MAST20004 Probability

Tutorial Set 2

Tutorial problems:

1. In lectures, we discussed the experiment where we repeatedly roll a die until we get a six. In this question we think about this experiment in more detail. Assume that the outcomes of successive rolls of the die are independent.
   (a) What is a suitable sample space Ω for the experiment?
   (b) For $i = 1, 2, \ldots$ let A_i be the event that it takes exactly i rolls to get a six for the first time. What is $\mathbb{P}(A_i)$?
   (c) Give a physical interpretation for the event $\bigcup_{i=1}^{\infty} A_i$.
   (d) What is $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i)$?
   (e) For this experiment, give an example of an outcome which is in Ω but not in $\bigcup_{i=1}^{\infty} A_i$.

   Solution:
   (a) We can take Ω to be the set of all infinite sequences of the numbers $1, \ldots, 6$.
   (b) For A_i to occur, we need to have $i - 1$ rolls in which a number other than 6 comes up and then a 6 on the ith roll. The probability of this is $\mathbb{P}(A_i) = (5/6)^{i-1}(1/6)$.
   (c) $\bigcup_{i=1}^{\infty} A_i$ is the event that we eventually roll a six.
   (d) The events A_i are disjoint, so

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} \mathbb{P}(A_i) \\
&= \sum_{i=1}^{\infty} (5/6)^{i-1}(1/6) \\
&= 1.
\end{aligned}$$

   Ask your tutor to explain the last equation if you don't understand it.
   (e) Any infinite sequence of the numbers 1 to 5 is not in $\bigcup_{i=1}^{\infty} A_i$.

2. An art dealer receives a shipment of five old paintings from abroad. On the basis of her past experience, she feels that the probabilities are, respectively, 0.76, 0.09, 0.02, 0.01, 0.02 and 0.10 that $0, 1, 2, 3, 4$ or all 5 of them are forgeries.
   (a) What is the probability of the event $A = \{\text{there is at least one forgery in the shipment}\}$?
   (b) Let $D = \{\text{all five paintings are forgeries}\}$. Calculate $\mathbb{P}(D|A)$.
   (c) Since the cost of authentication is fairly high, the dealer decides to select one of the paintings at random and send it away for authentication. If it turns out that this painting is a forgery, what probability should she then assign to the event D?

Solution:

(a) $\mathbb{P}(A) = 1 - 0.76 = 0.24$.

(b) $\mathbb{P}(D \cap A) = \mathbb{P}(D) = 0.10$. Therefore $\mathbb{P}(D|A) = 0.10/0.24 = 5/12 = 0.4167$.

(c) Let F be the event that the randomly-selected painting is a forgery. We want to calculate

$$\mathbb{P}(D|F) = \frac{\mathbb{P}(D \cap F)}{\mathbb{P}(F)}.$$

If D occurs then the randomly-selected painting will certainly be a forgery and so $\mathbb{P}(D \cap F) = \mathbb{P}(D)$. Now

$$\mathbb{P}(F) = \sum_{i=1}^{5} \mathbb{P}(F|C_i)\mathbb{P}(C_i),$$

where C_i is the event that there are i forgeries. $\mathbb{P}(C_i)$ is given and $\mathbb{P}(F|C_i) = i/5$, so

$$\begin{aligned} \mathbb{P}(F) &= \frac{0.09 + 2 \times 0.02 + 3 \times 0.01 + 4 \times 0.02 + 5 \times 0.10}{5} \\ &= \frac{37}{250}. \end{aligned}$$

Therefore $\mathbb{P}(D|F) = 25/37 = 0.6757$.

3. Consider the random experiment where a number is selected at random from the sample space $\Omega = \{2, 3, 4, \ldots, 25\}$. Let X denote the number of prime factors the number has. For example, any prime number has one prime factor, and since $9 = 3 \times 3$, 9 has two prime factors.

(a) The random variable X can take values 1, 2, 3, and 4. Write down the events $\{X = i\}$ (that is, list the outcomes in the events).

(b) Tabulate the probability mass function of $p_X(i)$ for $i = 1, 2, 3, 4$.

Solution:

(a)

$$\begin{aligned} \{X = 1\} &= \{\omega : X(\omega) = 1\} = \{2, 3, 5, 7, 11, 13, 17, 19, 23\}, \\ \{X = 2\} &= \{\omega : X(\omega) = 2\} = \{4, 6, 9, 10, 14, 15, 21, 22, 25\}, \\ \{X = 3\} &= \{\omega : X(\omega) = 3\} = \{8, 12, 18, 20\}, \\ \{X = 4\} &= \{\omega : X(\omega) = 4\} = \{16, 24\}. \end{aligned}$$

(b)

$$p_X(i) = \begin{cases} 9/24 = 3/8 & \text{if } i = 1, \\ 9/24 = 3/8 & \text{if } i = 2, \\ 4/24 = 1/6 & \text{if } i = 3, \\ 2/24 = 1/12 & \text{if } i = 4. \end{cases}$$

4. Haemophilia is a recessive sex-linked disease. If a woman who is a carrier (that is she has the recessive haemophilia gene but no symptoms) has a child with a man who is normal, then the child will be haemophiliac with probability 0.25.
   (a) Suppose such a couple in which the woman is a carrier intend to have three children. Let the random variable X denote the number of those children who will be haemophiliac. Tabulate the probability mass function $p_X(k)$.
   (b) The situation with haemophilia is actually more complicated. In reality, a male child of the couple described above has a probability 0.50 of being haemophiliac, whereas a female child will not suffer from the disease. For each of $m = 0, 1, 2, 3$, write down the probability mass function $p_X(k)$ of the number of haemophiliac children given the couple has m boys.

Solution:

(a)

$$p_X(k) = \begin{cases} (0.75)^3 & \text{if } k = 0, \\ 3.(0.25)(0.75)^2 & \text{if } k = 1, \\ 3.(0.25)^2(0.75) & \text{if } k = 2, \\ (0.25)^3 & \text{if } k = 3. \end{cases}$$

(b) Let m be the number of boys.
Then, if $m = 0$

$$p_X(0) = 1,$$

if $m = 1$

$$p_X(k) = \begin{cases} 0.5 & \text{if } k = 0, \\ 0.5 & \text{if } k = 1, \end{cases}$$

if $m = 2$

$$p_X(k) = \begin{cases} (0.5)^2 & \text{if } k = 0, \\ 2.(0.5)(0.5) & \text{if } k = 1, \\ (0.5)^2 & \text{if } k = 2, \end{cases}$$

and, if $m = 3$

$$p_X(k) = \begin{cases} (0.5)^3 & \text{if } k = 0, \\ 3.(0.5)(0.5)^2 & \text{if } k = 1, \\ 3.(0.5)^2(0.5) & \text{if } k = 2, \\ (0.5)^3 & \text{if } k = 3. \end{cases}$$

5. Peter rolls a fair die twice and notes the numbers X and Y he gets in the first and second trials, respectively. Define the events

$$A = \{X \leq 4\}, \quad B = \{X = Y\}, \quad C = \{X \text{ is even}\}.$$

   (a) Show that A is independent of B, and that A is independent of C.
   (b) Is B independent of C?
   (c) Are A, B, and C independent?
   (d) Are A and C independent if the die is biased with $\mathbb{P}(X = k) = 1/7$, $k = 1, \ldots, 5$, and $\mathbb{P}(X = 6) = 2/7$?

Solution:

$$\begin{aligned}
\Omega &= \{(1,1),(1,2)\ldots,(6,6)\},\\
A &= \{(1,1),\ldots,(1,6),(2,1),\ldots,(2,6),(3,1),\ldots,(3,6),(4,1)\ldots,(4,6)\},\\
B &= \{(1,1),\ldots,(6,6)\},\\
C &= \{(2,1),\ldots,(2,6),(4,1),\ldots,(4,6),(6,1)\ldots,(6,6)\}.
\end{aligned}$$

Since the dice are fair, it is reasonable to assume that the probability of an event E is given by $\#E/\#\Omega$

(a) $A \cap B = \{(1,1),\ldots,(4,4)\}$. Therefore $\mathbb{P}(A) = 2/3$, $\mathbb{P}(B) = 1/6$ and $\mathbb{P}(A \cap B) = 1/9 = \mathbb{P}(A)\mathbb{P}(B)$. Therefore A and B are independent.

    $A \cap C = \{(2,1),\ldots,(2,6),(4,1)\ldots,(4,6)\}$. $\mathbb{P}(C) = 1/2$ and $\mathbb{P}(A \cap C) = 1/3 = \mathbb{P}(A)\mathbb{P}(C)$. Therefore A and C are independent.

(b) $B \cap C = \{(2,2),(4,4),(6,6)\}$. $\mathbb{P}(B \cap C) = 1/12 = \mathbb{P}(B)\mathbb{P}(C)$. Therefore B and C are independent.

(c) Given the conclusions in (a) and (b), it remains to check whether $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$. Now, $A \cap B \cap C = \{(2,2),(4,4)\}$, so $\mathbb{P}(A \cap B \cap C) = 1/18$ and $\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 1/18$, so the events A, B, and C are mutually independent.

(d) For the biased die, $\mathbb{P}(A) = 4/7$ and $\mathbb{P}(C) = 4/7$ but $\mathbb{P}(A \cap C) = 2/7 \neq \mathbb{P}(A)\mathbb{P}(C)$. Therefore with the biased die, the events A and C are not independent.

MAST20004 Probability

Computer Lab 2

The aim of this lab is to

- use MATLAB to simulate a simple die experiment to help test whether two events A and B are positively or negatively related to each other;
- use MATLAB to investigate the Multiple Choice Exams example from lectures;
- use MATLAB to investigate the HIV testing example from lectures.

MATLAB programming

In this lab you will need to do a very limited amount of programming for yourself — primarily adding a few important lines to the programs that have been provided on the server. To help you with these tasks:

- continue to work in groups of 2 or 3;
- follow the hints built into the comment lines in the programs themselves;
- remember to make use of the help system which is very extensive;
- try to do it yourself but not for too long - ask a tutor for help if you get stuck. At this stage of semester we expect students new to programming to need a lot of assistance.

1. **Exercise A - Dice Throwing**

   Peter (from Tutorial Problem 5) plays a more sophisticated game now. First he rolls his fair die once and notes the score X on it. Then he rolls the die X times and notes all the scores; we denote them by $Y_1, \ldots, Y_X$. Put $Z = Y_1 + \cdots + Y_X$.

   Consider the events $A = \{X \leq 4\}$, $B = \{Z > 10\}$, and $C = \{Y_1 \geq 5\}$.

   [FYI: A possible probabilistic model for this experiment is as follows. Let

   $$\Omega = \bigcup_{k=1}^{6} \Omega_k, \qquad \text{where} \qquad \Omega_k = \big\{(s_1, \ldots, s_k) : s_i = 1, \ldots, 6,\ i \leq k\big\},$$

   so that $\#\Omega_k = 6^k$. Then put $\mathbb{P}(\Omega_k) = 1/6$ $(= \mathbb{P}(X = k))$, $k = 1, \ldots, 6$, and assume that, for each fixed $k \leq 6$, all the outcomes $\omega \in \Omega_k$ are equally likely and hence each of them has probability $1/6^{k+1}$. Think about that at home...]

   (a) Use your intuition (and/or common sense) to guess if A and B are positively related. It looks like calculating some of the relevant probabilities may be a hard task! (Please do not do any calculations. We will use simulation instead.) What about the events B and C?

   (b) Now you will use simulation to test if your guesses from part 2 were correct. Use the m-file **Lab2ExA.m**. Before running **Lab2ExA.m**, you should study the file in MATLAB editor and, following the hints in the program, type in the appropriate code to check for the occurrence of events A, B and $A \cap B$. Save the modified file.

Then run your program by typing `Lab2ExA` in the command window (and then pressing *Enter*, of course). The output of the program will give you the values of the relative frequencies for the above events as observed in `nreps` repetitions of the simulated experiment. Write them down and use the values to test if A and B are positively related (you may wish to add a few commands to the program to automate this task).

(c) Run the program several times with different `nreps` values and check the variability in the estimates obtained from the simulations.

2. **Exercise B - Multiple Choice Exams**

Recall the Multiple Choice Exam example discussed in Lecture Slides 87–89. This example is simulated in the m-file **Lab2ExB.m**. Study this program in the editor and make sure you understand how it works. In particular explain how the program determines if the answer is correct when the student guesses. Think of an alternative method for doing this check, program it and test that it produces the same answers as the original program.

(a) Now run the program for a few selected m and p values and check the empirical results against the theoretical formula on Slide 89.

Solution:

The formula from lectures said that the probability that the student was guessing if they had the correct answer is

$$\frac{1-p}{mp+1-p}.$$

(b) What percentage of questions would a student need to know the correct answer for if they wished to average around 73 percent correct answers overall ? (Solve theoretically and then test with your program).

Solution: The student wants to solve

$$\begin{aligned} 0.73 &= \mathbb{P}(\text{correct}|\text{knew})\mathbb{P}(\text{knew}) + \mathbb{P}(\text{correct}|\text{didn't know})\mathbb{P}(\text{didn't know}) \\ &= 1.p + 1/m.(1-p) \end{aligned}$$

This implies that

$$p = \frac{0.73m - 1}{m-1}.$$

Any value of p greater than or equal to this will do.

3. **Exercise C - Bayes formula - HIV testing example**

(a) Recall the disease testing example from lecture Slides 80–86: Suppose 1 in 1,000 has a certain disease. A test correctly diagnoses 90% of sick people and 95% of healthy people. Given a person tested positive, what is the probability that s/he has the disease? [Use Bayes' formula!]

A simulation for this example can be done using **Lab2ExC.m**. Study this program in the MATLAB editor and make sure you understand how it works. In particular, explain how the program determines (a) if the person has the disease and (b) if the test comes positive when the person is sick/healthy. Think of alternative ways for doing these checks, program them and test that they produce the same results (well, there will be some random variability, of course) as the original program.

(b) Run the program several times and note the results (are they close to our answer in lecture?). Next run it several times for `p=0.0001` (meaning?). Was there any difference in the variability of the results? If yes, any ideas re why?

(c) Now return to the case when `p=0.001` and run the program for a few values other than the 0.9 and 0.95 in the notes (they are called the *sensitivity* and the *specificity* of the test, resp.). Check the simulation results against the formula on Slide 85. If you wanted to increase the probability of the correct diagnosis for sick people, what would you prefer: to increase the sensitivity by 0.02 or to increase the specificity of the test by the same amount?