# INF 1004 Mathematics 2
# Tutorial #3

**Woon Jun Wei**
**2200624**

January 23, 2023

# Question 1

A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase "free money" is used. Whereas the phrase "free money" is used in only 1% of non-spam emails. A new email has just arrived, which does mention "free money". What is the probability that it is spam?

## My Solution

$$\begin{aligned}
P(\text{spam} \mid \text{free money}) &= \frac{P(\text{free money} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{free money})} \\
&= \frac{0.1 \cdot 0.8}{0.8 \cdot 0.1 + 0.2 \cdot 0.01} \\
&= \frac{40}{41} \\
&\approx 0.9756097561
\end{aligned}$$

# Question 2

A sample of people, who commute regularly from a town in surrey into London, were asked for an estimate of the time taken on their most recent journey. The replies are summarized in the given table.

| Time in minutes | Frequency |
|---|---|
| 35 to 45 | 12 |
| 45 to 55 | 54 |
| 55 to 65 | 68 |
| 65 to 85 | 41 |
| 85 to 105 | 23 |

a) Calculate the estimate of the mean and the standard deviation of these times.

b) What is the median?

c) What is the mode?

## My Solution

## Part 1

Using the midpoints of the intervals, we can calculate the mean and standard deviation as follows:

| Time in minutes | Frequency | $x_i f_i$ | $(x_i f_i)^2$ |
|---|---|---|---|
| 40 | 12 | 480 | 230400 |
| 50 | 54 | 2700 | 729000 |
| 60 | 68 | 4080 | 16646400 |
| 70 | 41 | 2870 | 8236900 |
| 95 | 23 | 2185 | 4774225 |
| total | 198 | 12315 | 30616925 |

$$\begin{aligned}\bar{x} &= \frac{40 \cdot 12 + 50 \cdot 54 + 60 \cdot 68 + 70 \cdot 41 + 95 \cdot 23}{12 + 54 + 68 + 41 + 23} \\ &= \frac{12735}{198} \\ &= 62.19697 \text{ minutes}\end{aligned}$$

$$\begin{aligned}s &= \sqrt{\frac{1}{198-1} \sum x_i^2 f_i - \frac{(\sum x_i f_i)^2}{n}} \\ &= 14.517505 \\ &\approx 14.52\end{aligned}$$

## Part 2

The median is the middle value of the data set. Since there are 198 data points, the median is the average of the 99th and 100th data points.

Since the 99th and 100th data points lie in the 55 to 65 interval,

$$
\begin{aligned}
Q &= L_m + \left(\frac{\frac{n}{2} - cf_{m-1}}{f_m}\right) w_m \\
&= 55 + \frac{\frac{198}{2} - 66}{68} \cdot 10 \\
&= 59.85294 \\
&\approx 59.85 \text{ minutes}
\end{aligned}
$$

## Part 3

The modal class is the class "55 to 65" with a frequency of 68.

# Question 3

Maurice works at home. At 2 pm he decides to take a break to buy a copy of the chronicle newspaper. There are three nearby newsagents: Arif, Bob and Carol. However, by 2 pm they may have sold all their chronicles and so have none available. The independent probabilities that they have a chronicle available at 2 pm are: Arif 0.4, Bob 0.7, Carol 0.25

(a) Find the probability that none of the three newsagents have the chronicles at 2 pm.?

(b) Maurice decides to visit the newsagent in turn until he obtains a chronicle or until he has visited all three. He tosses a coin. If it lands heads he will visit the three newsagents in the order Bob, Arif, Carol. If it lands tail, he will visit them in order Carol, Arif, Bob. Find the probability that he will obtain a chronicle from Arif.

## My Solution

## Part 1

$$P(\text{none available}) = 0.6 \cdot 0.3 \cdot 0.75 = 0.045$$

## Part 2

$$P(\text{Arif}) = 0.5 \cdot 0.3 \cdot 0.4 \cdot 0.75 + 0.5 \cdot 0.75 \cdot 0.4 \cdot 0.0.3 = 0.09$$

First statement: "Head", "bob", "arif", "carol"
Second statement: "Tail", "carol", "arif", "bob"

# Question 4

The following table summarizes the number of late arrivals of all students who attend a certain school on 5 mornings of particular week. (Show your steps for all of the following.)

| Number of late arrivals | Number of Pupils |
|---|---|
| 0 | 275 |
| 1 | 111 |
| 2 | 33 |
| 3 | 12 |
| 4 | 13 |
| 5 | 16 |

(a) Let random variable X=number of late arrivals. Is the random variable discrete or continuous?

(b) Draw the probability distribution for X. Calculate the mean and the standard deviation of the data in the table.

(c) Calculate the expected value.

## My Solution

### Part 1

The random variable is $\boxed{discrete}$.

### Part 2

| Number of late arrivals | Number of Pupils | $P(X)$ | $P(X) \cdot x$ |
|---|---|---|---|
| 0 | 275 | $\frac{265}{460} \approx 0.598$ | 0 |
| 1 | 111 | $\frac{111}{460} \approx 0.241$ | 0.241 |
| 2 | 33 | $\frac{33}{460} \approx 0.072$ | 0.143 |
| 3 | 12 | $\frac{12}{460} \approx 0.026$ | 0.5217 |
| 4 | 13 | $\frac{13}{460} \approx 0.028$ | 0.0565 |
| 5 | 16 | $\frac{16}{460} \approx 0.035$ | 0.0695 |

$$\text{mean} = \frac{275 \cdot 0 + 111 \cdot 1 + 33 \cdot 2 + 12 \cdot 3 + 13 \cdot 4 + 16 \cdot 5}{460} \approx 0.75$$

$$\begin{aligned} s &= \sqrt{\frac{1}{460-1} \sum x_i^2 p_i - \frac{(\sum x_i p_i)^2}{n}} \\ &= 1.23515 \end{aligned}$$

### Part 3

$$E(x) = 0.75$$

# Question 5

A Terrence walks to school every morning. The probability that he arrives late is 0.15 and independent of whether he arrives late on any other morning. For a week, in which he decides to walk to school on five mornings, find:

(a) Probability he arrives late on two or fewer mornings.

(b) Probability he arrives late on three or more mornings.

(c) Find the mean and standard deviation of the number of mornings on which he arrives late.?

## My Solution

Let X be the number of days he arrives late.
n = 5
p(X) = 0.15

### Part 1

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= \binom{5}{2}(o.15)^2(1-0.15)^{5-2} + \binom{5}{1}(0.15)(1-0.15)^{5-1} + \binom{5}{0}(1-0.15)^5 \\ &= 0.973388125 \\ &\approx 0.9734 \end{aligned}$$

### Part 2

$$\begin{aligned} P(X \geq 3) &= P(X=3) + P(X=4) + P(X=5) \\ &= \binom{5}{3}(0.15)^3(1-0.15)^{5-3} + \binom{5}{4}(0.15)^4(1-0.15)^{5-4} + \binom{5}{5}(0.15)^5 \\ &= 0.026611875 \\ &\approx 0.0266 \end{aligned}$$

### Part 3

$$E(x) = 5 \cdot 0.15 = 0.75$$
$$Var(X) = 0.75(1-0.15) = 0.6375$$
$$s = \sqrt{0.6375} \approx 0.799$$

# Question 6

Bob is a student who travels to and from University by bus. He believes that the probability of having to wait more than six minutes to catch a bus is 0.4 and is independent of the time of day and direction of travel.

(a) Assuming Bob beliefs are correct. Calculate values for the mean and standard deviation of the number of times he has to wait more than six minutes to catch a bus in a week when he catches 10 buses.

(b) During a thirteen-week period, the number of times (out of 10) he had to wait more than six minutes to catch a bus were as follows:

$$4, 8, 8, 9, 3, 2, 2, 7, 0, 1, 5, 2, 0$$

(i) Calculate the mean and standard deviation of this data.

(ii) Does the answer in (b)i support Bob's beliefs that the probability of having to wait more than six minutes to catch a bus is constant as 0.4 regardless of any factors?

## My Solution

**Given**:
n = 10
p(X) = 0.4

## Part 1

$$E(X) = 10 \cdot 0.4 = 4$$
$$Var(X) = 10 \cdot 0.4 \cdot 0.6 = 2.4$$
$$s = \sqrt{2.4} \approx 1.549$$

## Part 2

$$E(X) = \frac{4+8+8+9+3+2+2+7+0+1+5+2+0}{10} \approx 3.923076923$$
$$s = \sqrt{\frac{1}{10-1}\sum(x_i - 3.6923)^2} \approx 3.17441$$

## Part 3

The Mean of the data is 3.6923 which is not equal to 0.4. Therefore, Bob's belief is incorrect.

# Question 7

A shop sells two types of bath cubes: relaxing and invigorating. The owner of a hotel buys 50 cubes from this shop. For each cube, there is independently a probability of 0.4 that it will be relaxing.

(a) Find the probability that more than 90% of the 50 cubes will be relaxing cubes.

(b) The 50 cubes included 22 relaxing cubes. Give a reason, whether or not a binomial distribution will provide an appropriate model for the random variable, R, in each of the following cases:

(i) A guest at the hotel randomly selects one of the 50 cubes. If it is not a relaxing cube, he replaces it and again selects a cube at random. He continues this procedure until he obtains a relaxing cube. The random variable R denotes the number of cubes he selects until he obtains a relaxing cube.

(ii) The owner randomly selects 20 of the 50 cubes and places them in a bowl. The random variable R denotes the number of relaxing cubes placed in the bowl; each being evaluated as one trial.

## My Solution

### Part 1

Let X= the number of relaxing cubes.
**Given**:
p(X) = 0.4

90% of 50 = 45

$$\begin{aligned} p(X > 45) &= p(X = 46) + p(X = 47) + p(X = 48) + p(X = 49) + p(X = 50) \\ &= \binom{50}{46}(0.4)^{46}(0.6)^{50-46} + \binom{50}{47}(0.4)^{47}(0.6)^{50-47} \\ &+ \binom{50}{48}(0.4)^{48}(0.6)^{50-48} + \binom{50}{49}(0.4)^{49}(0.6)^{50-49} \\ &+ \binom{50}{50}(0.4)^{50}(0.6)^{50-50} \\ &= \end{aligned}$$

### Part 2

As p(R) is not an independent event, the binomial distribution will not be an appropriate model for the random variable R. The variable n is also not fixed, therefore, the binomial distribution will not be an appropriate model for the random variable R.

### Part 3

As the probability is not constant, the binomial distribution will not be an appropriate model for the random variable R.