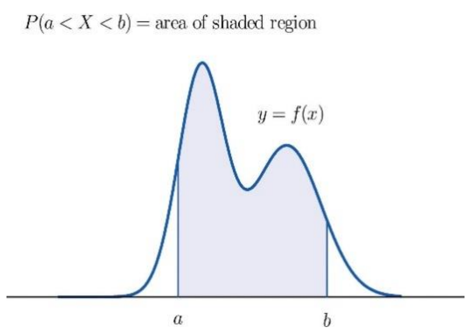
HAWKER COLLEGE

Engage | Inspire | Achieve

Learning Brief

SMM4

Goals



This fortnight we are going to:

- Distinguish between discrete and continuous random variables (CRV)
- Understand probability density functions and cumulative distributions for continuous random variables
- Compute the central tendency and variability (spread) of continuous distributions

Theoretical Components

Make notes on the following chapters:

Maths Quest 12 Mathematical Methods

- 12A - Continuous random variables
- 12B - Using a probability density function to find probabilities of continuous random variables
- 12C - Measures of central tendency and spread
- 12D - Applications to problem solving

Normal Curve:

- https://www.youtube.com/watch?v=McSFVzc8Swk

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. $P(a \leq x \leq b) = \int_a^b f(x)\, dx$

Practical Components

Do the following questions:
Organise your solutions neatly in your exercise book.

Chapter 12 of Maths Quest 12 Mathematical Methods (pdf – Google Classroom)

- 12A: even numbered questions
- 12B: even numbered questions
- 12C: even numbered questions
- 12D: All

Mathspace

Investigation

Questions 4 and 9 from Exercise 12D (Full worked solutions)

Other

Fun fact: The central limit is one of the most important results in all of probability theory. It states that, in many situations, the average of many samples tends towards a normal distribution even if the original variables are themselves not normally distributed. For example, flipping a coin many times will result in a normal distribution for the total number of heads observed (or, equivalently, the total number of tails).

Continuous Random Variable: a variable that can assume any value in one or more intervals.

Continuous Random Variables: variables whose values are measured (i.e. may assume any real value).

Probabilities are calculated from a continuous function known as a <u>probability density function</u> (PDF). Examples include Normal, Uniform, Exponential and Polynomial. The probability is determined by the Area under the curve because of this, we know that following:

1. $f(x) \geq 0$ for all x. *There cannot be any negative probabilities*
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ *Sum of all probabilities equals one*
3. $\Pr(a < x < b) = \int_a^b f(x)\,dx$ *The probability that X (a continuous variable) lies between a and b is determined by the area under the curve between the $x - axis$ and $x = a$ and $x = b$*

Mean:
$\mu = E(x) = \int_{-\infty}^{\infty} x \times f(x)\,dx$ *as per discrete variables where* $\mu = \sum x \times Pr\,(X = x)$

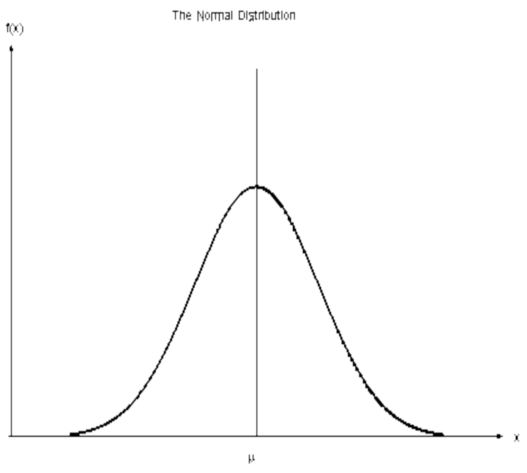
Median:
$\Pr(X \leq m) = \int_a^m f(x)\,dx = \frac{1}{2}$, *it is the value of x that divides the area in half*

Mode:
$\Pr(M) > \Pr\,(x)$, *the maximum value of the function*

Variance:

$$Var(X) = E(X - \mu)^2 = \int_a^b (x - \mu)^2 f(x)\,dx$$

Normal Distribution Properties of the Normal Distribution



The Normal Curve fairly and realistically models many observed frequency distributions such as height and weight, IQ, length of battery life and the distribution of errors in measurements.

$$x \sim N(\mu, \sigma^2)$$

The Normal Curve is characterised by a bell-shaped curve which is symmetrical about the mean.

The equation of the curve is given by the PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ where $x \in R$

The maximum value is obtained when $x = \mu$ i.e. $f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$

Mode, mean, and median are the same.

Many of the frequencies are clustered around the mean.

The graph extends indefinitely in both directions with the $x - axis$ as an asymptote.

Area under the curve is equal to 1 $\left(\int_{\infty}^{\infty} f(x)\, dx = 1\right)$

The probability: $\Pr(a < x < b) = \int_a^b f(x)\, dx$

<u>Verify the following results using CAS</u>

Let $\mu = 0$ and $\sigma = 1$

$\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99.7\%$

$\int_{-3}^{3} f(x)\, dx = 0.997$
$\int_{-2}^{2} f(x)\, dx = 0.954$
$\int_{-1}^{1} f(x)\, dx = 0.683$

Confidence Intervals Associated with a Normal Distribution:

1. Approximately 68.3% of all observations will lie within one standard deviation of the mean
   ($\int_{-1}^{1} f(x)\, dx = 0.683$ i.e. $\Pr(\mu - \sigma \leq X \leq \mu + \sigma) = 68.3\%$)

2. Approximately 95.4% of all observations will lie within two standard deviations of the mean
   ($\int_{-2}^{2} f(x)\, dx = 0.954$ i.e. $\Pr(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 95.4\%$)

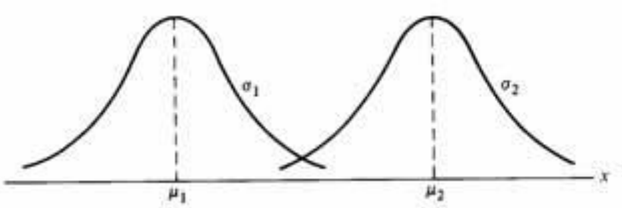
3. Approximately 99.7% of all observations will lie within three standard deviations of the mean
   ($\int_{-3}^{3} f(x)\, dx = 0.997$ i.e. $\Pr(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 99.7\%$)

Normal Distribution Curves will differ in location and degree of spread according to the values of the parameters μ and σ respectively.

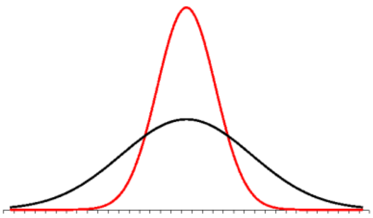
The mean μ and the standard deviation σ are used when dealing with a population and are thus called **population parameters**.

If the values are unknown, then the same mean $\bar{x}$ and sample deviation s are used.

To create a Standardised Normal Distribution, we convert all scores to $z - scores$, where $z = \frac{x-\mu}{\sigma}$, x is the raw score.



Above: Same standard deviation, different means



Above: Same mean, different standard deviations