messages are twice as likely as "a" messages. These messages are encoded into binary messages, in order to be transmitted over a digital channel. Assume that "a" and "b" are encoded as 0110 and 1000, respectively. However, the channel is noisy and each bit transmitted may be received incorrectly, according to the probabilities shown below. For example, if a zero is sent, there is probability 0.2 that a [5]



1 is received. Assume that errors during the transmission of different bits are statistically independent. Given that the receiver received the sequence 0011, find the probability that message "a" was transmitted.

key concepts: [Bayes rule 50%, Total probability theorem 50%]

8. The minimum number of members of the House of parliament required to be present for a proceeding to start is called Quorum. The Presiding Officer has power to prorogue a sitting in case of absence of quorum or suspend the sitting until there is quorum. The Presiding Officer decides that he will prorogue unless at least k of the n members are present. Each member will independently show up with probability p if the weather is good, and with probability q if the weather is bad. Given the probability of bad weather on a given day, obtain an expression for the probability that the Presiding officer will not prorogue the session on that day. Also find the probability that the Presiding officer will not prorogue the session on any day. [5]

key concepts: [Binomial distribution 50%, Total probability theorem 50%]

9. Albus Dumbledore, headmaster of Hogwarts School of Witchcraft and Wizardry founded the Order of the Phoenix when Lord Voldemort declared war on the wizarding world. Only 1% of the witches and wizards joined this secret society. The Order of the Phoenix decided that it must have its own secret code, so members may easily tell each other apart from non-members. Unfortunately, the code they choose is so complicated that a member has a 12% chance of thinking another member is an impostor, but a 93% chance of recognizing a non-member to be an impostor. Given that after administering the secret code, a member decides someone is also a member, find the probability that that person is in fact a real member. [5]

key concepts: [Bayes rule 50%, Total probability theorem 50%]

10. Most mornings, Victor checks the weather report before deciding whether to carry an umbrella. If the forecast is "rain," the probability of actually having rain that day is 80%. On the other [5]

hand, if the forecast is "no rain," the probability of it actually raining is equal to 10%. During fall and winter the forecast is "rain" 70% of the time and during summer and spring it is 20%.

(a) One day, Victor missed the forecast and it rained. What is the probability that the forecast was "rain" if it was during the winter? What is the probability that the forecast was "rain" if it was during the summer?

(b) The probability of Victor missing the morning forecast is equal to 0.2 on any day in the year. If he misses the forecast, Victor will flip a fair coin to decide whether to carry an umbrella. On any day of a given season he sees the forecast, if it says "rain" he will always carry an umbrella, and if it says "no rain," he will not carry an umbrella. Are the events "Victor is carrying an umbrella," and "The forecast is no rain" independent? Does your answer depend on the season?

(c) Victor is carrying an umbrella and it is not raining. What is the probability that he saw the forecast? Does it depend on the season?

key concepts: [Conditional probability 20%, Bayes rule 40%, Total probability theorem 40%]

11. **The problem of points** [5]

Tina and Vikram play a round of golf (18 holes) for a 1000 Rupees stake, and their probabilities of winning on any one hole are p and 1 - p, respectively, independent of their results in other holes. At the end of 10 holes, with the score 4 to 6 in favor of Vikram, Tina receives an urgent call and has to report back to work. They decide to split the stake in proportion to their probabilities of winning had they completed the round, as follows. If p_t and p_v are the conditional probabilities that Tina and Vikram, respectively, are ahead in the score after 18 holes given the 4-6 score after 10 holes, then Tina should get a fraction $\frac{p_t}{p_t+p_v}$ of the stake, and Vikram should get the remaining $\frac{p_v}{p_t+p_v}$. How much money should Tina get?

Note: This is an example of the, so-called, problem of points, which played an important historical role in the development of probability theory. The problem was posed by Chevalier de Mere in the 17th century to Pascal, who introduced the idea that the stake of an interrupted game should be divided in proportion to the players' conditional probabilities of winning given the state of the game at the time of interruption. Pascal worked out some special cases and through a correspondence with Fermat, stimulated much thinking and several probability-related investigations.

key concepts: [Independence 30%, Binomial distribution 70%]

12. A communication system transmits one of three signals, s_1, s_2 and s_3, with equal probabilities. [5]
The transmission is corrupted by noise, causing the received signal to be changed according to the following table of conditional probabilities:

		Receive, j		
	$P(s_j \mid s_i)$	s_1	s_2	s_3
Send, i	s_1	0.25	0.5	0.25
	s_2	0.04	0.9	0.06
	s_3	0.8	0.15	0.05

For example, if s_1 is sent, the probability of receiving s_3 is 0.25. The entries of the table list the probability of s_j received, given that s_i, is sent, i.e., P(s_j, received|s_i, sent).

(a) Compute the (unconditional) probability that s_j, is received for j = 1, 2, 3.

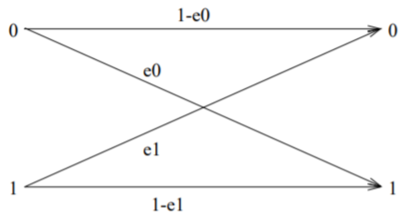
(b) Compute the probability P(s_i, sent|s_j, received) for i, j = 1, 2, 3.

(c) As seen from the numbers above, the transmitted signal can be very different from the received signal. Thus, when symbol s_j is received, it may not be a good idea to conclude that s_j was sent. We need a decision scheme to decide which signal was sent based on the signal we receive with the lowest possible error probability, that is, the probability of making a wrong decision. Find the scheme that minimizes the overall error probability.

key concepts: [Total probability theorem 50%, Bayes rule 50%]

13. **Communication through a noisy channel** [5]

A binary (0 or 1) message transmitted through a noisy communication channel is received incorrectly with probability e0 and e1, respectively. Errors in different symbol transmissions are independent. The channel source transmits a 0 with probability p and transmits a 1 with probability $1 - p$.



(a) What is the probability that a randomly chosen symbol is received correctly?

(b) Suppose that the string of symbols 1011 is transmitted. What is the probability that all the symbols in the string are received correctly?

(c) In an effort to improve reliability, each symbol is transmitted three times and the received symbol is decoded by majority rule. In other words, a 0 (or 1) is transmitted as 000 (or 111, respectively), and it is decoded at the receiver as a 0 (or 1) if and only if the received three-symbol string contains at least two 0s (or 1s, respectively). What is the probability that a transmitted 0 is correctly decoded?

(d) Suppose that the scheme of part (c) is used. What is the probability that a 0 was transmitted given that the received string is 101?

key concepts: [Independence 100%]

14. You have 2 five-sided dice. The sides of each are numbered from 1 to 5. The dice are "fair," and each die roll is independent of all others. [5]

(a) Event A is "you roll both dice, and the total is 10" (i.e., if you add the number that comes up on one die to the number on the other die, the total is 10).

   i. Is event A independent of the event "at least one of the dice has a 5 showing"?

   ii. Is event A independent of the event "at least one of the dice has a 1 showing"?

(b) Event B is "you roll both dice, and the total is 8."

   i. Is event B independent of getting "doubles" (i.e., both dice are showing the same number)?

   ii. Given that you rolled both dice and the total was 8, what is the probability that at least one of the dice has a 3 showing?

   iii. Given that you rolled both dice and the total was 8, what is the probability that at least one of the dice has a 5 showing?

key concepts: [Independence 70%, Conditional probability 30%]

15. A cellular phone system services a population of n_1 "voice users" (those who occasionally need a voice connection) and n_2 "data users" (those who occasionally need a data connection). We estimate that at a given time, each user will need to be connected to the system with probability P_1 (for voice users) or P_2 (for data users), independent of other users. The data [5]

rate for a voice user is r_1 bits/sec and for a data user is r_2 bits/sec. The cellular system has a total capacity of c bits/sec. What is the probability that more users want to use the system than the system can accommodate?

key concepts: [Independence 100%]

16. Your office parking lot consists of N parking spaces in a row ($N \geq 5$), where you and your colleague makes an equally likely choice among all available locations. Calculate the probability that the parking spaces you and your colleague selects are exactly 4 apart (or exactly three empty spaces between them). [5]

    key concepts: [Counting 100%]

17. **The hat problem** [5]

    The hats of n persons are thrown into a box. The persons then pick up their hats at random (i.e., so that every assignment of the hats to the persons is equally likely). What is the probability that

    (a) every person gets his or her hat back?

    (b) the first m persons who picked hats get their own hats back?

    (c) everyone among the first m persons to pick up the hats gets back a hat belonging to one of the last m persons to pick up the hats?

    key concepts: [Counting 100%]

18. An academic department offers 8 lower level courses: {L1 , L2, ... L8} and 10 higher level courses: {H1, H2,..., H10 }. A valid curriculum consists of 4 lower level courses and 3 higher level courses. [5]

    (a) How many different curricula are possible?

    (b) Suppose that {H1, ... , H5} have L1 as a prerequisite, and {H6 ,... H1O} have L2 and L3 as prerequisites. i.e .. any curricula which involve, say, one of {H 1, ... , H5} must also include L1 . How many different curricula are there?

    key concepts: [Counting 100%]

19. There are n people enrolled in the DA202 course. We are interested in the birthday of each of these people. Assume that every person has an equal probability of being born on any day during the year and ignore the additional complication presented by leap years (365 days every year). What is the probability that no two of them celebrate their birthday on the same day of the year (i.e., each person has a distinct birthday)? [5]
    key concepts: [Counting 100%]

20. Streets in many Bangalore localities are laid in a rectangular grid pattern with Main Road running north-south and Crosses running east-west. Consider a grid with 10 Main Roads and 10 Crosses with numbering from North (1) to South (10) for Crosses, and West (1) to East (10) for the Main Roads. Your house is at 1st Main, 1st Cross on the northwest corner. You want to go to the Metro station at 10th Main 10th Cross at the southeast corner. There is a Nandini stall at 6th Main, 7th Cross. If you randomly take turns to South or East only (turning North or West is not allowed as BTP has made it one way) in your daily ride to the Metro station, what is the probability that you cross the Nandini stall? [5]

    key concepts: [Counting 100%]