Certificate in Quantitative Finance

GLOBAL STANDARD IN FINANCIAL ENGINEERING

Certificate in Quantitative Finance
Probability and Statistics

1 Probability

1.1 Preliminaries

- An **experiment** is a repeatable process that gives rise to a number of outcomes.
- An **event** is a collection (or set) of one or more outcomes.
- An **sample space** is the set of all possible outcomes of an experiment, often denoted Ω.

Example

In an experiment a dice is rolled and the number appearing on top is recorded.

Thus

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

If E_1, E_2, E_3 are the events even, odd and prime occurring, then

$$\begin{aligned} E_1 &= \{2, 4, 6\} \\ E_2 &= \{1, 3, 5\} \\ E_3 &= \{2, 3, 5\} \end{aligned}$$

1.1.1 Probability Scale

Probability of an Event E occurring i.e. $P(E)$ is less than or equal to 1 and greater than or equal to 0.

$$0 \leq P(E) \leq 1$$

1.1.2 Probability of an Event

The probability of an event occurring is defined as:

$$P(E) = \frac{\text{The number of ways the event can occur}}{\text{Total number of outcomes}}$$

Example

A fair dice is tossed. The event A is defined as *the number obtained is a multiple of 3.* Determine $P(A)$

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$
$$A = \{3, 6\}$$

$$\therefore P(A) = \frac{2}{6}$$

1.1.3 The Complimentary Event E'

An event E occurs or it does not. If E is the event then E' is the complimentary event, i.e. not E where

$$P(E') = 1 - P(E)$$

1.2 Probability Diagrams

It is useful to represent problems diagrammatically. Three useful diagrams are:

- Sample space or two way table
- Tree diagram
- Venn diagram

Example

Two dice are thrown and their numbers added together. What is the probability of achieving a total of 8?

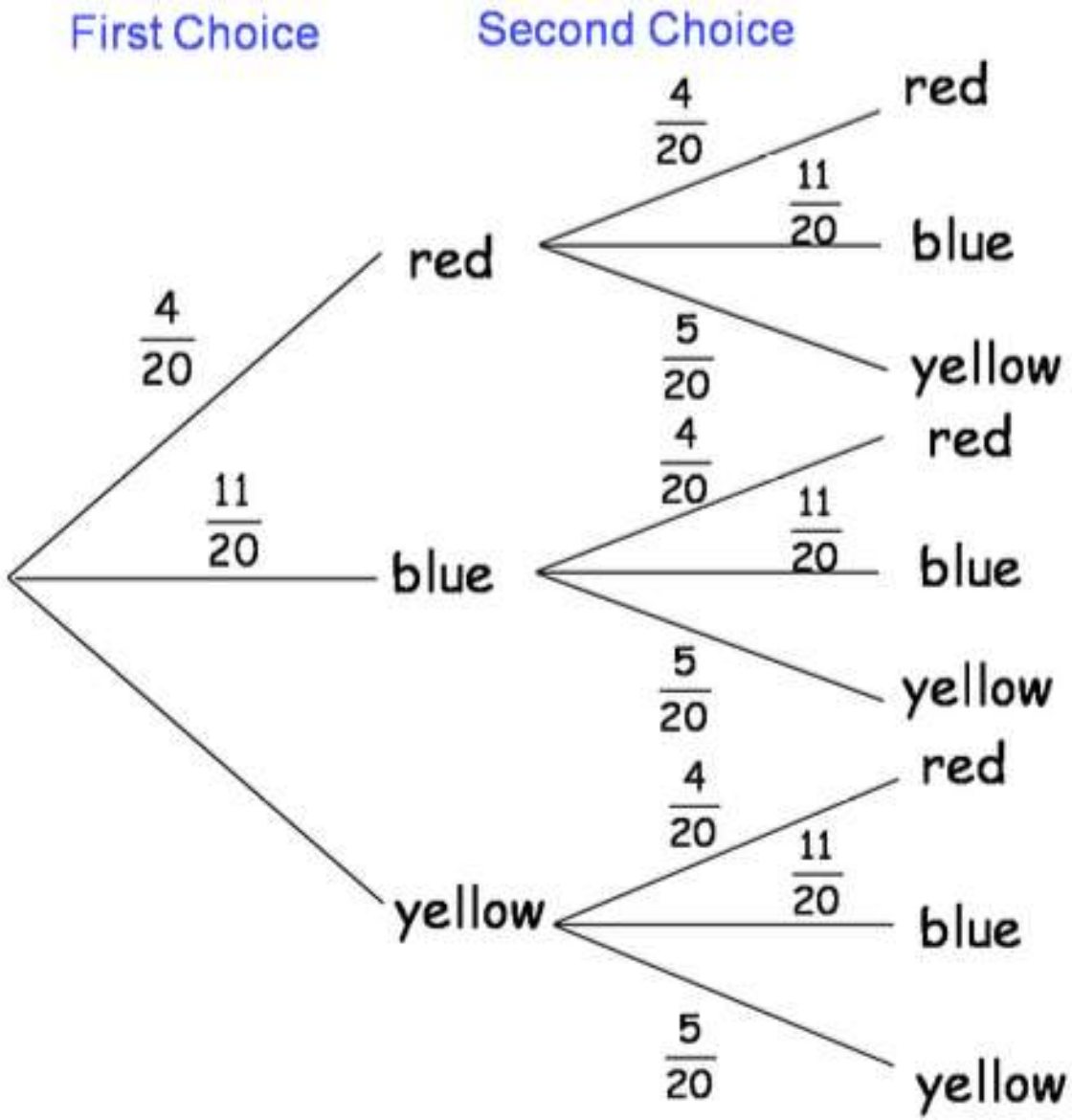
Die 1

Die 2 \ +	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$$P(8) = \frac{5}{36}$$

Example

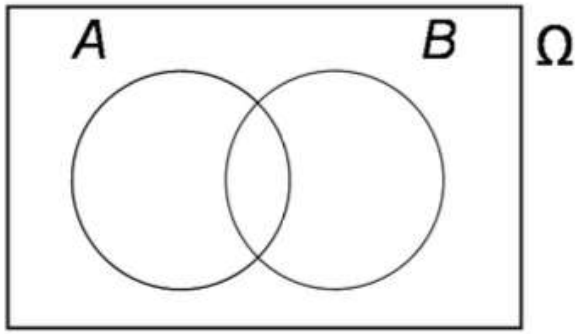
A bag contains 4 red, 5 yellow and 11 blue balls. A ball is pulled out at random, its colour noted and then replaced. What is the probability of picking a red and a blue ball in any order.



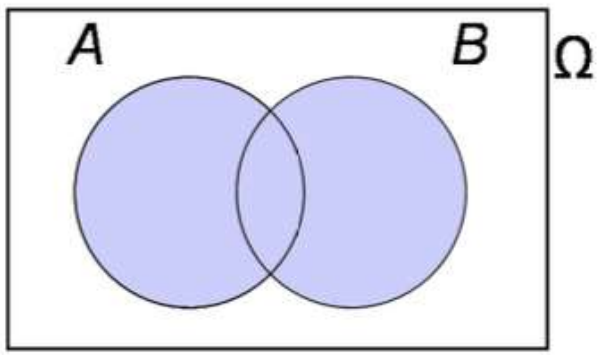
$$\text{P(Red and Blue) or P(Blue and Red)} = \left(\frac{4}{20} \times \frac{11}{20}\right) + \left(\frac{11}{20} \times \frac{4}{20}\right) = \frac{11}{50}$$

Venn Diagram

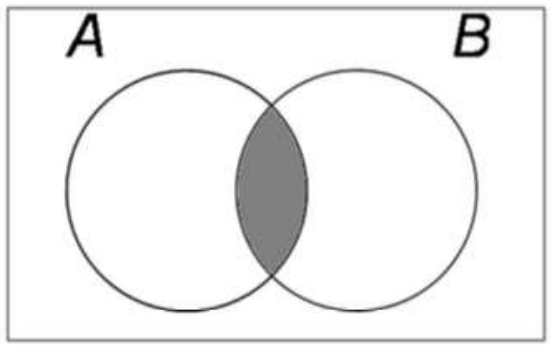
A Venn diagram is a way of representing data sets or events. Consider two events A and B. A Venn diagram to represent these events could be:



- $A \cup B$ "A or B"



- $A \cap B$ "A and B"



Addition Rule:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

or

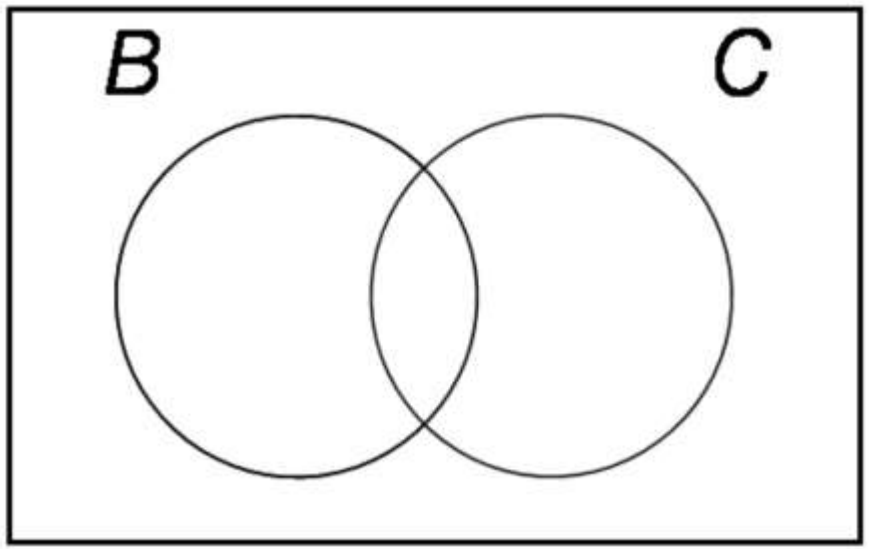
$$P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

Example

In a class of 30 students, 7 are in the choir, 5 are in the school band and 2 students are in the choir and the school band. A student is chosen at random from the class. Find:

a) The probability the student is not in the band

b) The probability the student is not in the choir nor in the band



$$P(\text{not in band}) = \frac{5+20}{30}$$
$$= \frac{25}{30} = \frac{5}{6}$$

$$P(\text{not in either}) = \frac{20}{30} = \frac{2}{3}$$

Example

A vet surveys 100 of her clients, she finds that:

(i) 25 own dogs

(ii) 53 own cats

(iii) 40 own tropical fish

(iv) 15 own dogs and cats

(v) 10 own cats and tropical fish

(vi) 11 own dogs and tropical fish

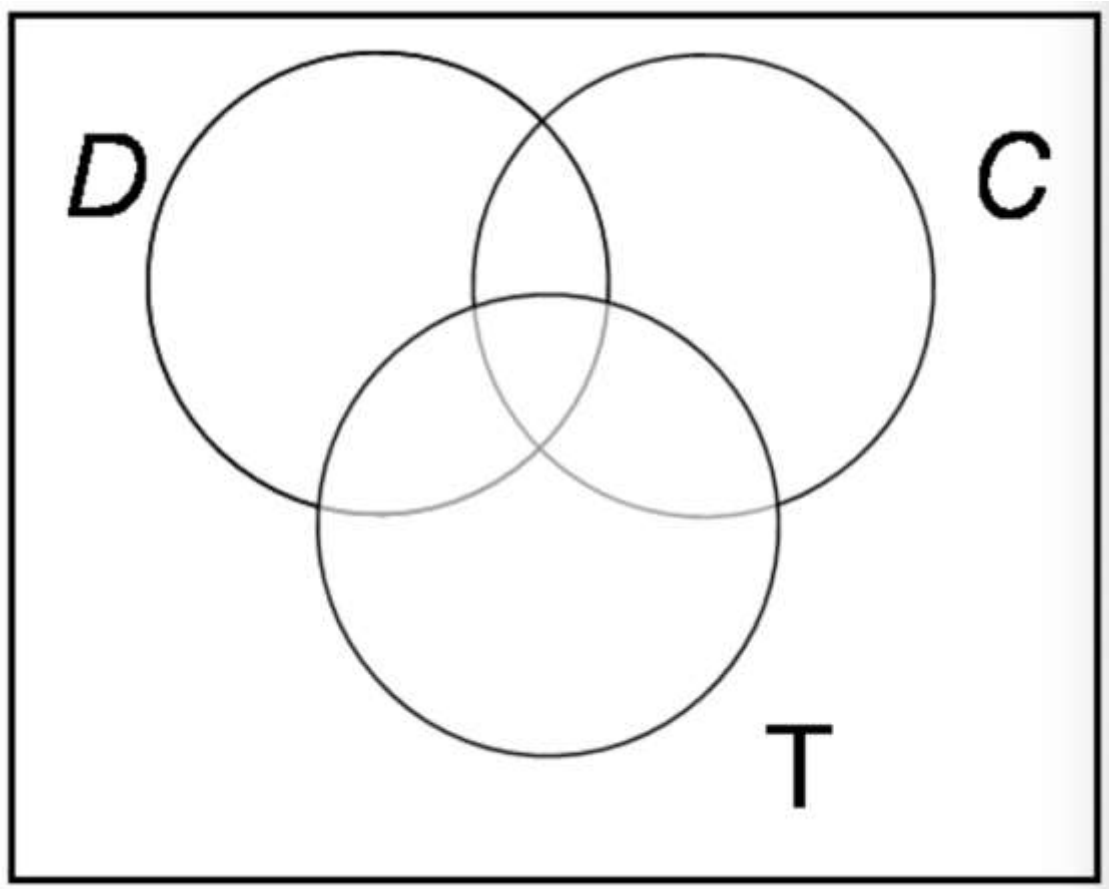
(vii) 7 own dogs, cats and tropical fish

If she picks a client at random, Find:

a) P(Owns dogs only)

b) P(Does not own tropical fish)

c) P(Does not own dogs, cats or tropical fish)



$$P(\text{Dogs only}) = \frac{6}{100}$$

$$P(\text{Does not own tropical fish}) = \frac{6+8+35+11}{100} = \frac{60}{100}$$

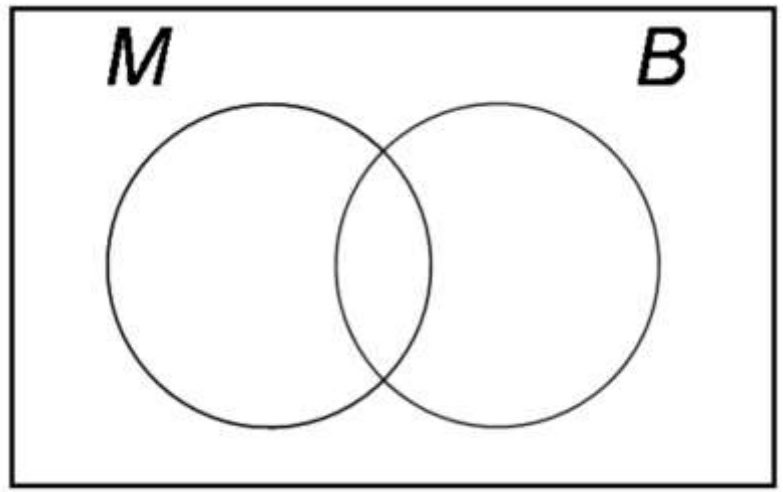
$$P(\text{Does not own dogs, cats or tropical fish}) = \frac{11}{100}$$

1.3 Conditional Probability

The probability of an event B may be different if you know that a dependent event A has already occurred.

Example

Consider a school which has 100 students in its sixth form. 50 students study mathematics, 29 study biology and 13 study both subjects. You walk into a biology class and select a student at random. What is the probability that this student also studies mathematics?



$$P(\text{study maths given they study biology}) = P(M|B) = \frac{13}{29}$$

In general, we have:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

or, **Multiplication Rule:**

$$P(A \cap B) = P(A|B) \times P(B)$$

Example

You are dealt exactly two playing cards from a well shuffled standard 52 card deck. What is the probability that both your cards are Kings ?

Tree Diagram!

$$P(K \cap K) = \frac{4}{52} \times \frac{3}{51} = \frac{1}{221} = \approx 0.5\%$$

or

$$P(K \cap K) = P(\text{2nd is King} \mid \text{first is king}) \times P(\text{first is king}) = \frac{3}{51} \times \frac{4}{52}$$

We know,

$$P(A \cap B) = P(B \cap A)$$

so

$$\begin{aligned} P(A \cap B) &= P(A|B) \times P(B) \\ P(B \cap A) &= P(B|A) \times P(A) \end{aligned}$$

i.e.

$$P(A|B) \times P(B) = P(B|A) \times P(A)$$

or

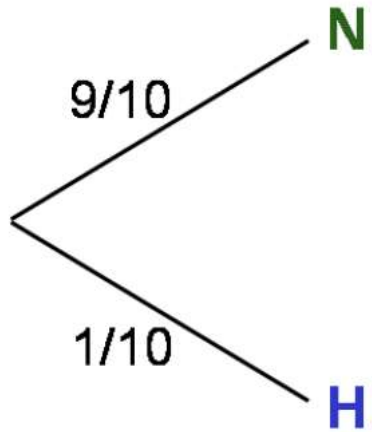
Bayes' Theorem:

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}$$

Example

You have 10 coins in a bag. 9 are fair and 1 is double headed. If you pull out a coin from the bag and do not examine it. Find:

1. Probability of getting 5 heads in a row
2. Probability that if you get 5 heads the you picked the double headed coin



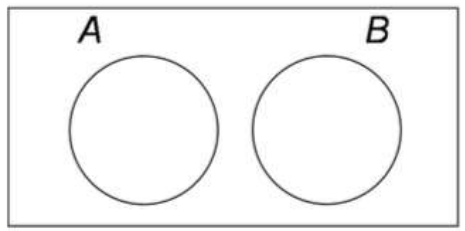
$$
\begin{aligned}
P(5heads) &= P(5heads|N) \times P(N) + P(5heads|H) \times P(H) \\
&= \left(\frac{1}{32} \times \frac{9}{10}\right) + \left(1 \times \frac{1}{10}\right) \\
&= \frac{41}{320} \\
&\approx 13\%
\end{aligned}
$$

$$
\begin{aligned}
P(H|5heads) &= \frac{P(5heads|H) \times P(H)}{P(5heads)} \\
&= \frac{1 \times \frac{1}{10}}{\frac{41}{320}} \\
&= \frac{320}{410} \\
&\approx 78\%
\end{aligned}
$$

1.4 Mutually exclusive and Independent events

When events can not happen at the same time, i.e. no outcomes in common, they are called **mutually exclusive**. If this is the case, then

$$P(A \cap B) = 0$$



and the addition rule becomes

$$P(A \cup B) = P(A) + P(B)$$

Example

Two dice are rolled, event A is 'the sum of the outcomes on both dice is 5' and event B is 'the outcome on each dice is the same'

When one event has no effect on another event, the two events are said to be **independent**, i.e.

$$P(A|B) = P(A)$$

and the multiplication rule becomes

$$P(A \cap B) = P(A) \times P(B)$$

Example

A red dice and a blue dice are rolled, if event A is 'the outcome on the red dice is 3' and event B 'is the outcome on the blue dice is 3' then events A and B are said to be independent.

1.5 Two famous problems

- **Birthday Problem** - What is the probability that at least 2 people share the same birthday

- **Monty Hall Game Show** - Would you swap ?

1.6 Random Variables

1.6.1 Notation

Random Variables X, Y, Z

Observed Variables x, y, z

1.6.2 Definition

Outcomes of experiments are not always numbers, e.g. two heads appearing; picking an ace from a deck of cards. We need some way of assigning real numbers to each random event. Random variables assign numbers to events.

Thus a random variable (RV) X is a function which maps from the sample space Ω to the number line.

Example

let $X =$ the number facing up when a fair dice is rolled,

or let X represent the outcome of a coin toss, where

$$X = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails} \end{cases}$$

1.6.3 Types of Random variable

1. Discrete - Countable outcomes, e.g. roll of a dice, rain or no rain
2. Continuous - Infinite number of outcomes, e.g. exact amount of rain in mm

1.7 Probability Distributions

Depending on whether you are dealing with a discrete or continuous random variable will determine how you define your probability distribution.

1.7.1 Discrete distributions

When dealing with a discrete random variable we define the probability distribution using a probaility mass fucntion or simply a probability function.

Example

The RV X is defined as' the sum of scores shown by two fair six sided dice'. Find the probability distribution of X

A sample space diagram for the experiment is:

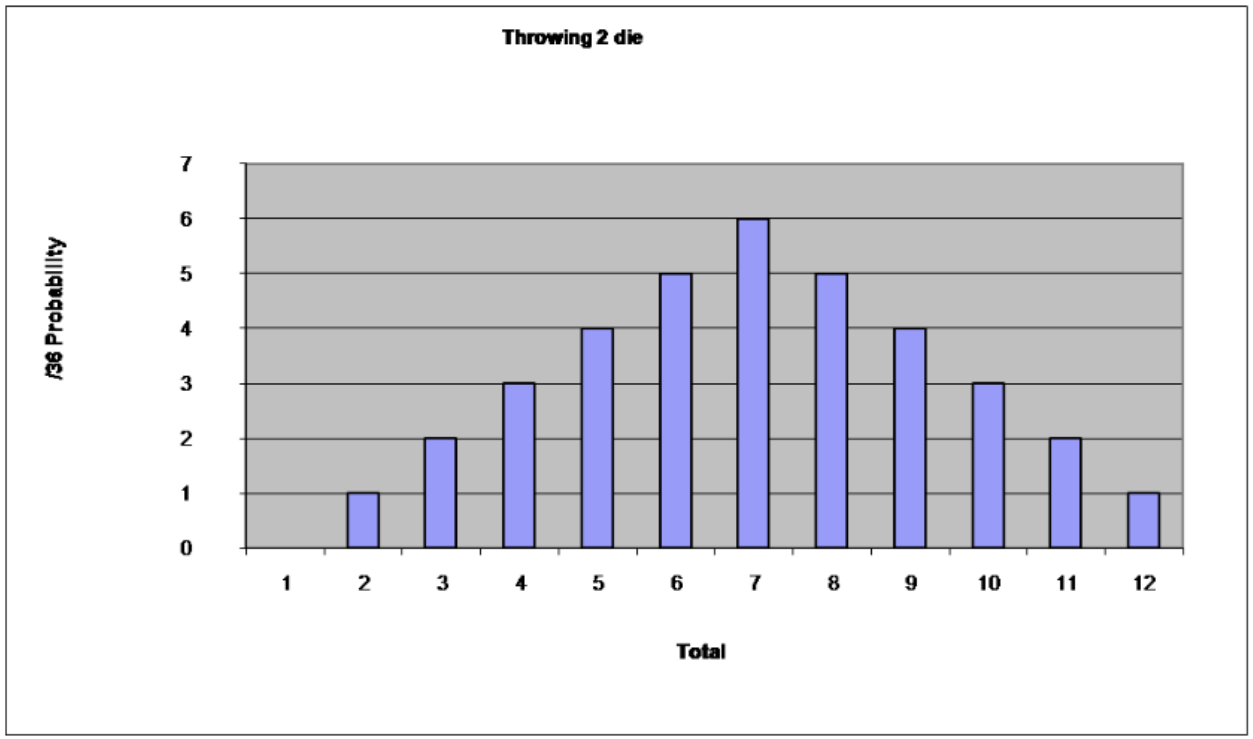
Die 1

Die 2	+	1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

The distribution can be tabulated as:

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

or can be represented on a graph as



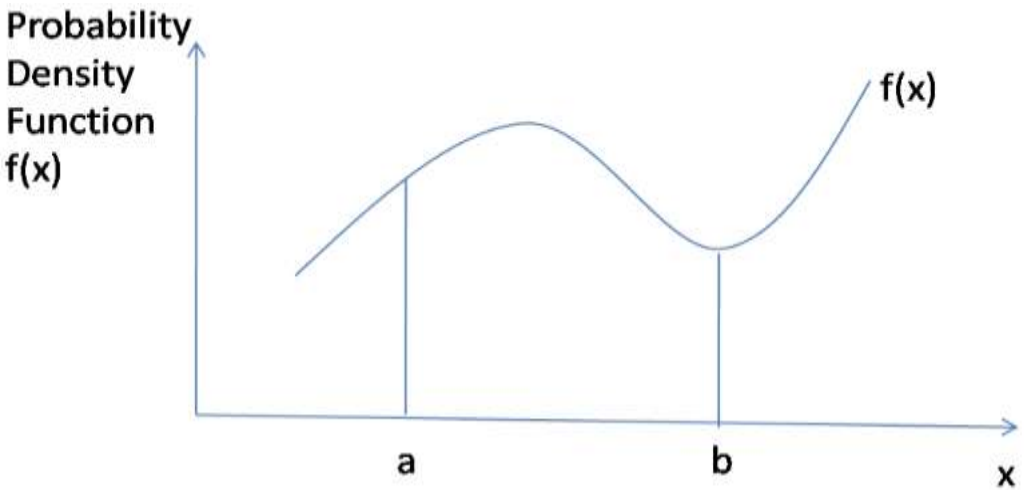
1.7.2 Continuous Distributions

As continuous random variables can take any value, i.e an infinite number of values, we must define our probability distribution differently.

For a continuous RV the probability of getting a specific value is zero, i.e

$$P(X = x) = 0$$

and so just as we go from bar charts to histograms when representing discrete and continuous data, we must use a probability density function (PDF) when describing the probability distribution of a continuous RV.



$$P(a < X < b) = \int_a^b f(x)dx$$

Properties of a PDF:

- $f(x) \geq 0$ since probabilities are always positive
- $\int_{\infty}^{+\infty} f(x)dx = 1$
- $P(a < X < b) = \int_a^b f(x)dx$

Example

The random variable X has the probability density function:

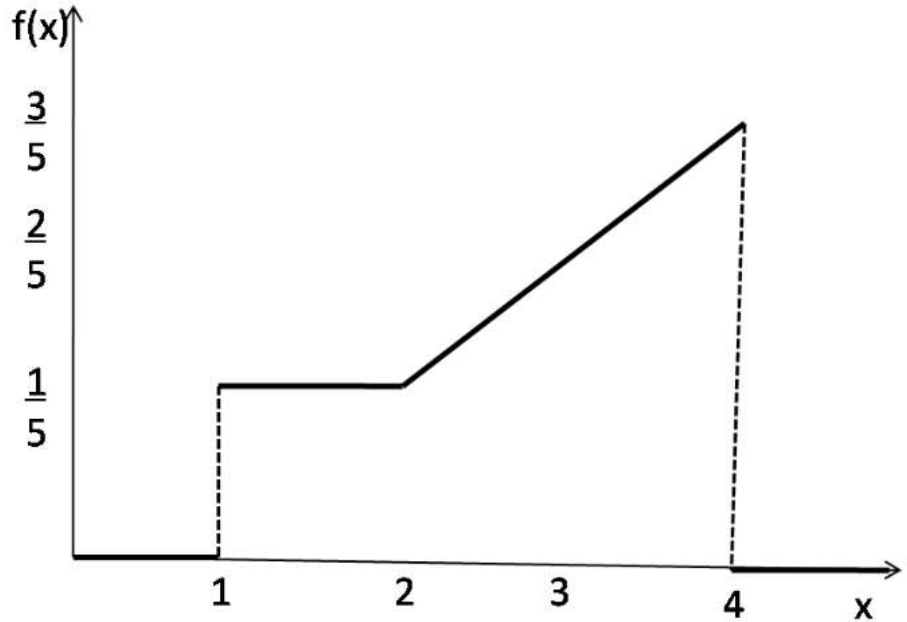
$$f(x) = \begin{cases} k & 1 < x < 2 \\ k(x-1) & 2 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

a) Find k and Sketch the probability distribution

b) Find $P(X \leq 1.5)$

a)

$$\begin{aligned} \int_{\infty}^{+\infty} f(x)dx &= 1 \\ 1 &= \int_1^2 k dx + \int_2^4 k(x-1)dx \\ 1 &= [kx]_1^2 + \left[\frac{kx^2}{2} - kx\right]_2^4 \\ 1 &= 2k - k + [(8k - 4k) - (2k - 2k)] \\ 1 &= 5k \\ \therefore k &= \frac{1}{5} \end{aligned}$$



b)

$$\begin{aligned} P(X \leq 1.5) &= \int_1^{1.5} \frac{1}{5} dx \\ &= \left[\frac{x}{5}\right]_1^{1.5} \\ &= \frac{1}{10} \end{aligned}$$

1.8 Cumulative Distribution Function

The CDF is an alternative function for summarising a probability distribution. It provides a formula for $P(X \leq x)$, i.e.

$$F(x) = P(X \leq x)$$

1.8.1 Discrete Random variables

Example

Consider the probability distribution

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$

$$F(X) = P(X \leq x)$$

Find:

a) $F(2)$ and

b) $F(4.5)$

a)

$$\begin{aligned} F(2) = P(X \leq 2) &= P(X = 1) + P(X = 2) \\ &= \frac{1}{2} + \frac{1}{4} \\ &= \frac{3}{4} \end{aligned}$$

b)

$$\begin{aligned} F(4.5) = P(X \leq 4.5) &= P(X \leq 4) \\ &= \frac{1}{16} + \frac{1}{8} + \frac{1}{4} + \frac{1}{2} \\ &= \frac{15}{16} \end{aligned}$$

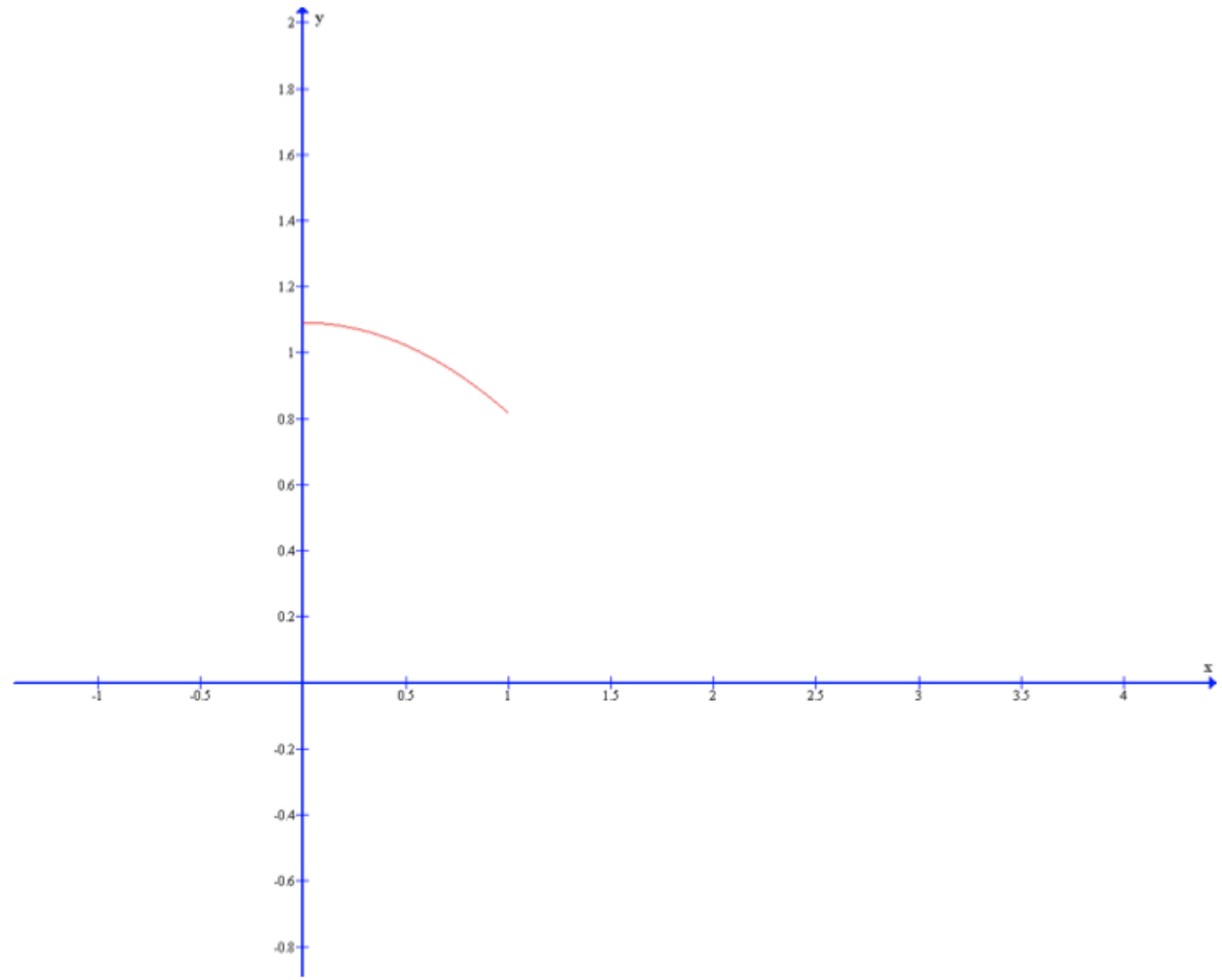
1.8.2 Continuous Random Variable

For continuous random variables

$$F(X) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

or

$$f(x) = \frac{d}{dx}F(x)$$



Example

A PDF is defined as

$$f(x) = \begin{cases} \frac{3}{11}(4 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the CDF

Consider:

From $-\infty$ to 0: $F(x) = 0$

From 1 to ∞: $F(x) = 1$

From 0 to 1 :

$$\begin{aligned} F(x) &= \int_0^x \frac{3}{11}(4 - x^2)dx \\ &= \frac{3}{11}\left[4x - \frac{x^3}{3}\right]_0^x \\ &= \frac{3}{11}\left[4x - \frac{x^3}{3}\right] \end{aligned}$$

i.e.

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{3}{11}\left[4x - \frac{x^3}{3}\right] & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Example

A CDF is defined as:

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{12}\left[x^2 + 2x - 3\right] & 1 \leq x \leq 3 \\ 1 & x > 3 \end{cases}$$

a) Find $P(1.5 \leq x \leq 2.5)$
b) Find $f(x)$

a)

$$\begin{aligned} P(1.5 \leq x \leq 2.5) &= F(2.5) - F(1.5) \\ &= \frac{1}{12}(2.5^2 + 2(2.5) - 3) - \frac{1}{12}(1.5^2 + 2(1.5) - 3) \\ &= 0.5 \end{aligned}$$

b)

$$f(x) = \frac{d}{dx}F(x)$$

$$f(x) = \begin{cases} \frac{1}{6}(x+1) & 1 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

1.9 Expectation and Variance

The expectation or expected value of a random variable X is the mean μ (measure of center), i.e.

$$E(X) = \mu$$

The variance of a random variables X is a measure of dispersion and is labeled σ^2, i.e.

$$Var(X) = \sigma^2$$

1.9.1 Discrete Random variables

For a discrete random variable

$$E(X) = \sum_{allx} xP(X = x)$$

Example

Consider the probability distribution

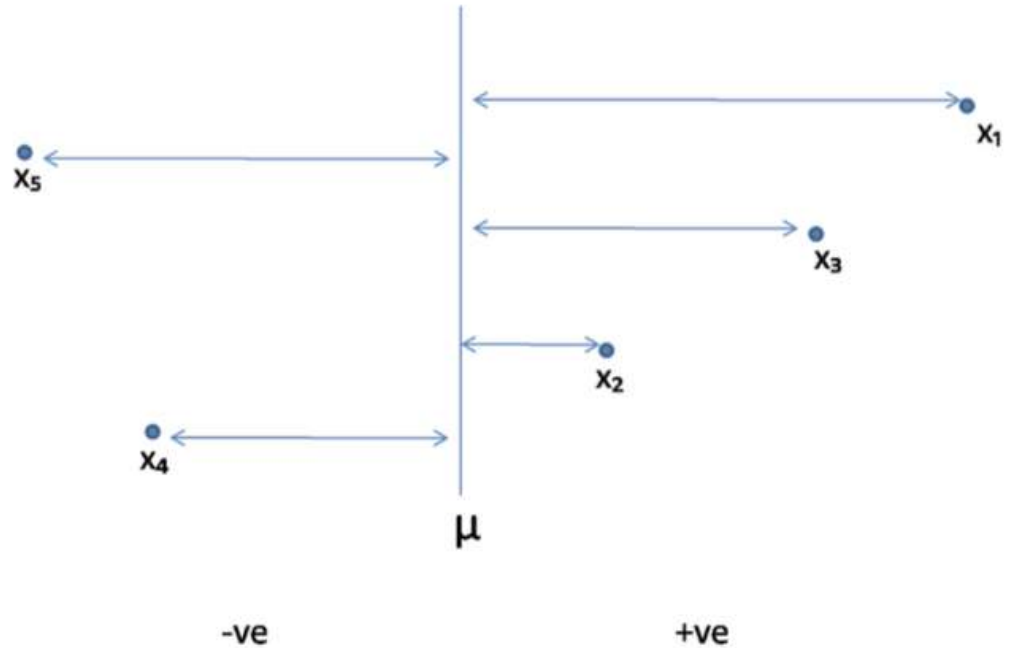
x	1	2	3	4
$P(X = x)$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$

then

$$\begin{aligned} E(X) &= (1 \times \frac{1}{2}) + (2 \times \frac{1}{4}) + (3 \times \frac{1}{8}) + (4 \times \frac{1}{8}) \\ &= \frac{15}{8} \end{aligned}$$

Aside

What is Variance?



$$\text{Variance} = \frac{\sum(x-\mu)^2}{n} = \frac{\sum x^2}{n} - \mu^2$$

$$\text{Standard deviation} = \sqrt{\frac{\sum(x-\mu)^2}{n}} = \sqrt{\frac{\sum x^2}{n} - \mu^2}$$

For a discrete random variable

$$Var(X) = E(X^2) - [E(X)]^2$$

Now, for previous example

$$\begin{aligned} E(X^2) &= 1^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} + 3^2 \times 18 + 4^2 \times \frac{1}{8} \\ E(X) &= \frac{15}{18} \\ \therefore Var(X) &= \frac{71}{64} \\ &= 1.10937... \\ \text{Standard Deviation} &= 1.05(3\text{s.f}) \end{aligned}$$

1.9.2 Continuous Random Variables

For a continuous random variable

$$E(X) = \int_{allx} x f(x) dx$$

and

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= \int_{allx} x^2 f(x) dx - \left[\int_{allx} x f(x) dx \right]^2 \end{aligned}$$

Example

if

$$f(x) = \begin{cases} \frac{3}{32}(4x - x^2) & 0 \leq x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

Find $E(X)$ and $Var(X)$

$$\begin{aligned} E(X) &= \int_0^4 x.\frac{3}{32}(4x - x^2)dx \\ &= \frac{3}{32}\int_0^4 4x - x^2 dx \\ &= \frac{3}{32}\left[\frac{4x^3}{3} - \frac{x^4}{4}\right]_0^4 \\ &= \frac{3}{32}\left[\left(\frac{4(4)^3}{3} - \frac{4^4}{4}\right) - (0)\right] \\ &= 2 \end{aligned}$$

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= \int_0^4 x^2.\frac{3}{32}(4x - x^2)dx - 2^2 \\ &= \frac{3}{32}\left[\frac{4x^4}{4} - \frac{x^5}{5}\right]_0^4 - 4 \\ &= \frac{3}{32}\left[4^4 - \frac{4^5}{5}\right] - 4 \\ &= \frac{4}{5} \end{aligned}$$

1.10 Expectation Algebra

Suppose X and Y are random variables and a,b and c are constants. Then:

- $E(X + a) = E(X) + a$
- $E(aX) = aE(X)$
- $E(X + Y) = E(X) + E(Y)$
- $Var(X + a) = Var(X)$
- $Var(aX) = a^2 Var(X)$
- $Var(b) = 0$

If X and Y are independent, then

- $E(XY) = E(X)E(Y)$
- $Var(X + Y) = Var(X) + Var(Y)$

1.11 Moments

The first moment is $E(X) = \mu$

The n^{th} moment is $E(X^n) = \int_{allx} x^n f(x) dx$

We are often interested in the moments about the mean, i.e. central moments.

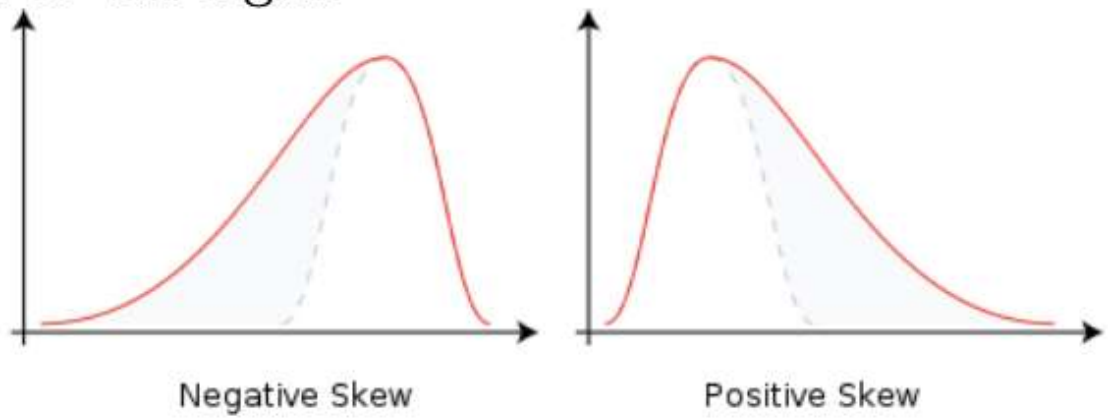
The 2^{nd} central moment about the mean is called the variance $E[(X - \mu)^2] = \sigma^2$

The 3^{rd} central moment is $E[(X - \mu)^3]$

So we can compare with other distributions, we scale with σ^3 and define **Skewness**.

$$\text{Skewness} = \frac{E[(X - \mu)^3]}{\sigma^3}$$

This is a measure of asymmetry of a distribution. A distribution which is symmetric has skew of 0. Negative values of the skewness indicate data that are skewed to the left, where positive values of skewness indicate data skewed to the right.



The 4^{th} normalised central moment is called **Kurtosis** and is defined as

$$\text{Kurtosis} = \frac{E[(X-\mu)^4]}{\sigma^4}$$

A normal random variable has Kurtosis of 3 irrespective of its mean and standard deviation. Often when comparing a distribution to the normal distribution, the measure of **excess Kurtosis** is used, i.e. Kurtosis of distribution -3.

Intiution to help understand Kurtosis

Consider the following data and the effect on the Kurtosis of a continuous distribution.

$x_i < \mu \pm \sigma$:

The contribution to the Kurtosis from all data points within 1 standard deviation from the mean is low since

$$\frac{(x_i - \mu)^4}{\sigma^4} < 1$$

e.g consider

$$x_1 = \mu + \frac{1}{2}\sigma$$

then

$$\frac{(x_1 - \mu)^4}{\sigma^4} = \frac{\left(\frac{1}{2}\right)^4 \sigma^4}{\sigma^4} = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$$

$x_i > \mu \pm \sigma$:

The contribution to the Kurtosis from data points greater than 1 standard deviation from the mean will be greater the further they are from the mean.

$$\frac{(x_i - \mu)^4}{\sigma^4} > 1$$

e.g consider

$$x_1 = \mu + 3\sigma$$

then

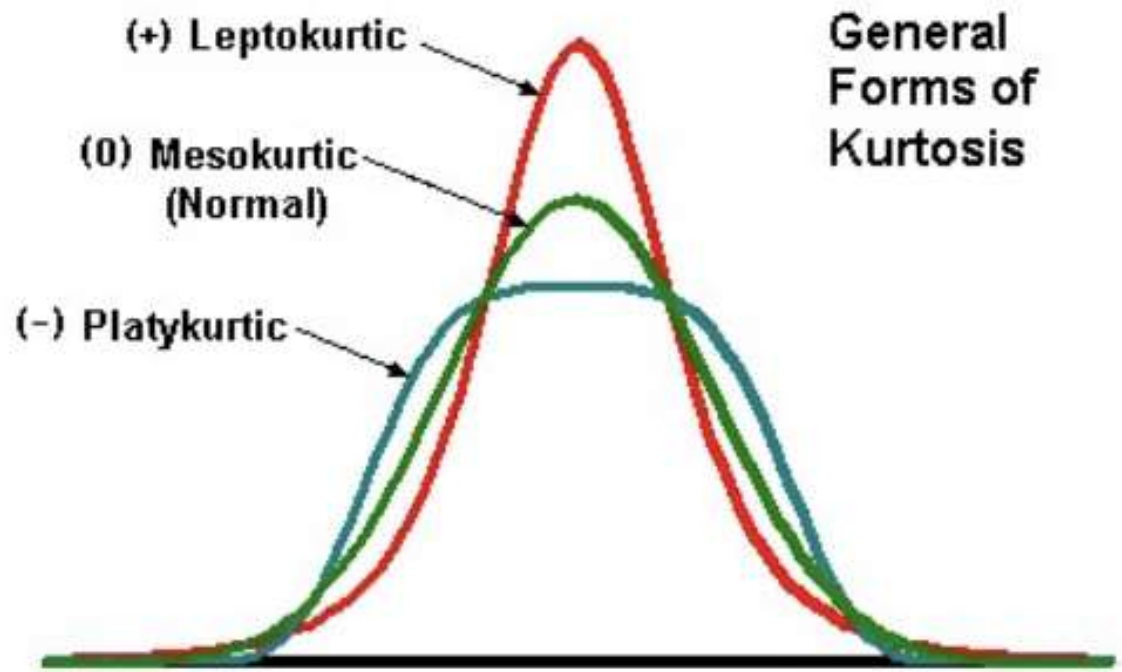
$$\frac{(x_1 - \mu)^4}{\sigma^4} = \frac{(3\sigma)^4}{\sigma^4} = 81$$

This shows that a data point 3 standard deviations from the mean would have a much greater effect on the Kurtosis than data close to the mean value. Therefore, if the distribution has more data in the tails, i.e. fat tails then it will have a larger Kurtosis.

Thus Kurtosis is often seen as a measure of how 'fat' the tails of a distribution are.

If a random variable has Kurtosis greater than 3 is is called *Leptokurtic*, if is has Kurtosis less than 3 it is called *platykurtic*

Leptokurtic is associated with PDF's that are simultaneously peaked and have fat tails.



1.12 Covariance

The covariance is useful in studying the statistical dependence between two random variables. If X and Y are random variables, then theor covariance is defined as:

$$\begin{aligned} Cov(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Intuition

Imagine we have a single sample of X and Y, so that:

$X = 1,\ E(X) = 0$

$Y = 3,\ E(Y) = 4$

Now

$$X - E(X) = 1$$

and

$$Y - E(Y) = -1$$

i.e.

$$Cov(X, Y) = -1$$

So in this sample when X was above its expected value and Y was below its expected value we get a negative number.

Now if we do this for every X and Y and average this product, we should find the Covariance is negative. What about if:

$X = 4,\ E(X) = 0$

$Y = 7,\ E(Y) = 4$

Now

$$X - E(X) = 4$$

and

$$Y - E(Y) = 3$$

i.e.

$$Cov(X, Y) = 12$$

i.e positive

We can now define an important dimensionless quantity (used in finance) called the correlation coefficient and denoted $\rho_{XY}(X, Y)$ where

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}; \qquad -1 \leq \rho_{XY} \leq 1$$

If $\rho_{XY} = -1 \implies$ perfect negative correlation

If $\rho_{XY} = 1 \implies$ perfect positive correlation

If $\rho_{XY} = 0 \implies$ uncorrelated

1.13 Important Distributions

1.13.1 Binomial Distribution

The Binomial distribution is a discrete distribution and can be used if the following are true.

- A fixed number of trials, n
- Trials are independent
- Probability of success is a constant p

We say $X \sim B(n, p)$ and

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Example

If $X \sim B(10, 0.23)$, find

a) $P(X = 3)$

b) $P(X < 4)$

a)

$$\begin{aligned} P(X = 3) &= \binom{10}{3}(0.23)^3(1-0.23)^7 \\ &= 0.2343 \end{aligned}$$

b)

$$\begin{aligned} P(X < 4) &= P(X \leq 3) \\ &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \binom{10}{0}(0.23)^0(0.77)^{10} + \binom{10}{1}(0.23)^1(0.77)^9 \\ &+ \binom{10}{2}(0.23)^2(0.77)^8 + \binom{10}{3}(0.23)^3(0.77)^7 \\ &= 0.821(3 \text{ d.p}) \end{aligned}$$

Example

Paul rolls a standard fair cubical die 8 times. What is the probability that he gets 2 sixes.

Let X be the random variable equal to the number of 6's obtained, i.e $X \sim B(8, \frac{1}{6})$

$$P(X = 2) = \binom{8}{2} \left(\frac{1}{6}\right)^2 \left(\frac{1}{6}\right)^6 = 0.2604(4 \text{ d.p})$$

It can be shown that for a binomial distribution where $X \sim B(n, p)$

$$E(X) = np$$

and

$$Var(X) = np(1 - p)$$

1.13.2 Poisson Distribution

The Poisson distribution is a discrete distribution where the random variable X represents the number of events that occur 'at random' in any interval. If X is to have a Poisson distribution then events must occur

- Singly, i.e. no chance of two events occurring at the same time
- Independently of each other
- Probability of an event occurring at all points in time is the same

We say $X \sim Po(\lambda)$.

The Poisson distribution has probability function:

$$P(X = r) = \frac{e^{-\lambda}\lambda^r}{r!} \qquad r = 0, 1, 2...$$

It can be shown that:

$$E(X) = \lambda$$

$$Var(X) = \lambda$$

Example

Between 6pm and 7pm, directory enquiries receives calls at the rate of 2 per minute. Find the probability that:

(i) 4 calls arrive in a randomly chosen minute

(ii) 6 calls arrive in a randomly chosen two minute period

(i) Let X be the number of call in 1 minute, so

$$\lambda = 2, \qquad i.e. \quad E(X) = 2$$

and

$$X \sim Po(2) = \frac{e^{-2}2^r}{r!}$$

$$P(X = 4) = \frac{e^{-2}2^4}{4!} = 0.090(3 \text{ d.p})$$

(ii) Let Y be the number of calls in 2 minutes, so

$$\lambda = 4, \qquad i.e. \quad E(Y) = 4$$

and

$$P(Y = 6) = \frac{e^{-4}4^6}{6!} = 0.104(3 \text{ d.p})$$

1.13.3 Normal Distribution

The Normal distribution is a continuous distribution. This is the **most** important distribution. If X is a random variable that follows the normal distribution we say:

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

where

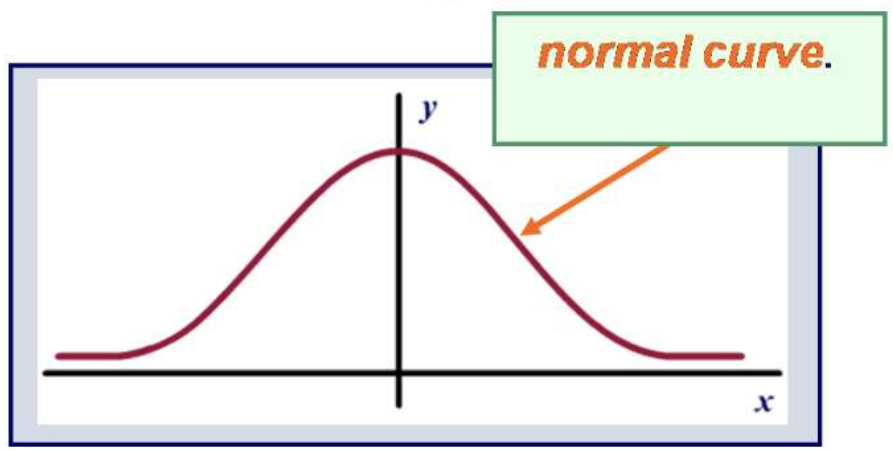
$$E(X) = \mu$$
$$Var(X) = \sigma^2$$

and the PDF is described as

$$\text{PDF} = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

i.e.

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(s-\mu)^2}{2\sigma^2}} ds$$



The Normal distribution is symmetric and area under the graph equals 1, i.e.

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

To find the probabilities we must integrate under $f(x)$, this is not easy to do and requires numerical methods. In order to avoid this numerical calculation we define a standard normal distribution, for which values have already been documented.

The Standard Normal distribution is just a transformation of the Normal distribution.

1.13.4 Standard Normal distribution

We define a standard normal random variable by Z, where $Z \sim N(0,1)$, i.e.

$$E(Z) = 0$$

$$Var(Z) = 1$$

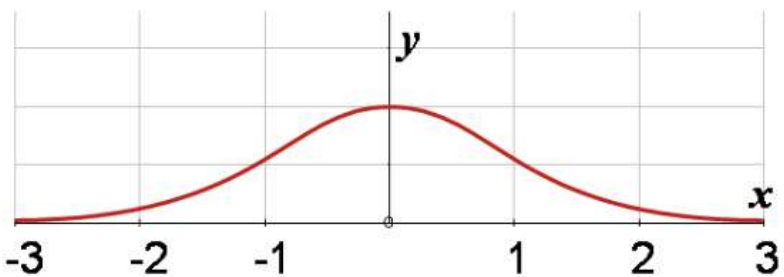
thus the PDF is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

and

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-s^2}{2}} ds$$

N[0, 1]



To transform a Normal distribution into a Standard Normal distribution, we use:

$$Z = \frac{X - \mu}{\sigma}$$

Example

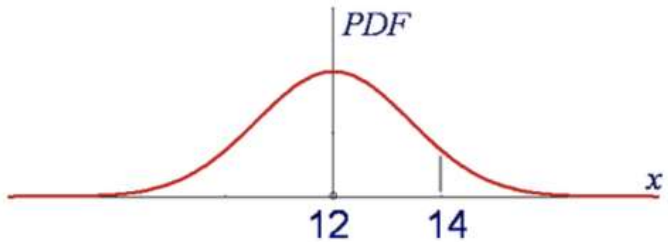
Given $X \sim N(12, 16)$ find:

a) $P(X < 14)$

b) $P(X > 11)$

c) $P(13 < X < 15)$

a)



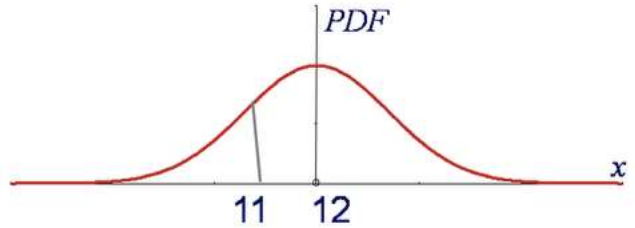
$$Z = \frac{X - \mu}{\sigma} = \frac{14 - 12}{4} = 0.5$$

Therefore we want

$$\begin{aligned} P(Z \leq 0.5) &= \Phi(0.5) \\ &= 0.6915 \end{aligned}$$

(from tables)

b)



$$Z = \frac{11 - 12}{4} = -0.25$$

Therefore we want

$$P(Z > -0.25)$$

but this is not in the tables. From symmetry this is the same as

$$P(Z < 0.25)$$

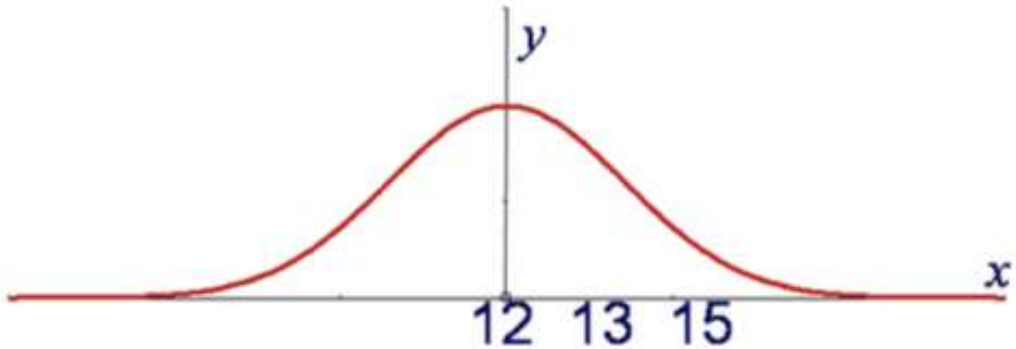
i.e.

$$\Phi(0.25)$$

thus

$$\begin{aligned} P(Z > -0.25) &= \Phi(0.25) \\ &= 0.5987 \end{aligned}$$

c)



$$Z_1 = \frac{13-12}{4} = 0.25$$

$$Z_2 = \frac{15-12}{4} = 0.75$$

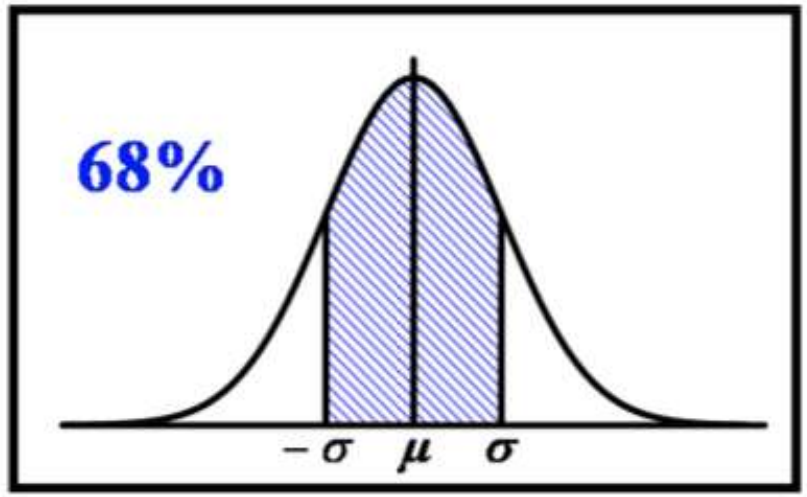
Therefore

$$\begin{aligned} P(0.25 < Z < 0.75) &= \Phi(0.75) - \Phi(0.25) \\ &= 0.7734 - 0.5987 \\ &= 0.1747 \end{aligned}$$

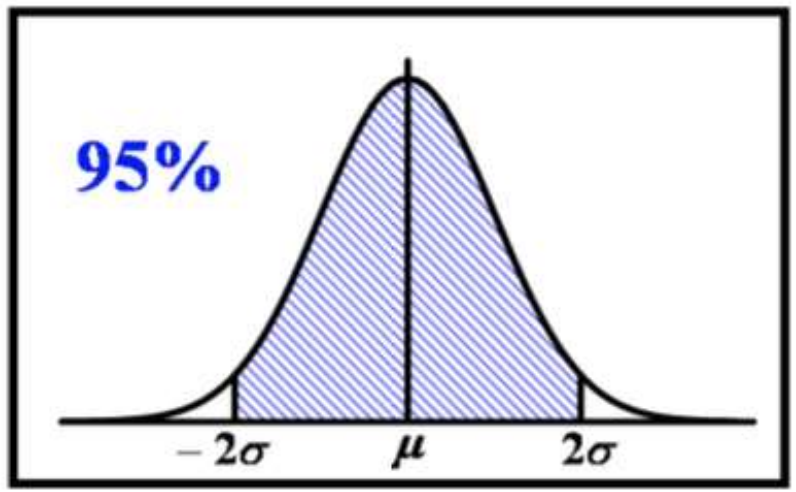
1.13.5 Common regions

The percentages of the Normal Distribution lying within the given number of standard deviations either side of the mean are approximately:

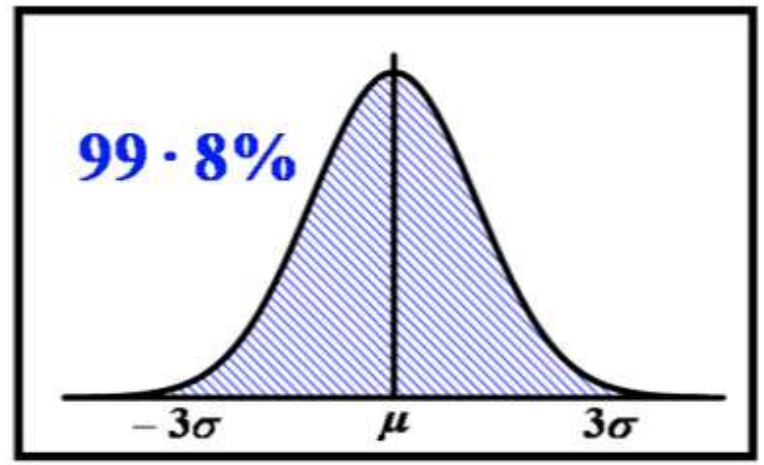
One Standard Deviation:



Two Standard Deviations:



Three Standard Deviations:



1.14 Central Limit Theorem

The Central Limit Theorem states:

Suppose $X_1, X_2, ..., X_n$ are n **independent** random variables, each having the **same** distribution. Then as n increases, the distributions of

$$X_1 + X_2 + ... + X_n$$

and of

$$\frac{X_1 + X_2 + ... + X_n}{n}$$

come increasingly to resemble normal distributions.

Why is this important ?

The importance lies in the fact:

(i) The common distribution of X is not stated - it can be any distribution

(ii) The resemblance to a normal distribution holds for remarkably small n

(iii) Total and means are quantities of interest

If X is a random variable with mean μ and standard devaition σ fom an unknown distribution, the central limit theorem states that the distribution of the sample means is **Normal**.

But what are it's mean and variance ?

Let us consider the sample mean as another random variable, which we will denote $\bar{X}$. We know that

$$\bar{X} = \frac{X_1 + X_2 + ...X_n}{n} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n$$

We want $E(\bar{X})$ and $Var(\bar{X})$

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n\right) \\ &= \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + ... + \frac{1}{n}E(X_n) \\ &= \frac{1}{n}\mu + \frac{1}{n}\mu + ... + \frac{1}{n}\mu \\ &= n\left(\frac{1}{n}\mu\right) \\ &= \mu \end{aligned}$$

i.e. the expectation of the sample mean is the population mean !

$$
\begin{aligned}
Var(\bar{X}) &= Var\left(\frac{1}{n}X_1 + \frac{1}{n}X_2 + ... + \frac{1}{n}X_n\right) \\
&= Var\left(\frac{1}{n}X_1\right) + Var\left(\frac{1}{n}X_2\right) + ... + Var\left(\frac{1}{n}X_n\right) \\
&= \left(\frac{1}{n}\right)^2 Var(X_1) + \left(\frac{1}{n}\right)^2 Var(X_2) + ... + \left(\frac{1}{n}\right)^2 Var(X_n) \\
&= \left(\frac{1}{n}\right)^2 \sigma^2 + \left(\frac{1}{n}\right)^2 \sigma^2 + ... + \left(\frac{1}{n}\right)^2 \sigma^2 \\
&= n\left(\frac{1}{n}\right)^2 \sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

Thus CLT tells us that where n is a sufficiently large

number of samples.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Standardising, we get the equivalent result that

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

This analysis could be repeated for the sum $S_n = X_1 + X_2 + ... + X_n$ and we would find that

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0, 1)$$

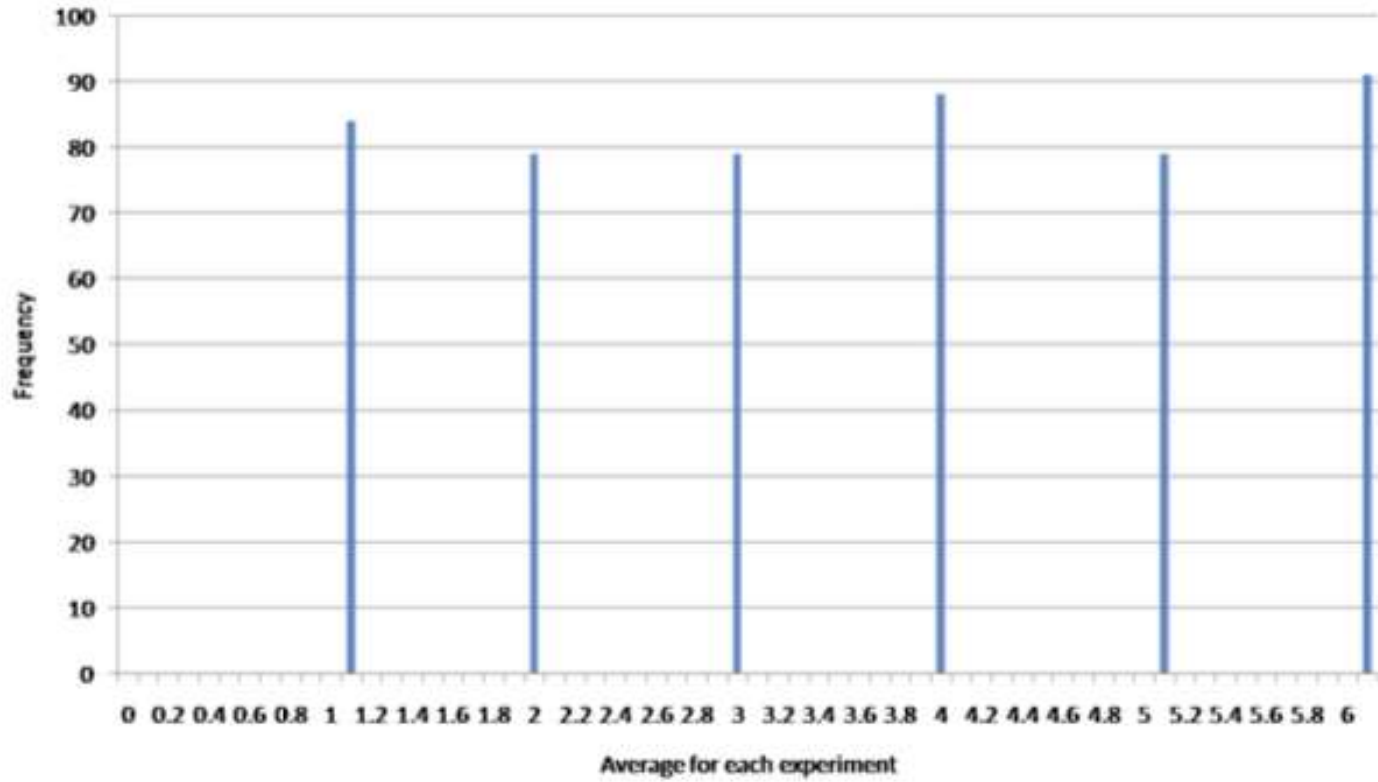
Example

Consider a 6 sided fair dice. We know that $E(X) = 3.5$ and $Var(X) = \frac{35}{12}$.

Let us now consider an experiment. The experiment consists of rolling the dice n times and calculating the average for the experiment. We will run 500 such experiments and record the results in a Histogram.

n=1

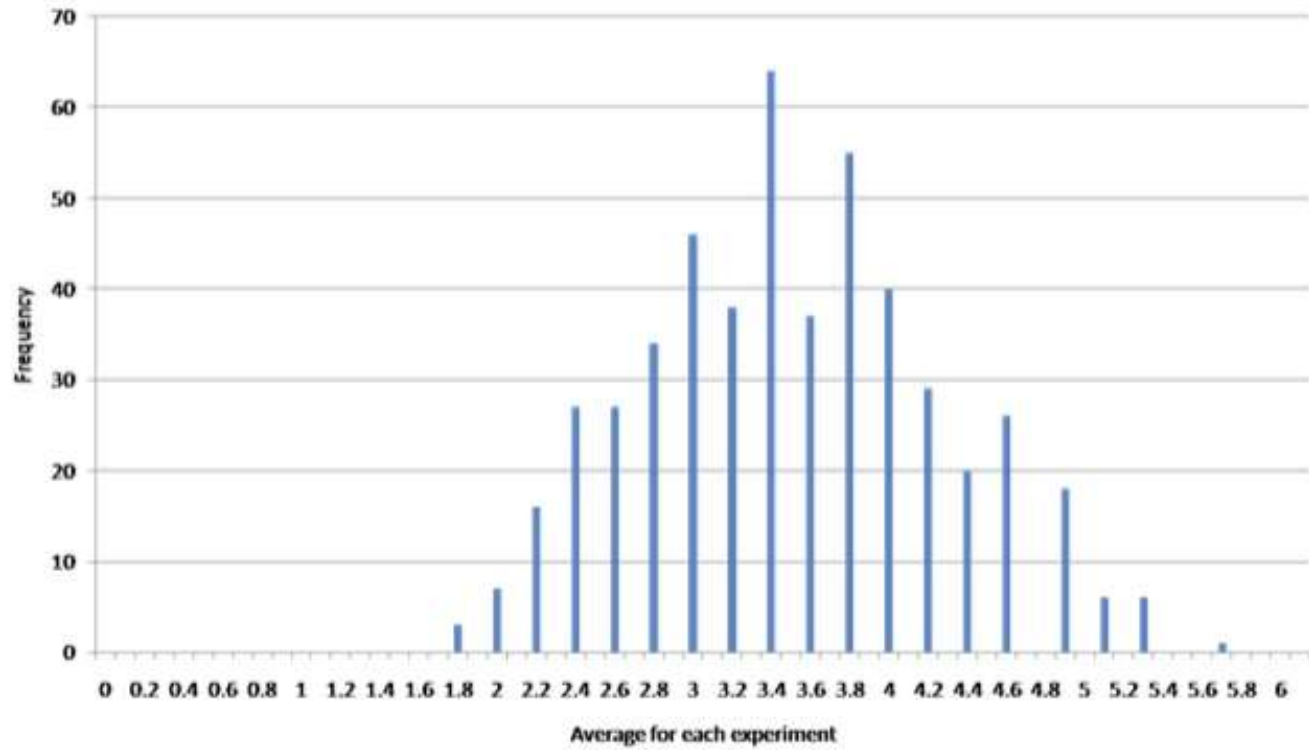
In each experiment the dice is rolled once only, this experiment is then repeated 500 times. The graph below shows the resulting frequency chart.



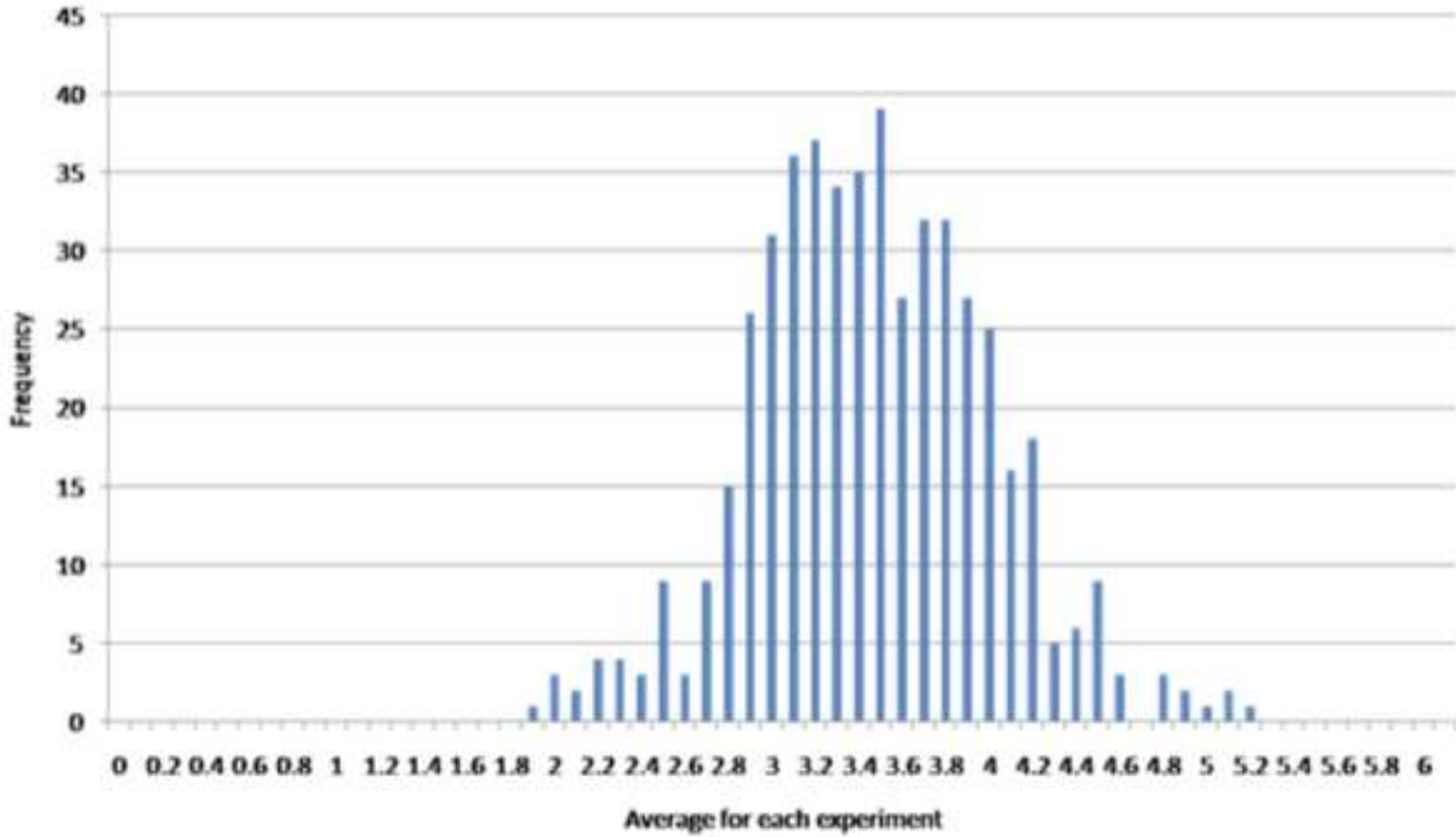
This clearly resembles a uniform distribution (as expected).

Let us now increase the number of rolls, but continue to carry out 500 experiments each time and see what happens to the distribution of $\bar{X}$

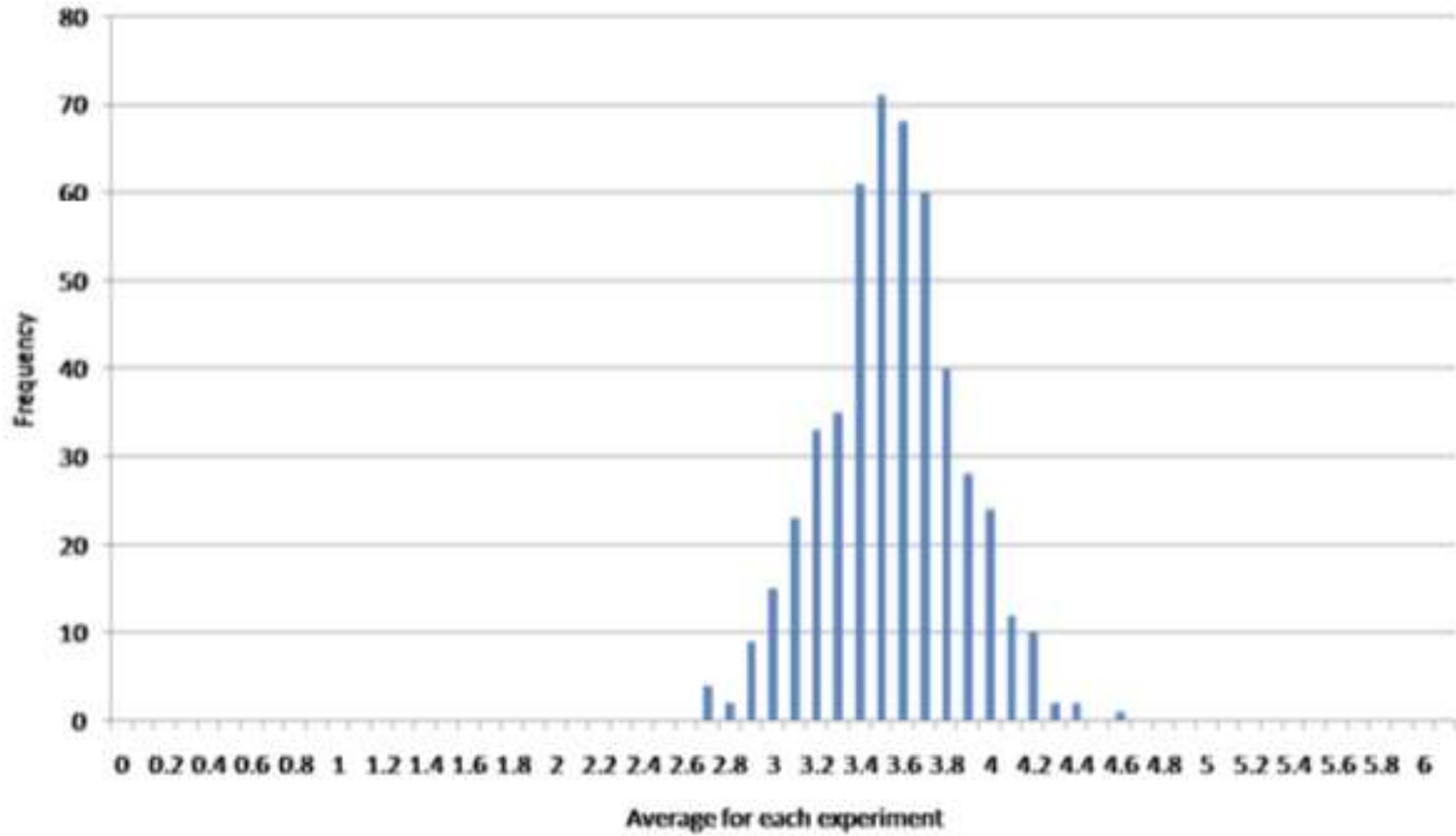
n=5



n=10


45
40
35
30
25
20
15
10
5
0
Frequency
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3 3.2 3.4 3.6 3.8 4 4.2 4.4 4.6 4.8 5 5.2 5.4 5.6 5.8 6
Average for each experiment

n=30


80
70
60
50
40
30
20
10
0
Frequency
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3 3.2 3.4 3.6 3.8 4 4.2 4.4 4.6 4.8 5 5.2 5.4 5.6 5.8 6
Average for each experiment

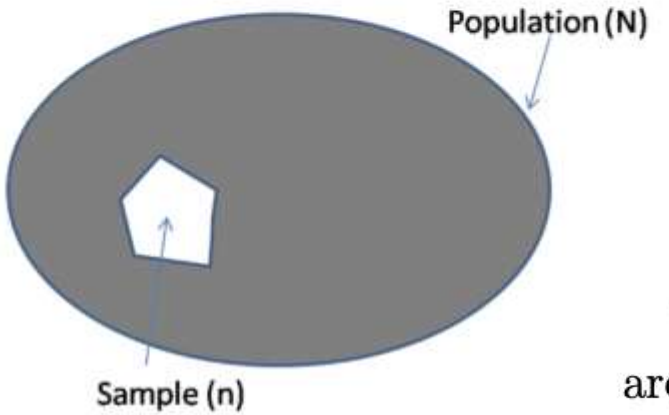
We can see that even for small sample sizes (number of dice rolls), our resulting distribution begins to look more like a Normal distribution. we can also note that as n increases our distribution begins to narrow, i.e. the variance becomes smaller $\frac{\sigma^2}{n}$, but the mean remains the same μ.

2 Statistics

2.1 Sampling

So far we have been dealing with populations, however sometimes the population is too large to be able to analyse and we need to use a sample in order to estimate the population parameters, i.e. mean and variance.

Consider a population of N data points and a sample taken from this population of n data points.



We know that the mean and variance of a population are given by:

$$\text{population mean,} \qquad \mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

and

$$\text{population variance,} \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}$$

But how can we use the sample to estimate our population parameters?

First we define an unbiased estimator. An unbiased estimator is when the expected value of the estimator is exactly equal to the corresponding population parameter, i.e.

if $\bar{x}$ is the sample mean then the unbiased estimator is

$$E(\bar{x}) = \mu$$

where the sample mean is given by:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{n}$$

If S^2 is the sample variance, then the unbiased estimator is

$$E(S^2) = \sigma^2$$

where the sample variance is given by:

$$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

2.1.1 Proof

From the CLT, we know:

$$E(\bar{X}) = \mu$$

and

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Also

$$Var(\bar{X}) = E(\bar{X}^2) - [E(\bar{X})]^2$$

i.e.

$$\frac{\sigma^2}{n} = E(\bar{X}^2) - \mu^2$$

or

$$E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$$

For a single piece of data $n = 1$, so

$$E(\bar{X}_i^2) = \sigma^2 + \mu^2$$

Now

$$\begin{aligned}
E\left[\sum (X_i - \bar{X})^2\right] &= E\left[\sum X_i^2 - n\bar{X}^2\right] \\
&= \sum E(X_i^2) - nE(\bar{X})^2 \\
&= n\sigma^2 + n\mu^2 - n\left(\frac{\sigma^2}{n} + \mu^2\right) \\
&= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 \\
&= (n-1)\sigma^2 \\
\therefore \sigma^2 &= \frac{E\left[\sum (X_i - \bar{X})^2\right]}{n-1}
\end{aligned}$$

2.2 Maximum Likelihood Estimation

The Maximum Likelihood Estimation (MLE) is a statistical method used for fitting data to a model (Data analysis).

We are asking the question:

"Given the set of data, what model parameters is most likely to give this data?"

MLE is well defined for the standard distributions, however in complex problems, the MLE may be unsuitable or even fail to exist.

Note: When using the MLE model we must first assume a distribution, i.e. a parametric model, after which we can try to determine the model parameters.

2.2.1 Motivating example

Consider data from a Binomial distribution with random variable X and parameters $n = 10$ and $p = p_0$. The parameter p_0 is fixed and unknown to us. That is:

$$f(x; p_0) = P(X = x) = \binom{10}{x} P_0^x (1 - p_0)^{10-x}$$

Now suppose we observe some data $X = 3$.

Our goal is to estimate the actual parameter value p_0 based on the data.

Thought Experiments:

let us assume $p_0 = 0.5$, so probability of generating the data we saw is

$$\begin{aligned} f(3; 0.5) &= P(X = 3) \\ &= \binom{10}{3}(0.5)^3(0.5)^7 \\ &\approx 0.117 \end{aligned}$$

Not very high !

How about $p_0 = 0.4$, again

$$\begin{aligned} f(3; 0.4) &= P(X = 3) \\ &= \binom{10}{3}(0.4)^3(0.6)^7 \\ &\approx 0.215 \end{aligned}$$

better.....

So in general let $p_0 = p$ and we want to maximise $f(3;p)$, i.e.

$$f(3;p) = P(X = 3) = \binom{10}{3} P^3(1-p)^7$$

Let us define a new function called the likelihood function $\ell(p;3)$ such that $\ell(p;3) = f(3;p)$. Now we want to maximise this function.

Maximising this function is the same as maximising the log of this function (we will explain why we do this

later!), so let

$$L(p;3) = \log \ell(p;3)$$

therefore,

$$L(p;3) = 3\log p + 7\log(1-p) + \log \binom{10}{3}$$

To maximise we need to find $\frac{dL}{dp} = 0$

$$\begin{aligned} \frac{dL}{dp} &= 0 \\ \frac{3}{p} - \frac{7}{1-p} &= 0 \\ 3(1-p) - 7p &= 0 \\ p &= \frac{3}{10} \end{aligned}$$

Thus the value of p_0 that maximises $L(p;3)$ is $p = \frac{3}{10}$. This is called the **Maximum Likelihood estimate** of p_0.

2.2.2 In General

If we have n pieces of **iid** data $x_1, x_2, x_3, ...x_n$ with probability density (or mass) function $f(x_1, x_2, x_3, ...x_n; \theta)$, where θ are the unknown parameter(s). Then the Maximum likelihood function is defined as

$$\ell(\theta; x_1, x_2, x_3, ...x_n) = f(x_1, x_2, x_3, ...x_n; \theta)$$

and the log-likelihood function can be defined as

$$L(\theta; x_1, x_2, x_3, ...x_n) = \log \ell(\theta; x_1, x_2, x_3, ...x_n)$$

Where the maximum likelihood estimate of the parameter(s) θ_0 can be obtained by maximising $L(\theta; x_1, x_2, x_3, ...x_n)$

2.2.3 Normal Distribution

Consider a random variable X such that $X \sim N(\mu, \sigma^2)$. Let $x_1, x_2, x_3, ...x_n$ be a random sample of iid observations. To find the maximum likelihood estimators of μ and σ^2 we need to maximise the log-likelihood function.

$$
\begin{aligned}
f(x_1, x_2, x_3, ...x_n; \mu, \sigma) &= f(x_1; \mu, \sigma).f(x_2; \mu, \sigma).....f(x_n; \mu, \sigma) \\
\ell(\mu, \sigma; x_1, x_2, x_3, ...x_n) &= f(x_1; \mu, \sigma).f(x_2; \mu, \sigma).....f(x_n; \mu, \sigma) \\
\therefore L(\mu, \sigma; x_1, x_2, x_3, ...x_n) &= \log \ell(\mu, \sigma; x_1, x_2, x_3, ...x_n) \\
&= \log f(x_1; \mu, \sigma) + \log f(x_2; \mu, \sigma) + ... + \log f(x_n \\
&= \sum_{i=1}^{n} log f(x_i; \mu, \sigma)
\end{aligned}
$$

For the Normal distribution

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

so

$$\begin{aligned} L(\mu, \sigma; x_1, x_2, x_3, ...x_n) &= \log\left[\sum_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right] \\ &= -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned}$$

To maximise we differentiate partially with respect to μ and σ set the derivatives to zero and solve. If we were to do this, we would get:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and

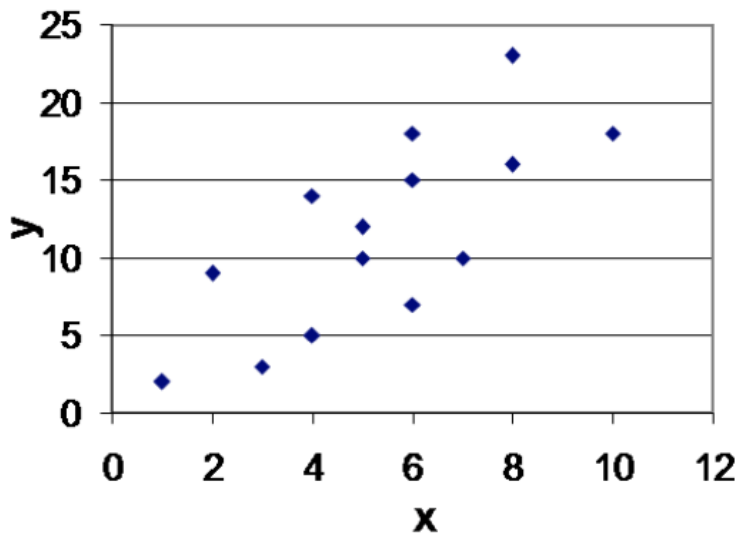
$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

2.3 Regression and Correlation

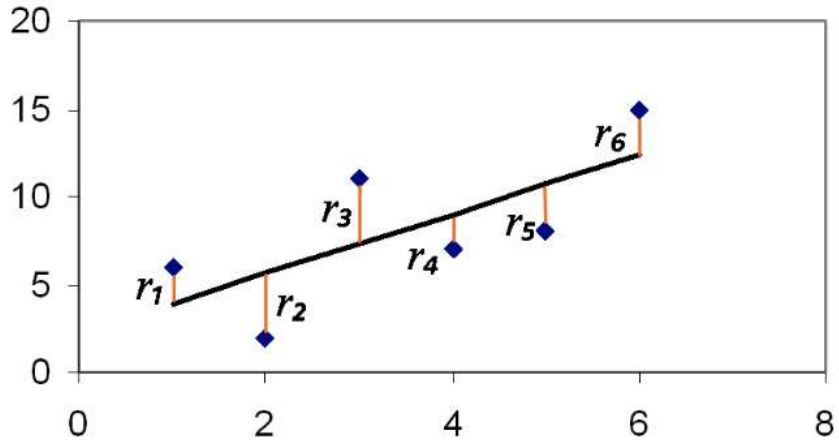
2.3.1 Linear regression

We are often interested in looking at the relationship between two variables (bivariate data). If we can model this relationship then we can use our model to make predictions.

A sensible first step would be to plot the data on a scatter diagram, i.e. pairs of values (x_i, y_i)



Now we can try to fit a straight line through the data. We would like to fit the straight line so as to minimise the sum of the squared distances of the points from the line. The different between the data value and the fitted line is called the residual or error and the technique of often referred to as the method of least squares.



If the equation of the line is given by

$$y = bx + a$$

then the error in y, i..e the residual of the i^{th} data point (x_i, y_i) would be

$$\begin{aligned} r_i &= y_i - y \\ &= y_i - (bx_i + a) \end{aligned}$$

We want to minimise $\sum_{n=1}^{n=\infty} r_i^2$, i.e.

$$S.R = \sum_{n=1}^{n=\infty} r_i^2 = \sum_{n=1}^{n=\infty} [y_i - (bx_i + a)]^2$$

We want to find the b and a that minimise $\sum_{n=1}^{n=\infty} r_i^2$.

$$
\begin{aligned}
S.R &= \sum \left[y_i^2 - 2y_i(bx_i + a) + (bx_i + a)^2 \right] \\
&= \sum \left[y_i^2 - 2by_i x_i - 2ay_i + b^2 x_i^2 + 2bax_i + a^2 \right]
\end{aligned}
$$

or

$$
= n\bar{y^2} - 2bn\bar{x}\bar{y} - 2an\bar{y} + b^2 n\bar{x^2} + 2ban\bar{x} + na^2
$$

To minimise, we want

(i) $\frac{\partial(S.R)}{\partial b} = 0$

(ii) $\frac{\partial(S.R)}{\partial a} = 0$

(i)

$$\frac{\partial(S.R)}{\partial b} = -2n\bar{x}\bar{y} + 2bn\bar{x^2} + 2an\bar{x} = 0$$

(ii)

$$\frac{\partial(S.R)}{\partial a} = -2n\bar{y} + 2bn\bar{x} + 2an = 0$$

These are linear simultaneous equations in b and a and can be solved to get

$$b = \frac{S_{xy}}{S_{xx}}$$

where

$$S_{xx} = \sum(x_i - \bar{x})^2 = \sum(x_i^2) - \frac{\sum(x_i)^2}{n}$$

and

$$S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$a = \bar{y} - b\bar{x}$$

Example

x	5	10	15	20	25	30	35	40
y	98	90	81	66	61	47	39	34

$$\sum x_i = 180 \quad \sum y_i = 516 \quad \sum x_i^2 = 5100 \quad \sum y_i^2 = 37228 \quad \sum x_i y_i = 9585$$

$$S_{xy} = 9585 - \frac{180 \times 516}{8} = -2025$$

$$S_{xx} = 5100 - \frac{180^2}{8} = 1050$$

$$\therefore b = \frac{-2025}{1050} = -1.929$$

$$\bar{x} = \frac{180}{8} = 22.5 \qquad \bar{y} = \frac{516}{8} = 64.5$$

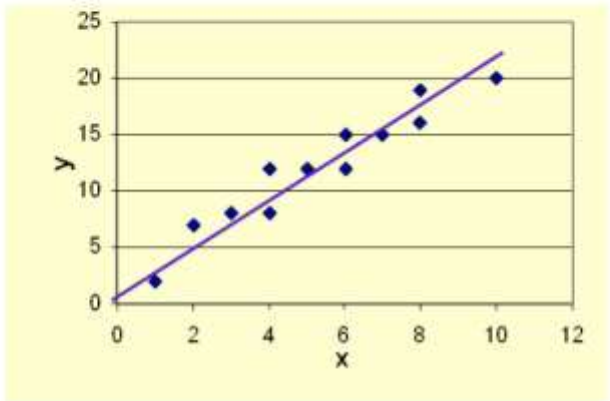
$$\therefore a = 64.5 - (-1.929 \times 22.5) = 107.9$$

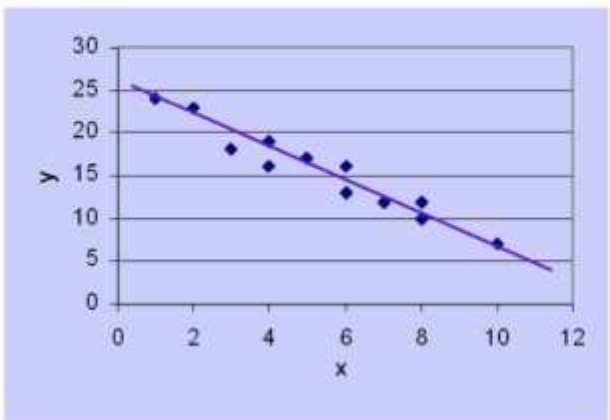
i.e.

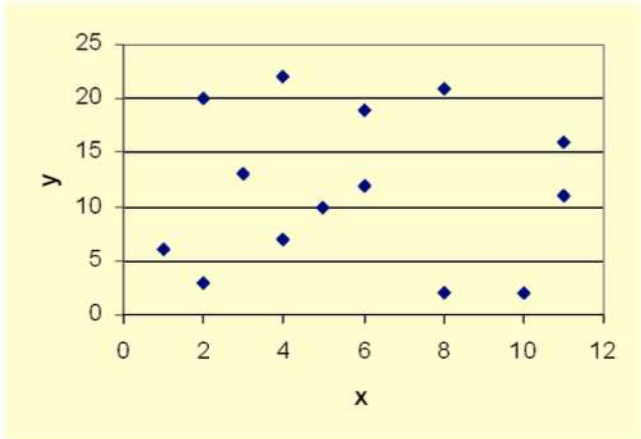
$$y = -1.929x + 107.9$$

2.3.2 Correlation

A measure of how two variables are dependent is their *correlation*. When viewing scatter graphs we can often determine if their is any correlation by sight, e.g.







It is often advantageous to try to quantify the correlation between between two variables, this can be done in a number of ways, two such methods are described.

2.3.3 Pearson Product-Moment Correlation Coefficient

A measure often used within statistics to quantify this is the **Pearson product-moment correlation coefficient**. This correlation coefficient is a measure of linear dependence between two variables, giving a value between $+1$ and -1.

$$\text{PMCC} \quad r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Example

Consider the previous example, i.e.

x	5	10	15	20	25	30	35	40
y	98	90	81	66	61	47	39	34

We calculated,

$$S_{xy} = -2025 \quad \text{and} \quad S_{xx} = 1050$$

also,

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum (y_i^2) - \frac{\sum (y_i)^2}{n}$$

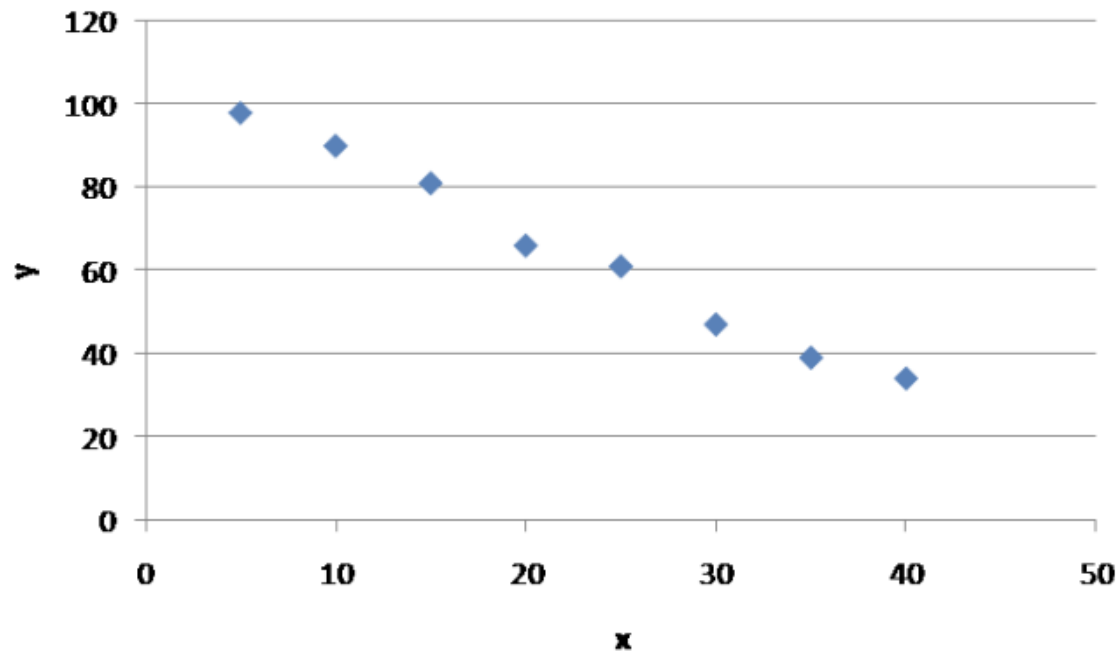
i.e

$$S_{yy} = 37228 - \frac{516^2}{8} = 3946$$

therefore,

$$r = \frac{-2025}{\sqrt{1050 \times 3946}} = -0.995$$

This shows a strong negative correlation and if we were to plot this using a scatter diagram, we can see this visually.



2.3.4 Spearman's Rank Correlation Coefficient

Another method of measuring the relationship between two variables is to use the **Spearman's rank correlation coeffieint**. Instead of dealing with the values of the variables as in the product moment correlation coefficient, we assign a number (rank) to each variable. We then calculate a correlation coefficient based on the ranks. The calculated value is called the Spearmans Rank Correlation Coefficient, r_s, and is an approximation to the PMCC.

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

where d is the difference in ranks and n is the number of pairs.

Example

Consider two judges who score a dancing championship and are tasked with ranking the competitors in order. The following table shows the ranking that the judges gave the competitors.

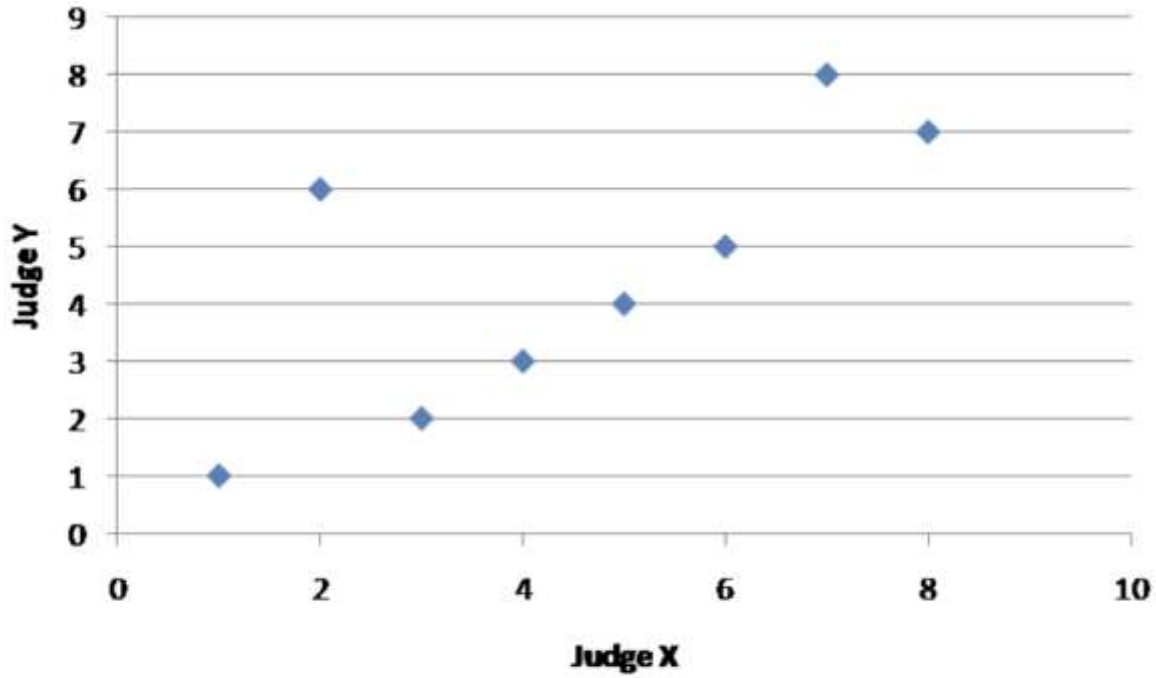
Competitor	A	B	C	D	E	F	G	H
JudgeX	3	1	6	7	5	4	8	2
JudgeY	2	1	5	8	4	3	7	6

calculating d^2, we get

$difference \quad d$	1	0	1	1	1	1	1	4
$difference^2 \quad d^2$	1	0	1	1	1	1	1	16

$$\therefore \sum d_i^2 = 22 \quad \text{and} \quad n = 8$$

$$r_s = 1 - \frac{6 \times 22}{8(8^2 - 1)} = 0.738$$



. strong positive correlation

2.4 Time Series

A time series is a sequence of data points, measured typically at successive times spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones index or the annual flow volume of the Nile River at Aswan.

Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

Two methods for modeling time series data are (i) Moving average models (MA) and (ii) Autoregressive models.

2.4.1 Moving Average

The moving average model is a common approach to modeling univariate data. Moving averages smooth the price data to form a trend following indicator. They do not predict price direction, but rather define the current direction with a lag.

Moving averages lag because they are based on past prices. Despite this lag, moving averages help smooth price action and filter out the noise. The two most popular types of moving averages are the Simple Moving Average (SMA) and the Exponential Moving Average (EMA).

Simple moving average

A simple moving average is formed by computing the average over a specific number of periods.

Consider a 5-day simple moving average for closing prices of a stock. This is the five day sum of closing prices divided by five. As its name implies, a moving average is an average that moves. Old data is dropped as new data comes available. This causes the average to move along the time scale. Below is an example of a 5-day moving average evolving over three days.

Closing Prices	5 day SMA
11	
12	
13	
14	
15	13
16	14
17	15

The first day of the moving average simply covers the last five days. The second day of the moving average drops the first data point (11) and adds the new data point (16). The third day of the moving average continues by dropping the first data point (12) and adding the new data point (17). In the example above, prices gradually increase from 11 to 17 over a total of seven days. Notice that the moving average also rises from 13 to 15 over a three day calculation period. Also notice that each moving average value is just below the last price. For example, the moving average for day one equals 13 and the last price is 15. Prices the prior four days were lower and this causes the moving average to lag.

Exponential moving average

Exponential moving averages reduce the lag by applying more weight to recent prices. The weighting applied to the most recent price depends on the number of periods choosen. There are three steps to calculating an exponential moving average. First, the simple moving average to be used as the previous-period's EMA in the first calculation. Second, calculate the weighting multiplier which depends on period length (below formula is for 10-day EMA, n=10). Third, calculate the rolling exponential moving average for the rest of observations.

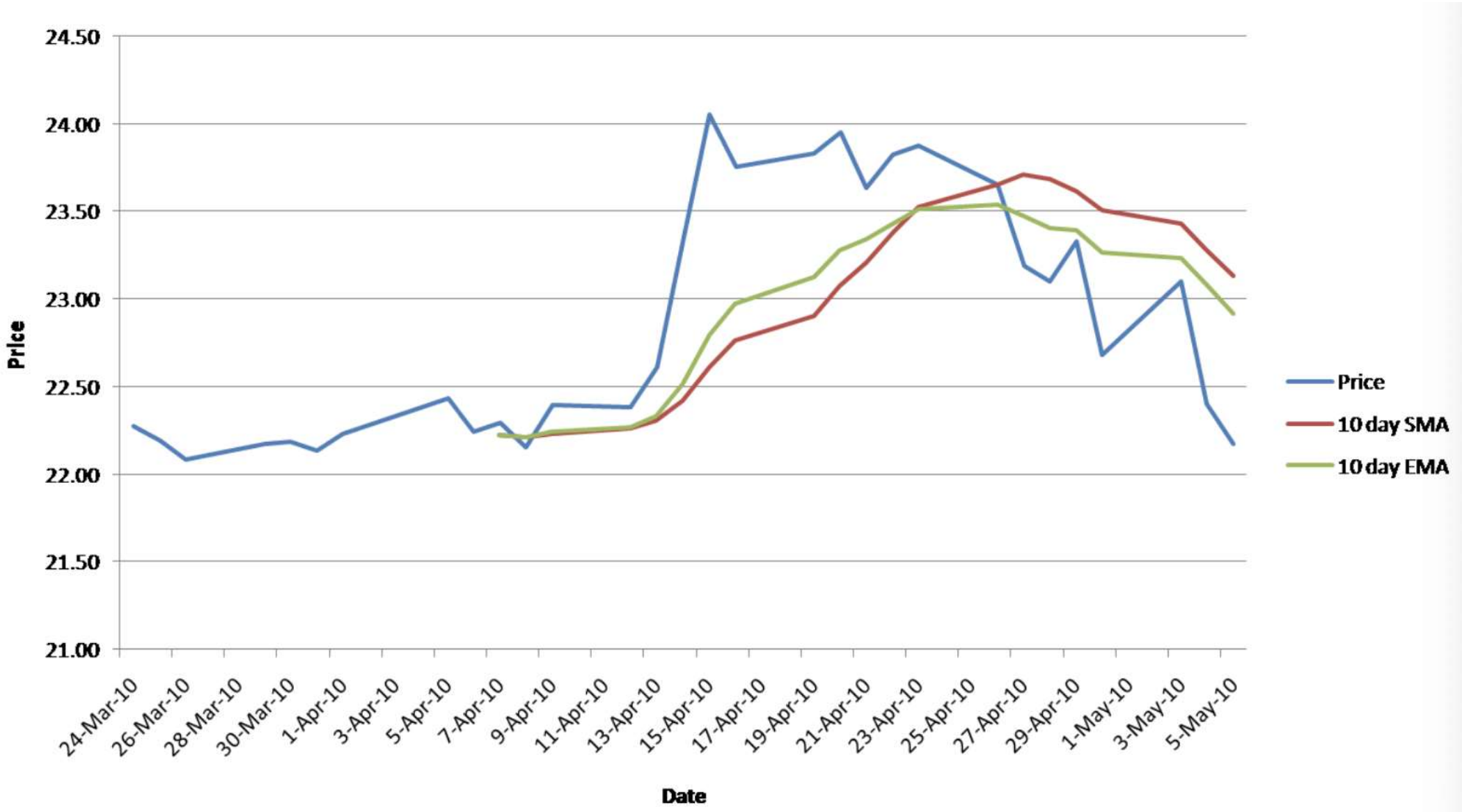
$$E_{i+1} = 2/(\text{n}+1)(P_{i+1} - E_i) + E_i$$

	Date	Price	10-day SMA	Smoothing Constant 2/(10 + 1)	10-day EMA
1	24-Mar-10	22.27			
2	25-Mar-10	22.19			
3	26-Mar-10	22.08			
4	29-Mar-10	22.17			
5	30-Mar-10	22.18			
6	31-Mar-10	22.13			
7	1-Apr-10	22.23			
8	5-Apr-10	22.43			
9	6-Apr-10	22.24			
10	7-Apr-10	22.29	22.22		22.22
11	8-Apr-10	22.15	22.21	0.1818	22.21
12	9-Apr-10	22.39	22.23	0.1818	22.24
13	12-Apr-10	22.38	22.26	0.1818	22.27
14	13-Apr-10	22.61	22.31	0.1818	22.33
15	14-Apr-10	23.36	22.42	0.1818	22.52
16	15-Apr-10	24.05	22.61	0.1818	22.80
17	16-Apr-10	23.75	22.77	0.1818	22.97
18	19-Apr-10	23.83	22.91	0.1818	23.13
19	20-Apr-10	23.95	23.08	0.1818	23.28
20	21-Apr-10	23.63	23.21	0.1818	23.34
21	22-Apr-10	23.82	23.38	0.1818	23.43
22	23-Apr-10	23.87	23.53	0.1818	23.51
23	26-Apr-10	23.65	23.65	0.1818	23.54
24	27-Apr-10	23.19	23.71	0.1818	23.47
25	28-Apr-10	23.10	23.69	0.1818	23.40
26	29-Apr-10	23.33	23.61	0.1818	23.39
27	30-Apr-10	22.68	23.51	0.1818	23.26
28	3-May-10	23.10	23.43	0.1818	23.23
29	4-May-10	22.40	23.28	0.1818	23.08
30	5-May-10	22.17	23.13	0.1818	22.92

10-period exponential moving average applies an 18.18% weighting to the most recent price. 10-period EMA can also be called an 18.18% EMA.

20-period EMA applies a 9.52% weighing to the most recent price $2/(20+1) = .0952$.

The weighting coefficient for the shorter 10-day EMA is higher than coefficient for the longer-perod 20-day EMA. The weighting drops by half every time the moving average period doubles.


24.50
24.00
23.50
23.00
22.50
22.00
21.50
21.00
Price
24-Mar-10
26-Mar-10
28-Mar-10
30-Mar-10
1-Apr-10
3-Apr-10
5-Apr-10
7-Apr-10
9-Apr-10
11-Apr-10
13-Apr-10
15-Apr-10
17-Apr-10
19-Apr-10
21-Apr-10
23-Apr-10
25-Apr-10
27-Apr-10
29-Apr-10
1-May-10
3-May-10
5-May-10
Date
Price
10 day SMA
10 day EMA