

15

NORMAL DISTRIBUTION

SYLLABUS

- Concept of a normal distribution as an example of a continuous probability model and its mean and variance; use of $N(\mu, \sigma^2)$ as a probability model
- Standard normal distribution
- Finding the value of $P(X < x)$ or a related probability, given the values of x, μ, σ
- Symmetry of the normal curve and its properties
- Finding a relationship between x, μ, σ given the value of $P(X < x)$ or a related probability
- Solving problems involving normal random variables
- Solving problems involving the use of $E(aX + b)$ and $\text{Var}(aX + b)$
- Solving problems involving the use of $E(aX + bY)$ and $\text{Var}(aX + bY)$, where X and Y are independent

Exclude normal approximation to binomial distribution

CONTENTS

1	Continuous Random Variables.....	3
1.1	Probability Density Function	3
1.2	Expectation of a Continuous Random Variable, $E(X)$	4
1.3	Variance of a Continuous Random Variable, $\text{Var}(X)$	4
2	Normal Distribution.....	5
2.1	Properties of the Probability Density Function of a Normal Random Variable	5
2.2	Effect of μ and σ on Probability Density Curve of a Normal Distribution	7
2.3	Finding Probabilities Using the Graphing Calculator.....	7
2.4	Finding the Inverse Normal	10
2.5	Standardising a Normal Distribution	13
3	Linear Combination of Independent Normal Random Variables...	15
Annex A:	Approximation of the Binomial Distribution by a Normal Distribution	22
Annex B:	Practice Questions on Normal Distribution	23

LECTURE 1

Lesson Outline

- Introduction to continuous random variables: definition, probability density function, expectation and variance.
 - Understand the properties of the probability density function of a normal distribution and how the parameters of the normal distribution affect its curve.
 - Understand and calculate probabilities involving normal distributions
-

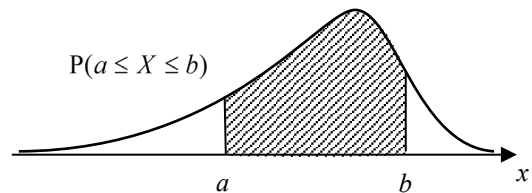
1 CONTINUOUS RANDOM VARIABLES

A **continuous random variable** (CRV) has random values which are defined over a range, i.e., the random values are not countable. Examples include the length of time a person waits for a taxi and the volume of water consumed by a household in a day.

1.1 Probability Density Function

A function $f(x)$ is the **probability density function (pdf)** of a **continuous** random variable X if

- (i) $f(x) \geq 0$ for all real values of x ,
- (ii) $P(a \leq X \leq b) = \int_a^b f(x) \, dx$, This is the shaded area.
- (iii) $\int_{-\infty}^{\infty} f(x) \, dx = 1$.



Notes

- $P(X = a) = \int_a^a f(x) \, dx = 0$
- $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$
($\because P(X = a) = P(X = b) = 0$)
- $P(X > a) = 1 - P(X \leq a)$
- $P(a \leq X \leq b) = P(X \leq b) - P(X < a)$

1.2 Expectation of a Continuous Random Variable, $E(X)$

Recall from Chapter 14 that for a discrete random variable X , its expectation $E(X)$ is given by

$$E(X) = \sum_{\text{all } x} x P(X = x).$$

For a continuous random variable X , the expectation of X (expected value or mean), $E(X)$, is given by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

Notes

- $E(X)$ is a constant.
- $E(X)$ represents the expected outcome if a sufficiently large number of the same experiment is repeated under the same conditions.

1.3 Variance of a Continuous Random Variable, $\text{Var}(X)$

The most common measure of spread of data is the variance (or standard deviation, where standard deviation = $\sqrt{\text{variance}}$). Like the expectation, variance makes use of the information contained in all of the observations. The variance is a measure of the **spread of the observations from the mean**. It is actually a mean of the squared deviations of the observations from the mean.

Consider the continuous random variable X and let $E(X) = \mu$.

The **variance** of X , written $\text{Var}(X)$, of a random variable X is given by

$$\text{Var}(X) = E((X - \mu)^2).$$

The **standard deviation** of X is

$$\sigma = \sqrt{\text{Var}(X)}.$$

The variance and the standard deviation are both at least zero. The greater the deviations, the greater the variance (and the standard deviation) will be.

2 NORMAL DISTRIBUTION

The normal distribution is **the most important** distribution in statistical theory because it is a suitable model for a **very large number** of distributions of data.

A continuous random variable X is said to have a **normal distribution** with **mean** μ and **variance** σ^2 if X has probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ where } -\infty < x < \infty.$$

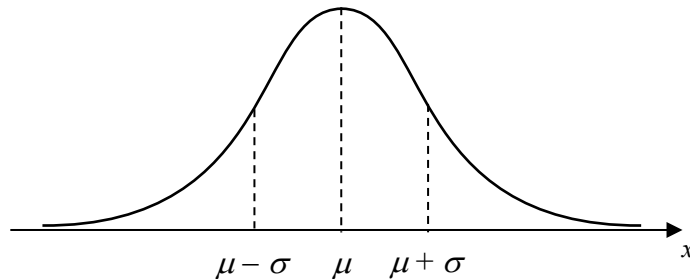
(You are not required to remember the formula.)

If X has a normal distribution with mean μ and variance σ^2 , we write it as

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

2.1 Properties of the Probability Density Function of a Normal Random Variable

The probability density function of a normal random variable is a bell-shaped curve that has the following properties:

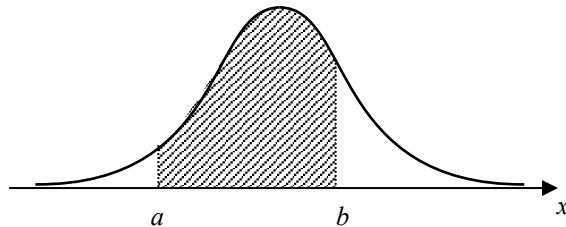


- (i) It is **symmetrical** about its mean μ .
- (ii) The mean, mode and median coincide at $x = \mu$.
- (iii) There is only one peak at $x = \mu$.
- (iv) The curve which lies completely above the x -axis (since $f(x) > 0$ for all real values x) and tapers off rapidly at each “tail” i.e. $f(x) \rightarrow 0^+$ as $x \rightarrow \pm\infty$.
- (v) Total area under the curve $= \int_{-\infty}^{\infty} f(x) dx = 1$.

As the probability is the area under the curve, therefore

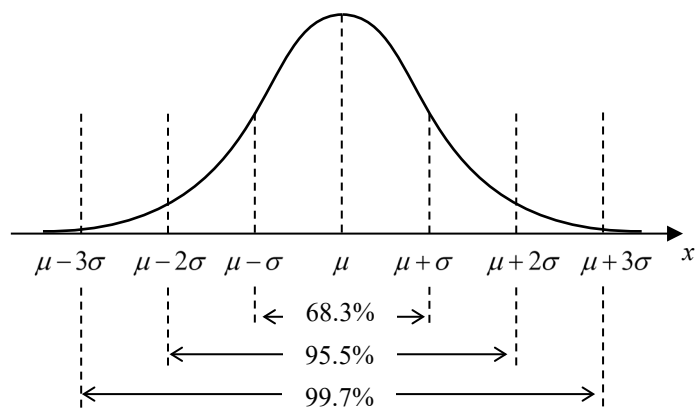
$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b) = \int_a^b f(x) \, dx,$$

where $f(x)$ is the probability density function of X in the normal distribution.



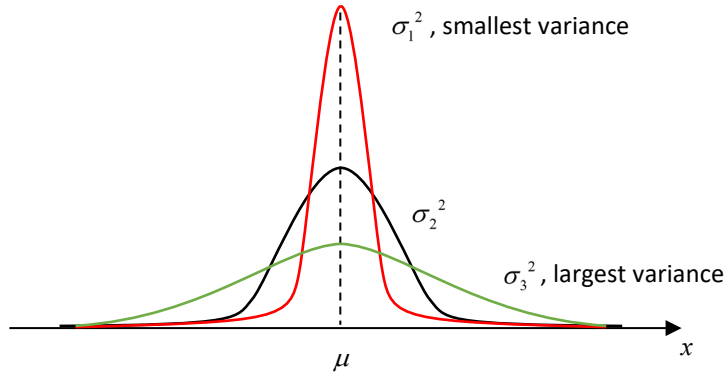
Notes

- Approximately 68.3% of the data falls within one standard deviation about the mean.
- Approximately 95.4% of the data falls within two standard deviations about the mean.
- Approximately 99.7% of the data falls within three standard deviations about the mean.

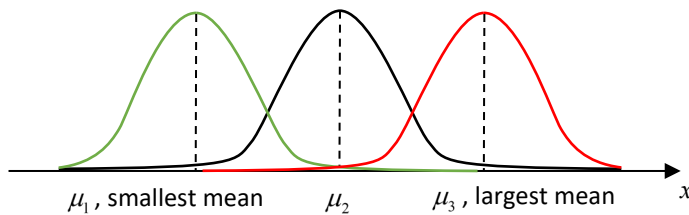


2.2 Effect of μ and σ on Probability Density Function of a Normal Distribution

Density Curves of Normal distributions with the same mean but different variances where $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$



Density Curves of Normal distributions with same variance but different means where $\mu_1 < \mu_2 < \mu_3$



Normal Distribution

An app showing the effect of changing the mean and standard deviation on the normal distribution curve.

<https://www.geogebra.org/m/E7R55dsW>

2.3 Finding Probabilities Using the Graphing Calculator

The function **normalcdf**(in the TI GC can be used to calculate $P(a < X < b)$, where a is the **lower** bound and b is the **upper** bound. If the values μ and σ are not specified, the default values are 0 and 1 respectively.

Example 1

It is given that $X \sim N(10, 4^2)$. For each of the following probabilities,

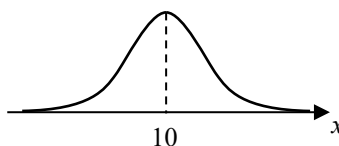
(a) $P(8 < X \leq 12)$ (b) $P(X \leq 14.8)$ (c) $P(X > 6)$

(i) sketch the curve of the distribution of X and **shade** the region representing the probability,

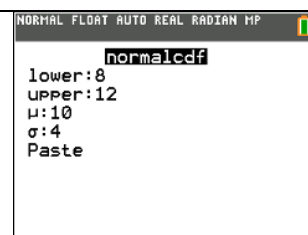
(ii) find the probability, giving your answer correct to 3 decimal places.

Solution

(a)(i) $P(8 < X \leq 12)$

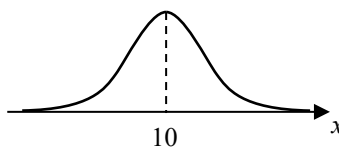


- 1) Press **[2nd] [VARS]** and select **2:normalcdf**.
- 2) Enter the values of a , b , μ and σ .
- 3) Select Paste into the main screen and press **[ENTER]**

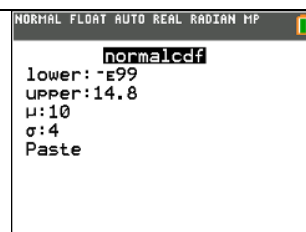


(ii) $P(8 < X \leq 12) =$

(b)(i) $P(X \leq 14.8)$

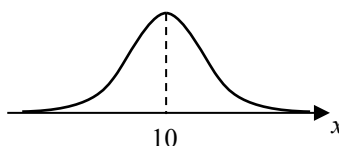


- 1) Follow the same three steps as in (a) except that the lower bound is $-\infty$. This is represented by -10^{99} in the calculator.
- 2) It is keyed in using the keystrokes **[(-)] [2nd] [.] [9] [9]**

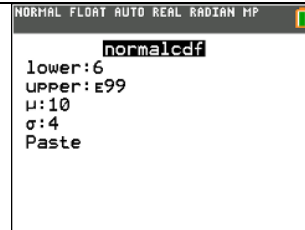


(ii) $P(X \leq 14.8) =$

(c)(i) $P(X > 6)$



- 1) This is the same as (b) except that the upper bound is ∞ . This is represented by 10^{99} in the calculator.
- 2) It is keyed in using the keystrokes **[2nd] [.] [9] [9]**



(ii) $P(X > 6) =$

■

Example 2

Given that $X \sim N(50, 36)$, find the following probabilities.

(a) $P(|X - 50| < 8)$

(b) $P(|X - 48| \geq 5)$

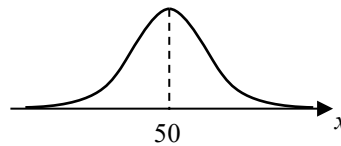
Solution

(a) $P(|X - 50| < 8)$

$$= P(-8 < X - 50 < 8)$$

$$= P(42 < X < 58)$$

=



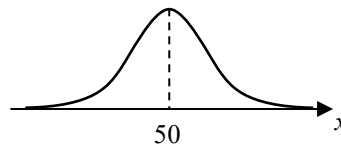
(b) $P(|X - 48| \geq 5)$

$$= P(X - 48 \geq 5 \text{ or } X - 48 \leq -5)$$

$$= P(X \geq 53 \text{ or } X \leq 43)$$

$$= P(X \geq 53) + P(X \leq 43)$$

=



Alternatively,

$$P(|X - 48| \geq 5)$$

$$= 1 - P(|X - 48| < 5)$$

$$= 1 - P(43 < X < 53)$$

=

■

LECTURE 2

Lesson Outline

- Understand and finding the inverse normal using G.C.
- Understand and apply the properties of expectation and variance
- Understand how and know when to standardise a normal distribution

2.4 Finding the Inverse Normal

For $X \sim N(\mu, \sigma^2)$, the function **invNorm**(can be used to find the value of a if the value of $P(X < a)$, $P(X > a)$ or $P(\mu - a < X < \mu + a)$ is known or given.

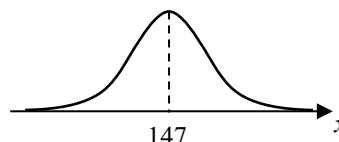
Example 3

Given that $X \sim N(147, 12.25)$, find the value of a (correct to 1 decimal place) such that:

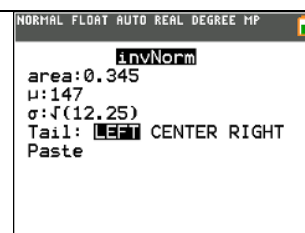
- $P(X \leq a) = 0.345$
- $P(X > a) = 0.82$
- $P(143 \leq X < a) = 0.72$

Solution

- $P(X \leq a) = 0.345$

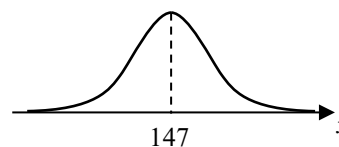


- Press **[2ND] [VARS]** and select **3:invNorm**(.
- Enter the values of area (probability), μ and σ . Under option for **Tail**, select **LEFT**.
- Select **Paste** and press **[ENTER]**.

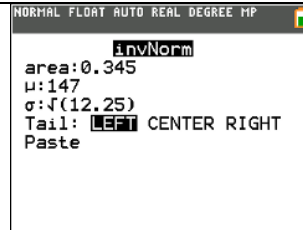


From GC, $a =$

- $P(X > a) = 0.82$

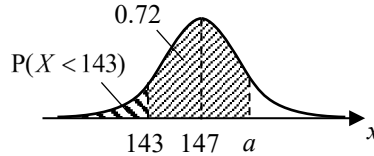


- 1) Enter the values of area (probability), μ and σ . As we are given the probability for the right-tail, under option for **Tail**, select **RIGHT**.
- 2) Select **Paste** and press **[ENTER]**.



From GC, $a =$

(c) $P(143 \leq X < a) = 0.72$



$$P(X < a) = 0.72 + P(X < 143) = 0.846549$$

From GC, $a = 150.6$ (1 d.p.)

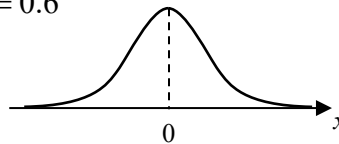
■

Example 4

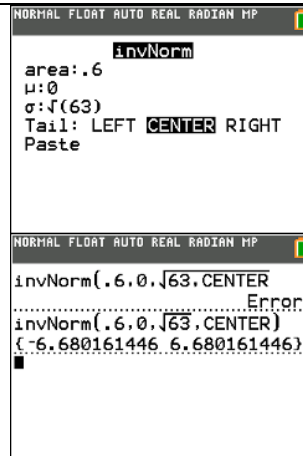
- (a) Given that $X \sim N(0, 63)$, find the value of a such that $P(|X| < a) = 0.6$.
- (b) Given that $Y \sim N(40, 5^2)$, find the value of b such that $P(|Y - 40| > b) = 0.24$.

Solution

(a) $P(|X| < a) = 0.6 \Rightarrow P(-a < X < a) = 0.6$



- 1) As we are given the probability for the central area, under option for **Tail**, select **CENTER**.
- 2) Select **Paste** and press **[ENTER]**.



From GC, $a = 6.68$ (3 s.f.)

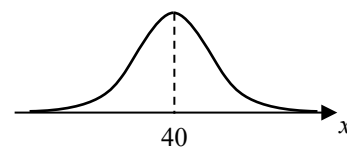
$$(b) \quad P(|Y - 40| > b) = 0.24$$

$$\Rightarrow P(Y - 40 > b \text{ or } Y - 40 < -b) = 0.24$$

$$\Rightarrow 1 - P(40 - b < Y < 40 + b) = 0.24$$

$$\Rightarrow P(40 - b < Y < 40 + b) = 0.76$$

$$\text{From GC, } 40 + b =$$



■

Recall from the previous chapter the following properties of $E(X)$ and $\text{Var}(X)$, which apply to both continuous and discrete random variables. For constants a and b ,

$$E(a) = a$$

$$\text{Var}(a) = 0$$

$$E(aX) = aE(X)$$

$$\text{Var}(aX) = a^2 \text{Var}(X)$$

$$E(aX \pm b) = aE(X) \pm b$$

$$\text{Var}(aX \pm b) = a^2 \text{Var}(X)$$

Example 5

The random variable X has a normal distribution with mean 10 and standard deviation 2. Another random variable Y is defined by $Y = aX + b$, where a and b are positive constants. It is given that $E(Y) = 70$ and $\text{Var}(Y) = 100$. Find a and b .

Solution

$$X \sim N(10, 4) \Rightarrow E(X) = 10, \text{Var}(X) = 4$$

$$Y = aX + b,$$

$$E(Y) = E(aX + b) =$$

$$\Rightarrow$$

$$\text{Var}(Y) = \text{Var}(aX + b) =$$

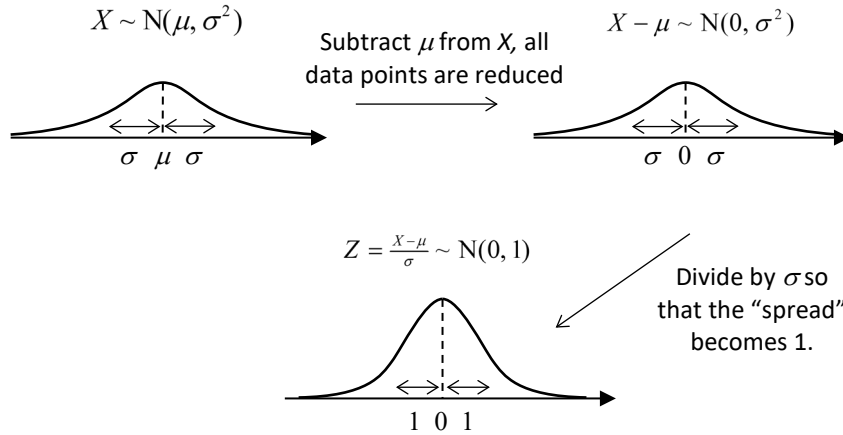
$$\Rightarrow$$

Solving, we get

■

2.5 Standardising a Normal Distribution

To standardise $X \sim N(\mu, \sigma^2)$, let $Z = \frac{X - \mu}{\sigma}$. Then $Z \sim N(0, 1)$.



$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}(E(X) - \mu) = \frac{1}{\sigma}(\mu - \mu) = 0$$

$$\text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2}[\text{Var}(X) + \text{Var}(-\mu)] = \frac{1}{\sigma^2}(\sigma^2 + 0) = 1$$

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) \\ &= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) \end{aligned}$$

Notes

- The standard normal distribution is the normal distribution with mean 0 and variance 1.
- Z is used to denote the standard normal distribution, i.e. $Z \sim N(0, 1)$.
- When solving problems where μ and/or σ are **unknown**, we need to standardise $X \sim N(\mu, \sigma^2)$ to $Z \sim N(0, 1)$ by using the formula

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



Standardisation of Normal Distribution

This applet shows the effect of standardization and how the "area" or probability is unchanged after standardization.

<https://www.geogebra.org/m/twuqGJHr>

Example 6

Given that $X \sim N(\mu, 25)$ and $P(X < 35) = 0.7257$, find the value of μ .

Solution

$$P(X < 35) = 0.7257$$

$$P\left(Z < \frac{35 - \mu}{5}\right) = 0.7257$$

From G.C.

$$\Rightarrow \mu =$$

■

Example 7

A random variable Y has a normal distribution. Calculate the mean and standard deviation of Y given that

$$P(Y > 5.82) = 0.04 \text{ and } P(Y < -3.74) = 0.18.$$

The random variable W , also normally distributed, has the same standard deviation as Y but its mean is unknown. Find the greatest possible value of $P(-3.74 < W < 5.82)$.

Solution

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

$$P(Y > 5.82) = 0.04$$

$$P(Y < -3.74) = 0.18$$

$$P\left(Z > \frac{5.82 - \mu}{\sigma}\right) = 0.04$$

$$P\left(Z < \frac{-3.74 - \mu}{\sigma}\right) = 0.18$$

From G.C.,

$$\frac{5.82 - \mu}{\sigma} = 1.7507 \text{ (5 s.f.)}$$

$$\frac{-3.74 - \mu}{\sigma} = -0.91537 \text{ (5 s.f.)}$$

$$\mu + 1.7507\sigma = 5.82$$

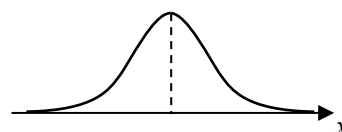
$$\mu - 0.91537\sigma = -3.74$$

Using G.C., we get

$$\mu = -0.45766 = -0.458 \text{ (3 s.f.)}$$

$$\sigma = 3.5858 = 3.59 \text{ (3 s.f.)}$$

$$W \sim N(\mu', 3.5858^2)$$



■

LECTURE 3

Lesson Outline

- Understand and practise problems involving linear combination of independent normal variables.
-

3 LINEAR COMBINATION OF INDEPENDENT NORMAL RANDOM VARIABLES

Linear combinations of independent **normal** variables yield variables that are also **normally distributed**.

Specifically, if X and Y are any two **independent** variables, and a and b are constants, then $aX \pm bY$ is also normally distributed with the means and variances calculated as follows:

$$E(aX \pm bY) = a E(X) \pm b E(Y)$$

$$\text{Var}(aX \pm bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$$

If $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$, and a and b are constants, then

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

$$aX - bY \sim N(a\mu_1 - b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

This result can be easily extended to linear combination of any number of independent normal variables.

In particular, if X_1, X_2 are **independent observations** of X where $X \sim N(\mu, \sigma^2)$, then X_1 and X_2 have the same distribution as X , i.e. $E(X_1) = E(X_2) = E(X) = \mu$ and $\text{Var}(X_1) = \text{Var}(X_2) = \text{Var}(X) = \sigma^2$. Therefore

$$X_1 + X_2 \sim N(\mu + \mu, \sigma^2 + \sigma^2) = N(2\mu, 2\sigma^2)$$

and the sum of n independent observations of X

$$X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2).$$

Example 8

Let X be a normal variable denoting the mass of a randomly chosen packet of biscuits with mean 100 g and variance 10 g^2 . Find the mean and variance of

Know the difference between the variables 'total weight of two packets' and 'twice the weight of one packet'

- (i) the total weight of two randomly chosen packets and
- (ii) twice the weight of a randomly chosen packet.

Write down the distributions of both the variables in (i) and (ii).

Solution

$$X \sim N(100, 10)$$

- (i) Total weight of two randomly chosen packets $= X_1 + X_2$

$$E(X_1 + X_2) =$$

$$\text{Var}(X_1 + X_2) =$$

- (ii) Twice the weight of a randomly chosen packet $= 2X$

$$E(2X) =$$

$$\text{Var}(2X) =$$

■

Example 9

The continuous random variables X and Y have independent normal distributions. The means of X and Y are 10 and 12 respectively, and the standard deviations are 2 and 3 respectively. Find

- (i) $P(Y < 10)$,
- (ii) $P(Y < X)$,
- (iii) $P(4X + 5Y > 90)$,
- (iv) the value a such that $P(X_1 + X_2 > a) = \frac{1}{4}$, where X_1 and X_2 are independent observations of X . [N87/II/9]

Solution

$$X \sim N(10, 2^2), Y \sim N(12, 3^2)$$

$$(i) \quad P(Y < 10) = 0.252 \text{ (3 s.f.)}$$

$$(ii) \quad P(Y < X) = P(Y - X < 0)$$

Move all the variables to one side of the inequality and proceed to find the new mean & variance

$$(iii) \quad 4X + 5Y \sim$$

$$(iv) \quad X_1 + X_2 \sim$$

■

Example 10

In a junior college, the male students have masses which are normally distributed with mean 70 kg and standard deviation 8 kg. Three male students are to be chosen at random from this junior college. Find the probability that

- (i) none of them will have a mass greater than 66kg,
- (ii) their mean mass will be greater than 72kg.

The female students in the same junior college have masses which are normally distributed with mean 60kg and standard deviation 6kg.

- (iii) Find the probability that a male student and a female student, selected at random, will have masses differing by less than 5kg.

Solution

- (i) Let X be random variable “mass of a male student.”

$$X \sim N(\quad, \quad)$$

- (ii) Mean mass of three male students

- (iii) Let Y be the random variable “mass of a female student.”

$$Y \sim N(\quad, \quad)$$

■

Example 11

The weight of contents of a randomly chosen packet of keropok brand A may be taken to have a normal distribution with mean 625 grams and standard deviation 15 grams. The weight of packaging may be taken to have an independent normal distribution with mean 25 grams and standard deviation 3 grams. Find

- (i) the probability that a randomly chosen packet of keropok brand A has a total weight exceeding 630 grams,
- (ii) the probability that the total weight of the contents of four randomly chosen packets of keropok brand A exceeds 2450 grams,

The weight of the contents of a randomly chosen packet of keropok brand B may be taken to have a normal distribution with mean 465 grams and standard deviation 10 grams.

- (iii) Find the probability that four times the contents of a randomly chosen packet of keropok brand B weigh more than the contents of three randomly chosen packets of keropok brand A . [J87/modified]

Solution

Let X be the random variable “weight of contents of a packet of keropok brand A ”

Let T be the random variable “weight of packaging.”

$$X \sim N(625, 15^2), T \sim N(25, 3^2)$$

- (i) $X + T$ is total weight of a packet of keropok of brand A

$$X + T \sim N(625 + 25, 15^2 + 3^2) = N(650, 234)$$

$$P(X + T > 630) = 0.904 \text{ (3 s.f.)}$$

- (ii) $X_1 + X_2 + X_3 + X_4 \sim N(4(625), 4(15^2)) = N(2500, 900)$

$$P(X_1 + X_2 + X_3 + X_4 > 2450) = 0.952 \text{ (3 s.f.)}$$

- (iii) Let Y be the random variable “weight of contents of a packet of keropok brand B .”

$$Y \sim N(465, 10^2)$$

■

Example 12

Melons are sold by weight at a price of \$1.50 per kg. The masses of melons are normally distributed with a mean of 0.8 kg and a standard deviation of 0.1 kg. Pumpkins are sold by weight at a price of \$0.50 per kg. The masses of pumpkins are normally distributed with a mean of 1.2 kg and a standard deviation of 0.2 kg. Find the probability that the total price of 5 randomly chosen melons and 3 randomly chosen pumpkins exceeds \$8.

Mrs Lee can carry a maximum load of 10 kg and she wants to buy a total of 10 melons and pumpkins. What is the maximum number of pumpkins that she can buy so that there is more than 95% chance that she is able carry the load? [N2002/II/28]

Solution

Let X and Y be the random variable “weight of a melon” and “weight of a pumpkin” respectively.

$$X \sim N(0.8, 0.1^2), Y \sim N(1.2, 0.2^2)$$

Use a new letter to define sum of variables to keep your working tidy.

$$\text{Let } T = 1.5(X_1 + X_2 + X_3 + X_4 + X_5) + 0.5(Y_1 + Y_2 + Y_3)$$

$$\begin{aligned} E(T) &= E(1.5(X_1 + X_2 + X_3 + X_4 + X_5) + 0.5(Y_1 + Y_2 + Y_3)) \\ &= 1.5E(X_1 + X_2 + X_3 + X_4 + X_5) + 0.5E(Y_1 + Y_2 + Y_3) \\ &= 1.5(5)E(X) + 0.5(3)E(Y) = 7.8 \end{aligned}$$

$$\begin{aligned} \text{Var}(T) &= \text{Var}(1.5(X_1 + X_2 + X_3 + X_4 + X_5) + 0.5(Y_1 + Y_2 + Y_3)) \\ &= 1.5^2 \text{Var}(X_1 + X_2 + X_3 + X_4 + X_5) + 0.5^2 \text{Var}(Y_1 + Y_2 + Y_3) \\ &= 1.5^2(5) \text{Var}(X) + 0.5^2(3) \text{Var}(Y) = 0.1425 \end{aligned}$$

$$\therefore T \sim N(7.8, 0.1425)$$

$$\therefore P(T > 8) = 0.29812 = 0.298 \text{ (3 s.f.)}$$

Let n be the number of pumpkins Mrs Lee buys.

Hence she will be buying $(10 - n)$ melons.

$$\text{Let } W = (X_1 + X_2 + \cdots + X_{10-n}) + (Y_1 + Y_2 + \cdots + Y_n)$$

$$\begin{aligned} E(W) &= E((X_1 + X_2 + \cdots + X_{10-n}) + (Y_1 + Y_2 + \cdots + Y_n)) \\ &= (10 - n)E(X) + nE(Y) \\ &= (10 - n)(0.8) + n(1.2) \\ &= 8 + 0.4n \end{aligned}$$

$$\begin{aligned} \text{Var}(W) &= \text{Var}((X_1 + X_2 + \cdots + X_{10-n}) + (Y_1 + Y_2 + \cdots + Y_n)) \\ &= (10 - n) \text{Var}(X) + n \text{Var}(Y) \\ &= (10 - n)(0.1^2) + n(0.2^2) \\ &= 0.1 + 0.03n \end{aligned}$$

$$\therefore W \sim N(8 + 0.4n, 0.1 + 0.03n)$$

$$P(\text{she can handle load}) = P(W \leq 10) > 0.95$$

$$P\left(Z \leq \frac{10 - (8 + 0.4n)}{\sqrt{0.1 + 0.03n}}\right) = P\left(Z \leq \frac{2 - 0.4n}{\sqrt{0.1 + 0.03n}}\right) > 0.95$$

$$\text{From G.C., } \frac{2 - 0.4n}{\sqrt{0.1 + 0.03n}} > 1.6449$$

Since n is an integer, we can solve this inequality using **table**:

$$\text{When } n = 3, \frac{2 - 0.4n}{\sqrt{0.1 + 0.03n}} = 1.84 > 1.6449$$

$$\text{When } n = 4, \frac{2 - 0.4n}{\sqrt{0.1 + 0.03n}} = 0.852 < 1.6449$$

Hence the maximum number of pumpkins she can buy is 3. ■

ANNEX A

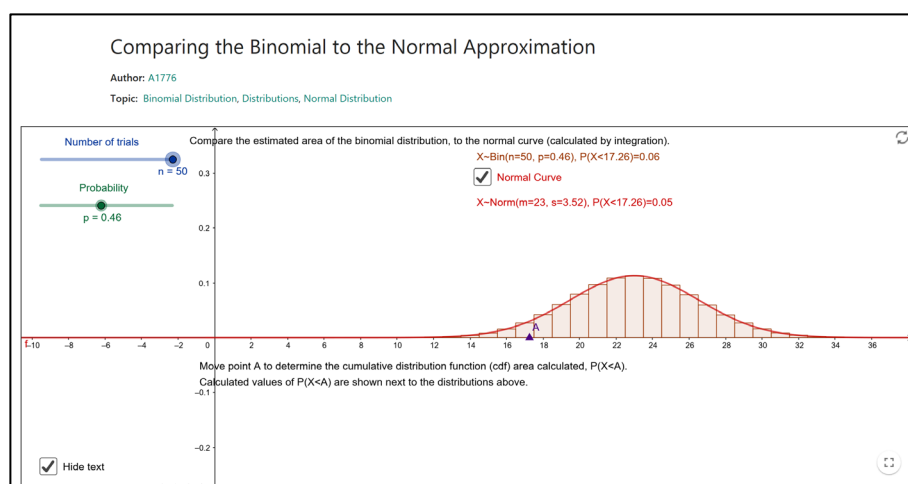
Approximation of a Binomial Distribution by a Normal Distribution

Suppose we plot the histogram for the distribution $B(50, 0.46)$ and the density curve of the distribution $N(23, 3.52^2)$. Notice the density curve of the normal distribution approximates the histogram of the binomial distribution quite closely.



You may also try out the following applets to understand the approximation better:

<https://www.geogebra.org/m/RsQZMejx>



In general, the discrete random variable X with distribution $B(n, p)$ where

n is large (> 50) and $np > 5$ and $nq > 5$

is approximately normal, with its mean $\mu = np$ and variance $\sigma^2 = npq$, where $q = 1 - p$ i.e.,

$X \sim N(np, npq)$ approximately.

Other Relevant Websites



Use the applet to visualise the probability as the area under the normal curve:

<https://www.geogebra.org/m/E7R55dsW>



Illustrating the concept of standardisation:

<https://www.geogebra.org/m/twuqGJHr>

ANNEX B

Practice Questions on Normal Distribution

- 1 Given that $X \sim N(5, 4)$, find
 - (i) $P(4 \leq X < 6)$
 - (ii) $P(X < 7.8)$
 - (iii) $P(|X| < 3)$
- 2 Given that $X \sim N(8, 2.5^2)$, find
 - (i) $P(0 \leq X < 4)$
 - (ii) $P(X \leq 6)$
 - (iii) $P(X > 9.2)$
 - (iv) $P(|X| > 9.2)$
- 3 Given that $X \sim N(400, 64)$, find a such that $P(|X - 400| < a) = 0.9$.
- 4 At two years of age, sardines inhabiting Japanese waters have a length distribution that is approximately normal with mean 20.20 cm and standard deviation 0.65 cm.
 - (a) What proportion of these sardines are between 19.0 and 21.0 cm long?
 - (b) How long are the longest 15% of all these sardines?
 - (c) How short are the shortest 15% of all these sardines?
 - (d) In what range do the middle 70% of all sardines fall?
- 5 The lengths of certain items follow a normal distribution with mean μ cm and standard deviation 6 cm. It is known that 4.78% of the items have a length greater than 82 cm. find the value of the mean μ .
- 6 The marks in an examination were normally distributed with mean μ and standard deviation σ . Of all the candidates, 10% had more than 75 marks and 20% had less than 40 marks. Find the values of μ and σ .

Answers

- 1 (i) 0.383 (ii) 0.919 (iii) 0.841 (iv) 0.159
- 2 (i) 0.0541 (ii) 0.212 (iii) 0.316 (iv) 0.316
- 3 $a = 13.2$
- 4 (a) 0.8585 (b) 20.88 cm (c) 19.52 cm (d) between 19.52 and 20.88 cm
- 5 72 cm
- 6 53.87, 16.48