Statistics – MIOT H6014

Worksheet on Probability

Question 1

A computer program generates a random integer number between 1 and 100, inclusive, where all numbers are equally likely to come up. Identify the experiment and the sample space in this example. Calculate the probability for each of the following events.

Solution:

The experiment is the generation of a random number.

The sample space is the list of numbers 1 to 100.

1. The integer is divisible by 4.

The number 100 is itself divisible by 4, so the list of numbers divisible by 4 is 4, 8, 12, … 100. How many numbers are these? 100/4 = 25.

The probability is then 25/100 = ¼.

Alternatively every 4th number, in a list of length divisible by 4, is counted, giving a probability of ¼.

2. The integer is divisible by 6.

Firstly, beware of the fact that 100 is not divisible by 6, so while the answer will be roughly 1/6, it will not be exactly this.

For the numbers 1 to 6, 1 to 36 or 1 to 96, the answer would be 1/6, since these numbers are divisible by 6.

For 96, there are 16 numbers so far (6 x 16 = 96).

The next number divisible by 6 is 102.

Therefore 16 of the numbers between 1 and 100 are divisible by 6.

Therefore the probability is 16/100 = 0.16.

3. The integer has ‘1’ as one of its digits.

First set of numbers with ‘1’ as a digit are ones with 1 as the second digit.

1, 11, 21, 31, … 91.

The second group are the ones between 10 and 19 with 1 as the first digit

10, 11, 12, … 19.

There is also the number 100.

But we have counted 11 twice, so this is 10 + 10 – 1 + 1 = 20.

Answer is 20/100 = 0.2

Question 2

A small club lottery offers a jackpot for matching the 4 numbers on a ticket to the 4 chosen at random from 30. Identify the experiment and the sample space. Calculate the probability of the following events.

1. The four numbers on a ticket come up in the draw.
2. Three of the four numbers on a ticket come up in a draw. Use your result to explain why the prize for matching 3 rather than 4 numbers is considerably lower than that for matching all 4.
3. Calculate the probability of the club suffering three jackpots in a row.

Solutions

The experiment is selecting 4 numbers at random.

The sample space is the full list of every combination of 4 numbers from the list of 14 names.

1. The four numbers on a ticket come up in the draw.

The number of ways of choosing the four numbers is ${}^{30}C_4$.

The number of ways of choosing the 4 numbers that come up for the jackpot is ${}^{4}C_4 = 1$.

The probability is then the second number divided by the first:

$$\frac{{}^{4}C_4}{{}^{30}C_4} = \frac{1}{\frac{30\times29\times28\times27}{4\times3\times2\times1}} = \frac{1}{5\times29\times7\times27} = \frac{1}{27{,}405}.$$

2. Three of the four numbers on a ticket come up in a draw.

The number of ways of choosing the four numbers is ${}^{30}C_4$.

The number of ways of choosing the 3 numbers from the winning 4 and then 1 from the remaining 26 is ${}^{4}C_3$ x ${}^{26}C_1$.

$$\frac{{}^{4}C_3\times{}^{26}C_1}{{}^{30}C_4} = \frac{4\times26}{27{,}405}.$$

This probability is 104 times higher than the previous, thus the prize has to be 2 orders of magnitude less.

Question 3

A small firm consists of 10 women and 4 men. Three names are chosen at random. Identify the experiment and the sample space in this example. Identify the Universal set. Calculate the probability of the following events.

Solutions

The experiment is selecting 3 names at random.

The sample space is the full list of every combination of 3 names from the list of 14 names.

3. The three names are all men.

Method A, using the laws of probability

Consider this as three choices, each time a man's name.

Let A_1 be the event that the first name chosen is a man's name, A_2 be the second and so on. Then using the multiplication rule for conditional probabilities:

$A = A_1$ and A_2 and A_3, so

$$P[A] = P[A_1]P[A_2]P[A_3] = \frac{4}{14}\frac{3}{13}\frac{2}{12} = \frac{1}{91}.$$

Method B, using a counting processes

The number of ways of choosing the three names is ${}^{14}C_3$.

The number of ways of choosing the 3 names, so that they are all men, is ${}^{4}C_3$.

The probability is then the second number divided by the first:

$$\frac{{}^{4}C_3}{{}^{14}C_3} = \frac{4}{\frac{14\times 13\times 12}{3\times 2\times 1}} = \frac{4}{14\times 13\times 2} = \frac{1}{7\times 13} = \frac{1}{91}.$$

4. The three names are two men and one woman.

Method A, using the laws of probability

Consider firstly getting two men's names and one woman's name as three events.

Let M be the event that the name chosen is a man's name, F be the event a woman's name. Then using the multiplication rule for conditional probabilities:

$$P[M_1]P[M_2]P[F] = \frac{4}{14}\frac{3}{13}\frac{10}{12} = \frac{5}{91}.$$

However, this is just one way in which two men's names and one woman's name can be drawn out, the other probabilities are:

$$P[M_1]P[F]P[M_2] = \frac{4}{14}\frac{10}{13}\frac{3}{12} = \frac{5}{91} \text{ and } P[F]P[M_1]P[M_2] = \frac{10}{14}\frac{4}{13}\frac{3}{12} = \frac{5}{91}.$$

These are all the same, so the probability of the event of 1 man's name and two women's names in any order is

$$P[A] = 3\frac{5}{91} = \frac{15}{91}.$$

Method B, using the counting processes

The number of ways of choosing the three names is ${}^{14}C_3$.

The number of ways of choosing the 3 names, so that they are two men's names and one woman's name is ${}^{4}C_2$ x ${}^{10}C_1$.

$$\frac{{}^{4}C_1 \times {}^{10}C_2}{{}^{14}C_3} = \frac{\frac{4}{2\times 1}\times 10}{\frac{14\times 13\times 12}{3\times 2\times 1}} = \frac{3\times 9}{7\times 13} = \frac{15}{91}.$$

5. The three names are one man and two women.

Method A, using the laws of probability

Consider firstly getting one man's name and two women's names as three events.

Let M be the event that the name chosen is a man's name, F be the event a women's name. Then using the multiplication rule for conditional probabilities:

$$P[M]P[F_1]P[F_2] = \frac{4}{14}\frac{10}{13}\frac{9}{12} = \frac{15}{91}.$$

However, this is just one way in which one man's name and two women's name can be drawn out, the other probabilities are:

$$P[F_1]P[M]P[F_2] = \frac{10}{14}\frac{4}{13}\frac{9}{12} \text{ and } P[M] = P[F_1]P[F_2]P[M] = \frac{10}{14}\frac{9}{13}\frac{4}{12}.$$

These are all the same, so the probability of the event of 1 man's name and two women's names in any order is

$$P[A] = 3\frac{15}{91} = \frac{45}{91}.$$

Method B, using the counting processes

The number of ways of choosing the three names is ${}^{14}C_3$.

The number of ways of choosing the 3 names, so that they are one man and two women is ${}^{4}C_1$ x ${}^{10}C_2$.

$$\frac{{}^{4}C_1 \times {}^{10}C_2}{{}^{14}C_3} = \frac{4 \times \dfrac{10 \times 9}{2 \times 1}}{\dfrac{14 \times 13 \times 12}{3 \times 2 \times 1}} = \frac{5 \times 9}{7 \times 13} = \frac{45}{91}.$$

6. The three names are all women.

Method A, using the laws of probability

Consider this as three choices, each time a woman's name.

Let A_1 be the event that the first name chosen is a woman's name, A_2 be the second and so on. Then using the multiplication rule for conditional probabilities:

$A = A_1$ and A_2 and A_3, so

$$P[A] = P[A_1]P[A_2]P[A_3] = \frac{10}{14}\frac{9}{13}\frac{8}{12} = \frac{30}{91}.$$

Method B, using a counting processes

The number of ways of choosing the three names is ${}^{14}C_3$.

The number of ways of choosing the 3 names, so that they are all women, is ${}^{10}C_3$.

The probability is then the second number divided by the first:

$$\frac{{}^{10}C_3}{{}^{14}C_3} = \frac{\frac{10\times9\times8}{3\times2\times1}}{\frac{14\times13\times12}{3\times2\times1}} = \frac{10\times9\times8}{14\times13\times2} = \frac{30}{91}.$$

7. Add up these probabilities and explain your answer.

The four probabilities add up to 1. They are four distinct events and between them they cover all eventualities, in other words they cover the full sample space. Therefore the sum must be 1.

Question 4

A standard fair dice is rolled three times in succession. Calculate the probability of the following two events:

1. The second number is larger then the first.
2. The third number is larger than the second.

Solutions

To answer the first question, look at the sample space for the roll of two dice, showing the difference:

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

From the event space: Total number of outcomes = 36

Whichever colour represents the second dice, then we see that one triangular sector above or below the line of zeros represents the result that one dice was above the other. Therefore the probability we need is 15/36.

The second probability is the same!

Consider this event:

The third number is larger than the second, given the second number is larger then the first.

Solution

Firstly, the sample space is the 3-dimensional equivalent of the one before, with size $6^3 = 216$.

Our probability is P[Third > Second | Second > First].

Question 5

The following questions require Bayes Rule

1. Dublin's marquee forward Diarmuid Connolly may be given a 12 week ban, meaning he will not play for them until a potential All-Ireland semi-final. He is appealing the ban; the probability his appeal succeeds and he will play is estimated at ¾. Pundits have estimated the following probabilities:
    - *P*[Dublin reach a semi-final with DC] = 0.8
    - *P*[Dublin reach a semi-final without DC] = 0.6

   After a lengthy sojourn on the moon, you return to find that Dublin did in fact reach the semi-final. Assuming Connolly's presence was the only factor affecting their chances of doing this, calculate the probability he played for them.

For this case, let C be the event Connolly plays, let S be the event Dublin reach a semi-final. The information we have is:

- The probability Connolly plays is $P[C] = 0.75$.
- The probability $P[S|C]$ is 0.8.
- The probability a test gives a positive result given a driver is not over the limit , $P[S|C'] = 0.6$.

Now calculate the relevant probability $P[C|S]$ with Bayes rule:

$$P[C\mid S] = \frac{P[S\mid C]P[C]}{P[S\mid C]P[C] + P[S\mid C']P[C']}.$$

Looking at the top line we have all the probabilities in $P[S|C]\ P[C]$, then in the bottom line we just need:

$$P[C'] = 1 - P[C] = 0.25.$$

Therefore

$$P[C \mid S] = \frac{0.8 \times 0.75}{0.8 \times 0.75 + 0.6 \times 0.25} = 0.8.$$

This is 0.8, a high probability but not very different from the original 0.75!

2. A blood-test on a driver determines whether they are over the legal limit of alcohol for driving. Anonymous polling has suggested that 24% of drivers will still drive after drinking enough to put them over this limit. It is proposed to apply random breathalyser-tests to drivers. It is known that:
    - The probability the breathalyser test will give a positive result if the subject is over the limit is 0.95,
    - The probability the breathalyser test will give a positive result if the subject is not over the limit is 0.02.

   Calculate the probability a driver was not over the limit despite giving a positive result.

For this case the data be as shown:

- The probability a driver chosen at random is over the limit is $P[D]$ = 0.24.
- The probability of getting a positive result given a driver is over the limit is $P[T|D]$ is 0.95.
- The probability a test gives a positive result given a driver is not over the limit , $P[T|D']$ = 0.02.

Now calculate the relevant probability $P[D|T]$ with Bayes rule:

$$P[D \mid T] = \frac{P[T \mid D]P[D]}{P[T \mid D]P[D] + P[T \mid D']P[D']}.$$

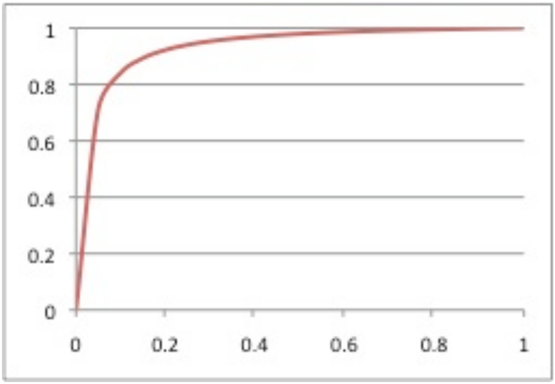
Looking at the top line we have all the probabilities in $P[T|D]\ P[D]$, then in the bottom line we just need:

$$P[D'] = 1 - P[D] = 0.76.$$

Therefore

$$P[D|T] = \frac{0.95 \times 0.24}{0.95 \times 0.24 + 0.02 \times 0.76} = \frac{0.228}{0.228 + 0.0152}.$$

This is 0.9375, a high probability but not very high.



Here is a graph of $P[D|T]$ vs $P[D]$; as can be seen it changes dramatically at the lower values, but soon settles down to a high probability.