RAFFLES INSTITUTION
H2 Mathematics (9758)
2017 Year 6

Chapter S3: Normal Distribution

SYLLABUS INCLUDES

- 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 $P(X < x_1)$ or a related probability, given the values of x_1, μ, σ
- Symmetry of the normal curve and its properties
- Finding a relationship between x_1, μ, σ given the value of $P(X < x_1)$ or a related probability
- Solving problems involving the use of $E(aX + b)$ and $Var(aX + b)$
- Solving problems involving the use of $E(aX + bY)$ and $Var(aX + bY)$, where X and Y are independent

PRE-REQUISITES

- Concepts of random variable, expectation, variance/standard deviation

CONTENT

INTRODUCTION

In Chapter S2, we learnt about discrete random variables and a special discrete probability distribution, the binomial distribution.

In this chapter, we shall learn about continuous random variables and the most important continuous distribution in statistics – the normal distribution.

1 CONTINUOUS RANDOM VARIABLE

Recall that a random variable is a quantity that takes different numerical values according to the outcome of a random experiment.

A continuous random variable can take any value in a given range, and it best describes data such as height, mass, time, distance, etc.

While a discrete random variable is defined by its probability distribution, a continuous random variable is defined by its probability density function.

The probability density function is represented by a curve $y = \mathrm{f}(x)$, and the **probabilities are given by the area under the curve.**

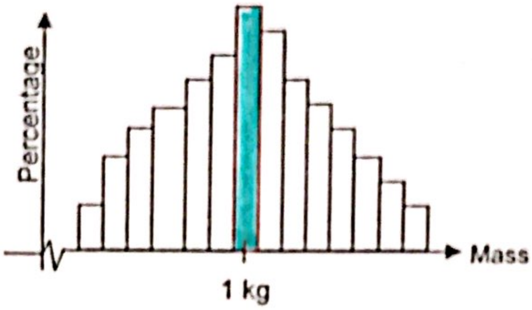
As in the case of discrete random variables, we are also interested in the expectation and variance of continuous random variables.

(Refer to Appendix 1 for more information on the probability density function, expectation and variance of continuous random variables)

2 NORMAL DISTRIBUTION AND NORMAL CURVE

2.1 Normal Distribution

Consider the following situation:
A large number of 1kg bags of sugar are weighed to check how accurately they have been filled, and the results are shown in the histogram below.



This histogram is similar to the histograms you might obtain for a variety of different types of data, such as the heights of 5 year-old girls, or the time taken to run 2.4km by 18 year-old boys, or the lifespans of batteries, etc,

Observe that
- the histogram is (almost) symmetrical about the mean, and
- the percentages of bags with weight close to the mean is higher than the percentage of bags with weight further away from the mean.

If a smooth curve is drawn through the top of the columns, this will give a distribution which is roughly 'bell'-shaped.

Such a distribution can be modelled by the **normal distribution**, whose probability density function is given by

$$\mathrm{f}(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\quad x\in\mathbb{R}.$$

Note : There is no need to memorize the complicated function above.

> If a continuous random variable X follows a normal distribution, we write $X\sim \mathrm{N}(\mu,\sigma^2)$, where
> - $\mathrm{E}(X)=\mu$,
> - $\mathrm{Var}(X)=\sigma^2$.
>
> mean variance
> σ = standard deviation

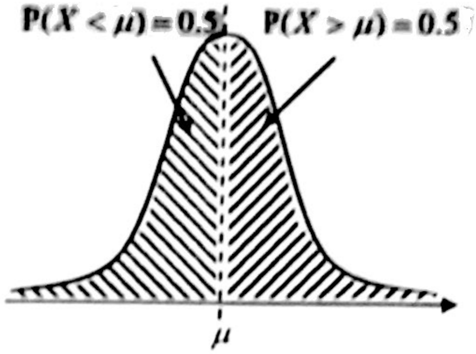
Remarks :
Recall that in section 1.4 of Chapter S2B, we looked at the graphs of the probability distribution of the random variable X, where $X\sim \mathrm{B}(n,p)$. It can be observed that for large n and value of p which is not too close to 0 or 1, the shape of the graph of the probability distribution of X becomes symmetrical and bell-shaped. In fact, for large n and value of p which is not too close to 0 or 1, the Binomial distribution can be approximated using a normal distribution. (Refer to Appendix 2 for more details)

2.2 Normal Curve and its Properties

Let $X \sim \mathrm{N}(\mu,\sigma^2)$.

Properties of the normal curve are as follow:

(1) It is **symmetrical** about the line $x=\mu$.



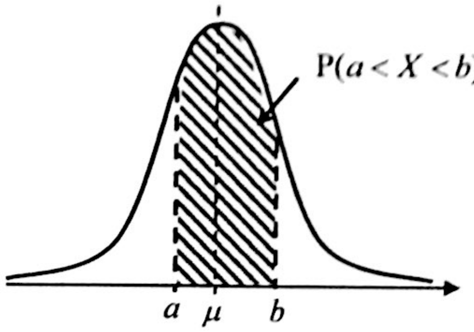
(2) The **mean, median and mode** are all equal to μ.

(3) It approaches the x – axis as $x \to \pm\infty$.

(4) Area under the graph gives the probabilities, i.e.,

$$\mathrm{P}(a<X<b)=\int_a^b \mathrm{f}(x)\,\mathrm{d}x$$

where $y=\mathrm{f}(x)$ represents the probability density function of the normal curve.

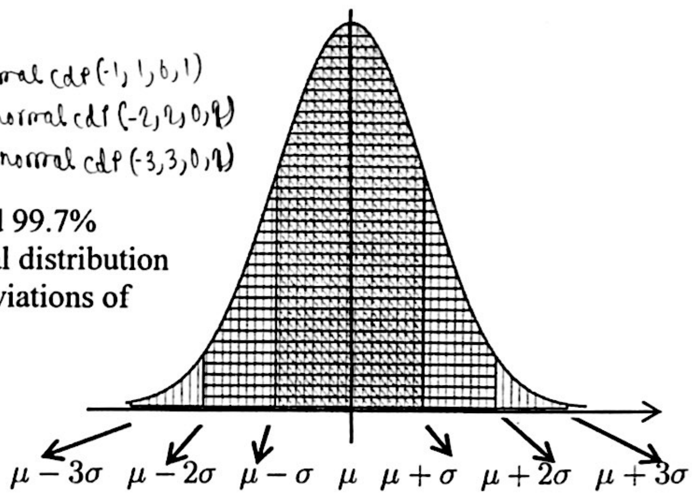


Hence $\mathrm{P}(a<X<b)$ is given by area under the ~~graph~~ graph from x=a to x=b.

(5) Total area under the curve is 1.

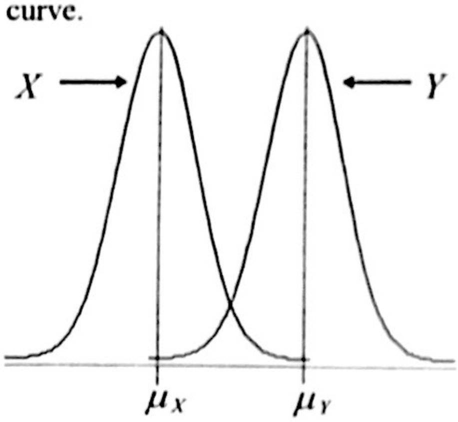
(6) $\mathrm{P}(\mu-\sigma<X<\mu+\sigma)\approx 0.68$ normal cdf (-1, 1, 0, 1)
$\mathrm{P}(\mu-2\sigma<X<\mu+2\sigma)\approx 0.95$ normal cdf (-2, 2, 0, 1)
$\mathrm{P}(\mu-3\sigma<X<\mu+3\sigma)\approx 0.997$ normal cdf (-3, 3, 0, 1)

i.e., approximately 68%, 95% and 99.7% of the values drawn from a normal distribution lies within 1, 2 and 3 standard deviations of the mean respectively.



The shape of the normal curve is completely determined by the values of μ and σ.

The following figures illustrate how the mean and the standard deviation affect the shape of the normal curve.



same "spread": σ

Figure 1

X and Y have the **same standard deviation** but **different means**, with $\mu_Y > \mu_X$

↳ Y

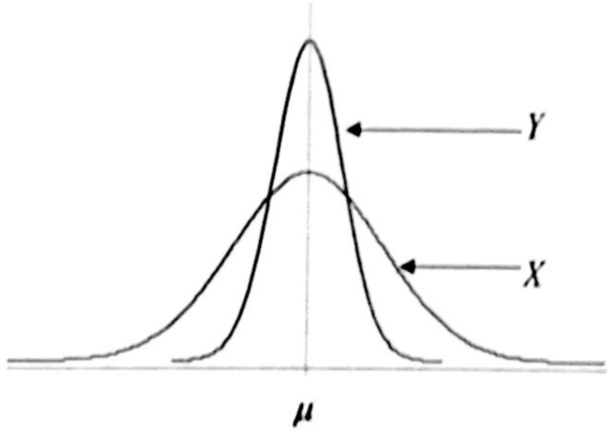


Figure 2

X and Y have the **same mean** but **different standard deviations**, with $\sigma_X > \sigma_Y$

↳ wider, 'fatter' spread

(Refer to Appendix 3 for the steps to sketch normal curves using GC and explore the curves with different means and variances)

2.3 Use of GC to Evaluate Normal Probabilities

Let's revisit our example of 1kg bags of sugar.
Suppose the mass of a 1kg bag of sugar follows a normal distribution with mean 1kg and standard deviation 50g.

What is the probability that a randomly chosen 1 kg bag of sugar weighs
(i) at most 980g? (ii) less than 980g?

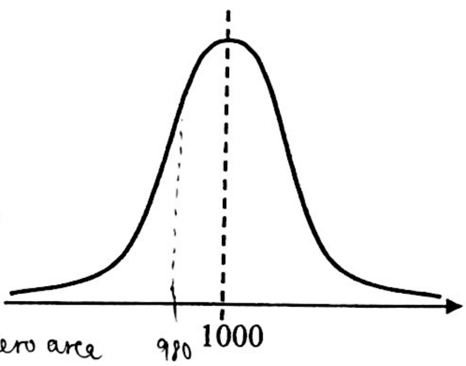
Let X be the mass of a 1kg bag of sugar in grams.
Then $X \sim \mathrm{N}(1000, 50^2)$.

To evaluate $\mathrm{P}(X \leq 980)$, we need to find the shaded area under the graph (refer to diagram),
i.e. we need to find $\int_{-\infty}^{980} \mathrm{f}(x)\, \mathrm{d}x$,
where $y = \mathrm{f}(x)$ represents the probability density function of the normal curve.

(i) From GC,
$$\mathrm{P}(X \leq 980) = 0.345 \text{ (3 s.f.)}$$

(ii) From the sketch, it is clear that
$$\mathrm{P}(X < 980) = \mathrm{P}(X \leq 980) = 0.345 \text{ (3 s.f.)}$$

P(X = 980) = 0
↳ boundary values, lines have zero area



From the above example, we can conclude that if X is a **continuous random variable**, then

$$P(a<X<b) = P(a \le X \le b) = P(a < X \le b) = P(a \le X < b)$$

as $P(X=a) = P(X=b) = 0$

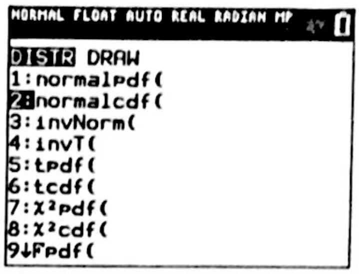
in discrete random variables $P(X \le 2) \ne P(X<2)$

To evaluate $P(X \le 980)$, given $X \sim N(1000, 50^2)$:

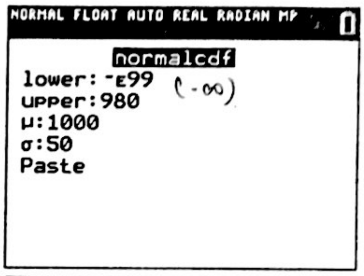
$= \int_{-\infty}^{980} f(x)\,dx$

1 Press [2nd][VARS]

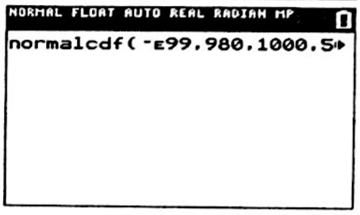
Select **2: normalcdf(**



2 Leave lower as –E99, which represents -1×10^{99}
Enter 980 for upper
1000 for μ
50 for σ
Highlight **Paste** and press [ENTER] twice



$(-\infty)$



Note: To find $P(a \le X \le b)$,
Enter a for lower
b for upper

To find $P(X \ge c)$,
Enter c for lower
E99 [2nd][,][9][9] for upper or 10^{99}

Example 1

$\mu = 65$ $\sigma = 3$

The weight of a randomly chosen student from a population may be assumed to have a normal distribution with mean 65 kg and standard deviation 3 kg. A student was randomly chosen from this population. Find the probability that

(i) his weight exceeds 70 kg,
(ii) his weight is below 62 kg,
(iii) his weight is between 60 kg and 75 kg,
(iv) his weight is within one standard deviation from the mean weight.

Solution

Important!
Always define the random variable in the context of the question.

Let X be the weight of a student in kg (always include units)

then $X \sim N(65, 3^2)$ $\hookrightarrow \sigma^2$

i) $P(X > 70) = 0.0478$ (3 s.f.)

ii) $P(X < 62) = 0.159$ (3 s.f.)

iii) $P(60 < X < 75) = 0.952$ (3 s.f.)

i) lower 70, upper E99, μ 65, σ 3

iv) $P(62 < x < 68) =$

$P(|X-65| < 3) = P(-3 < X - 65 < 3) = P(62 < X < 68)$
$= 0.683$ (3 s.f.)

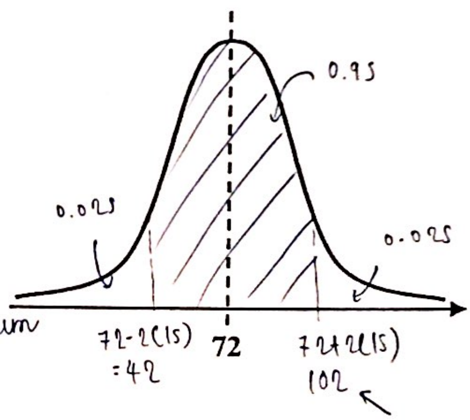
don't rly need to open up the modulus

Example 2 [8863/2008/Q7 part (modified)]

An examination, taken by a large number of candidates, is marked out of 100. The mean mark for all candidates is 72 and the standard deviation is 15.
Give a reason why a normal distribution, with this mean and standard deviation, would not give a good approximation to the distribution of marks.

Solution

Let X be the mark of a candidate.
If $X \sim N(72, 15^2)$,
then $P(72 - 2(15) < X < 72 + 2(15)) \approx 0.95$



From the diagram,

$P(X > 72 + 2(15)) \approx 0.025$ (by symmetry)

but $72 + (15)(2) = 102$ is already greater than the maximum mark of 100.

$\Rightarrow P(X > 72 + 2(15)) = 0$ which contradicts the above

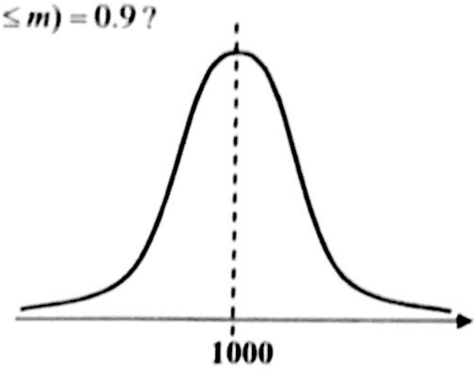
Hence a normal distribution would not give a good approximation to the distribution of marks.

2.4 Use of GC to Evaluate Inverse Normal Values

Back to our example of 1kg bags of sugar.

Recall that $X \sim \mathrm{N}(1000, 50^2)$ where X denotes the mass of a 1kg bag of sugar in grams.

What is the value of m (to nearest gram) such that $\mathrm{P}(X \le m) = 0.9$?



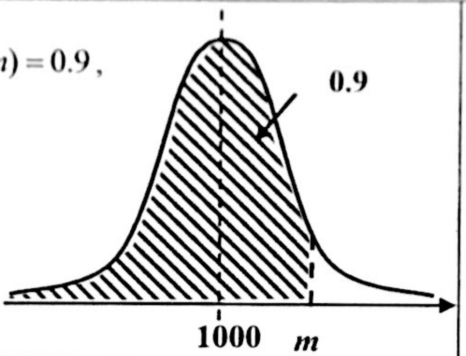
This time we need to find m such that the shaded area (refer to diagram) is 0.9.

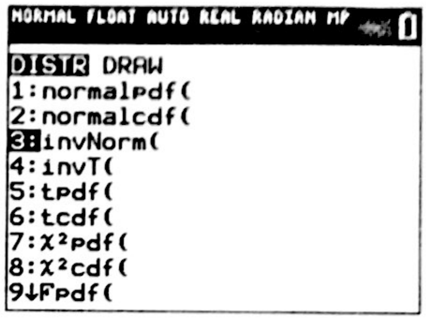
From GC, $m = 1064$ (nearest gram)

To find m such that $\mathrm{P}(X \le m) = 0.9$, given $X \sim \mathrm{N}(1000, 50^2)$:

1 Press [2nd][VARS]

Select **3: invNorm(**

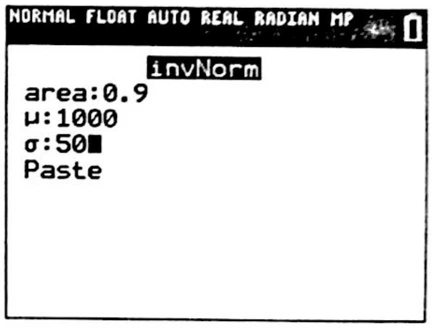


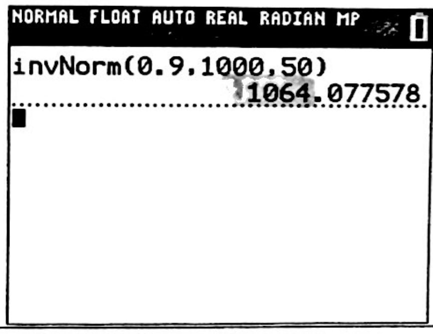


2 Enter 0.9 for area
1000 for μ
50 for σ

Highlight **Paste** and press [ENTER] twice

Note: For invNorm, the area is shaded from the left, i.e. lower tail.





Example 3 mean variance standard dev = ?

Given $X \sim \mathrm{N}(10,4)$, find, correct to 3 significant figures, the values of a and b such that

(i) $\mathrm{P}(X \geq a) = 0.23$, (ii) $\mathrm{P}(9 < X < b) = 0.64$.

Solution

$X \sim \mathrm{N}(10, 2^2)$

$\mu = 10$, $\sigma = 2$

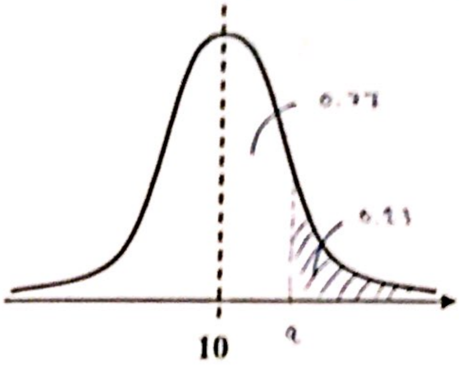
(i) $\mathrm{P}(X \geq a) = 0.23$

From the diagram,

$\mathrm{P}(X < a) = 1 - 0.23 = 0.77$

From GC, $\mathrm{P}(X < a) = 0.77$

$a = 11.5$ (3s.f.)




```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm
area:0.77
μ:10
σ:2
Paste
```

```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm(0.77,10,2)
                    11.47769371
```

(ii) $\mathrm{P}(9 < X < b) = 0.64$

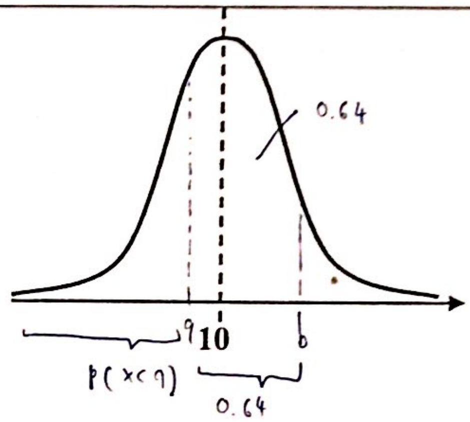
From the diagram,

$\mathrm{P}(X < b) = 0.64 + \mathrm{P}(X < 9)$

$= 0.94854$ (5s.f.)

From GC, $\mathrm{P}(X < 13.3) = 0.94854$

$\therefore b = 13.3$ (3s.f.)




```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm
area:0.94854
μ:10
σ:2
Paste
```

```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm(0.94854,10,2)
                    13.26171871
```

3 STANDARD NORMAL DISTRIBUTION

Let $X \sim \mathrm{N}(\mu, \sigma^2)$.

The random variable Z is defined by $Z = \dfrac{X-\mu}{\sigma}$.

Then $Z \sim \mathrm{N}(0,1)$.

Remarks:

- This process of converting $X \sim \mathrm{N}(\mu, \sigma^2)$ into $Z \sim \mathrm{N}(0,1)$ is known as **standardization.**

  It can be viewed as a re-scaling on the normal curve of X.

  It can be proven that $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$. (Refer to Appendix 4 for details)

- Z is known as the **standard normal random variable** and its distribution is called the **standard normal distribution.**

- $\boxed{\mathrm{P}(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right)}$

  use when μ and σ are unknown, GC cannot help

  $\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{Var}(X-\mu) = \frac{1}{\sigma^2}(\mathrm{Var}(X)$

In the past, we used to refer to the standard normal distribution table for calculations of probability involving the standard normal random variable. For non-standard normal random variables, we perform standardization first before looking up the table.

(Refer to Appendix 5 for more information on how to use the standard normal distribution table)

Now, with the availability of the graphing calculator, standardization is usually applied to solve problems where μ and/or σ are unknown (see example 4).

Example 4

The diameter of copper wires found in a certain type of equipment may be assumed to have a normal distribution with mean μ mm and standard deviation σ mm.

It was known that 99% of all wires produced have diameters exceeding 2.24 mm and 5% have diameters exceeding 2.30 mm.

Find the values of μ and σ.

Solution

Let X be the diameter of a copper wire in mm.
Then $X \sim N(\mu, \sigma^2)$.

$P(X > 2.24) = 0.99$

~~$P(X > 2.30) > 0.05$~~

$P(X \le 2.24) = 0.01$ → lower tail

Standardisation

$P\left(Z \le \frac{2.24-\mu}{\sigma}\right) = 0.01$

From GC, ||

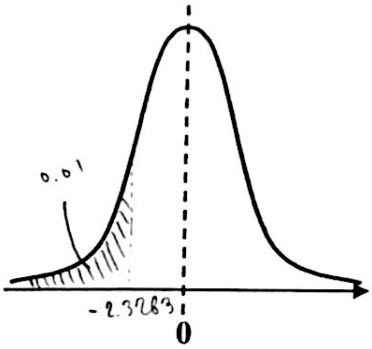
$P(Z \le -2.3263) = 0.01$

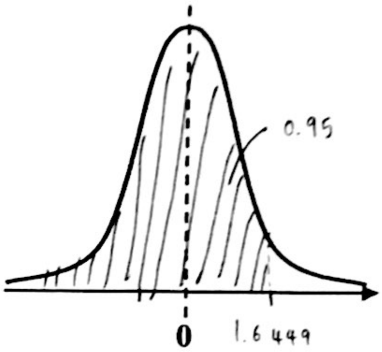
$\frac{2.24-\mu}{\sigma} = -2.3263$ (5 s.f.)

$2.3263\sigma - \mu = -2.24$ — ①

```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm
area:0.01
μ:0
σ:1
Paste
```

```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm(0.01,0,1)
                    -2.326347877
```





$P(X > 2.30) = 0.05$

$P(X \le 2.30) = 0.95$

$P\left(Z \le \frac{2.30-\mu}{\sigma}\right) = 0.95$

From GC,

$P(Z \le 1.6449) = 0.95$

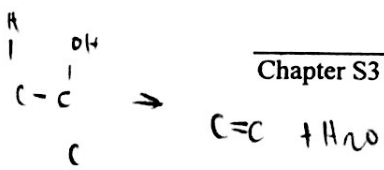
$\frac{2.30-\mu}{\sigma}$ 1.6449 (5 s.f.)

~~2.30~~ $1.6449\sigma + \mu = 2.30$ — ②

```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm
area:0.95
μ:0
σ:1
Paste
```

```
NORMAL FLOAT AUTO REAL RADIAN MP
invNorm(0.95,0,1)
                    1.644853626
```

Solving (1) and (2), $\mu = 2.28$ (3 s.f.) and $\sigma = 0.0151$ (3 s.f.).



4 LINEAR COMBINATIONS OF INDEPENDENT NORMAL RANDOM VARIABLES

4.1 Properties of Expectation and Variance of Random Variables

In Chapter S2A, we were introduced to the properties of expectation and variance for discrete random variables. The properties of expectation and variance for continuous random variables are similar and are as follows :

Let X and Y be random variables and a and b be constants.
We have

(1) $\mathrm{E}(a) = a$

(2) $\mathrm{E}(aX) = a\mathrm{E}(X)$

(3) $\mathrm{E}(aX \pm b) = a\mathrm{E}(X) \pm b$

(4) $\mathrm{E}(aX \pm bY) = a\mathrm{E}(X) \pm b\mathrm{E}(Y)$

Let X and Y be **independent** random variables and a and b be constants.
We have

(1) $\mathrm{Var}(a) = 0$

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

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

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

4.2 Properties of Independent Normal Random Variables

Let $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$ be **independent** random variables and a and b be constants. We have

(1) $X \pm Y \sim \mathrm{N}(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$

(2) $aX \pm b \sim \mathrm{N}(a\mu_1 \pm b, a^2\sigma_1^2)$

(3) $aX \pm bY \sim \mathrm{N}(a\mu_1 \pm b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

$E(aX \pm bY)$

$E(aX+bY) = aE(X)+bE(Y) = a\mu_1 + b\mu_2$

$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y)$
$= a^2\sigma_1^2 + b^2\sigma_2^2$

The mean and variance of the distributions can be shown using the properties in section 4.1.

Example 5

Let X and Y be 2 independent normal random variables. The means of X and Y are 10 and 12 respectively, and the standard deviations are 2 and 3 respectively.

$P(y-x<0)$ ← moves all variables outside.

Find **(a)** $P(X+Y\le 18)$, **(b)** $P(Y<X)$, **(c)** $P(4X+5Y>90)$, stating clearly the means and variances in each case.

Solution

Given $X \sim N(10, 2^2)$ and $Y \sim N(12, 3^2)$, where X and Y are independent.

(a) $E(X+Y) = E(x)+E(y) = 10+12 = 22$

$Var(X+Y) = Var(x)+Var(y) = 2^2+3^2 = 13$

$\therefore x+y \sim N(22, 13)$

normalcdf

$P(X+Y\le 18) = 0.134$ (3s.f.)

(b) $E(y-x) = E(y)-E(x) = 12-10 = 2$

$Var(y-x) = Var(y)+Var(x) = 4+9 = 13$

$\therefore y-x \sim (2, 13)$

$y-x \sim N(\quad)$ ← or calculate inside bracket

normal cdf

$P(Y<X) = P(y-x<0) = 0.290$ (3s.f.)

(c) $E(4X+5Y) = 4E(x)+5E(y) = 4(10)+5(12) = 100$

$Var(4X+5Y) = 4^2 Var(x)+5^2 Var(y) = 16(4)+25(9) = 289$

$\therefore 4x+5y \sim N(100, 289)$

$P(4X+5Y>90) = 0.722$ (3s.f.)

4.3 Random variable $X_1 + X_2$ versus random variable $2X$

Let $X \sim \mathrm{N}(\mu, \sigma^2)$.

We write X_1 and X_2 to denote 2 independent and identically distributed (i.i.d.) observations of X, i.e. $X_1 \sim \mathrm{N}(\mu, \sigma^2)$, $X_2 \sim \mathrm{N}(\mu, \sigma^2)$, X_1 and X_2 are independent.

We write $2X$ to denote twice the value of one observation of X.

Now, if X denotes the height of a Year 6 male student in RI, then

(i) X_1 may denote the height of a male student in 3A,

(ii) X_2 may denote the height of a male student in 3B and

(iii) $2X$ will denote twice the height of a Year 6 male student.

Is it therefore obvious that $X_1 + X_2$ is not the same as $2X$?

If not, let us consider the means and variances of $X_1 + X_2$ and $2X$:

	$X_1 + X_2$	$2X$
Mean	$\mathrm{E}(X_1 + X_2) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = \mu + \mu = 2\mu$	$\mathrm{E}(2X) = 2\mathrm{E}(X) = 2\mu$
Variance	$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 2\sigma^2$	$\mathrm{Var}(2X) = 4\mathrm{Var}(X) = 4\sigma^2$
Distribution	$X_1 + X_2 \sim \mathrm{N}(2\mu, 2\sigma^2)$	$\cancel{X}$ $2X \sim \mathrm{N}(2\mu, 4\sigma^2)$

Conclusion:

The variances of $X_1 + X_2$ and $2X$ are different. Thus, $X_1 + X_2$ is not the same as $2X$.

As such, extra care must be taken to distinguish between the random variable $2X$ and the random variable $X_1 + X_2$, where X_1 and X_2 are two independent observations of the random variable X.

Example 6

A factory produces sweets such that the mass of a packet of sweets is normally distributed with mean 60g and standard deviation 7g.

(i) If three packets of sweets are chosen at random, find the probability that the total mass exceeds 160g.

(ii) If one packet of sweet is chosen at random, find the probability that three times of its mass will exceed 160g.

Solution	**Remarks :**
Let X be the mass of a packet of sweets in grams. Then $X \sim \mathrm{N}(60, 7^2)$. let $Y = X_1 + X_2 + X_3$ $Y \sim N(3(60), 3(7^2))$ $Y \sim N(180, 147)$ $P(Y > 160) = 0.950$ ii) let $T = 3X$ $T \sim N(3(60), 3^2(7^2))$ $T \sim N(180, 441)$ $P(T > 160) = 0.830$	$E(Y) = E(X_1 + X_2 + X_3)$ $= E(X_1) + E(X_2) + E(X_3)$ $= 3E(X)$ $= 3(60)$ $= 180$ $Var(Y) = Var(X_1 + X_2 + X_3)$ $= Var(X) + Var(X) + Var(X)$ $= 3Var(X)$ $= 3(49)$ $= 147$ $E(T) = E(3X)$ $= 180$ $Var(T) = Var(3X)$ $= 3^2 VarX$ $= 9(49)$ $= 441$

Example 7

The thickness, X cm of a randomly chosen paperback may be assumed to follow a normal distribution with mean 3 and variance 0.67. The thickness, Y cm of a randomly chosen hardback book may also be assumed to follow a normal distribution with mean 5.9 and variance 1.87.

(i) Find the probability that the combined thickness of four randomly chosen paperbacks is greater than the combined thickness of two randomly chosen hardbacks.

(ii) Find the probability that a randomly chosen paperback is less than half as thick as a randomly chosen hardback.

(iii) Find the probability that the difference in thickness between a randomly chosen hardback and the combined thickness of two randomly chosen paperbacks is more than 1.5 cm.

Solution

Given $X \sim \mathrm{N}(3, 0.67)$ and $Y \sim \mathrm{N}(5.9, 1.87)$.

(i) let $W = (X_1+X_2+X_3+X_4) - (Y_1+Y_2)$

$W \sim N(0.2, 6.42)$

$P(W>0) = 0.531$ (3s.f.)

↳ normalcdf

(ii) let $V = X - \frac{1}{2}Y$

$V \sim N(0.05, 1.1375)$

$P(V<0) = 0.481$ (3s.f.)

(iii) $E(D) = E(Y) - 2E(X)$

$= 5.9 - 2(3)$

$= -0.1$

$Var(D) = Var(Y) + 2Var(X)$

$= 1.87 + 2(0.67)$

$= 3.21$

let $D = Y - (X_1+X_2)$

$\therefore D \sim N(-0.1, 3.21)$

$P(|D|>1.5) = P(D<-1.5) + P(D>1.5)$

$= 1 - P(-1.5<D<1.5)$

$= 0.403$ (3s.f.)

Remarks :

$E(W) = 4E(X) - 2E(Y)$

$= 4(3) - 2(5.9)$

$= 0.2$

$Var(W) = 4Var(X) + 2Var(Y)$

$= 4(0.67) + 2(1.87)$

$= 6.42$

$E(V) = E(X) - \frac{1}{2}E(Y)$

$= 3 - 0.5(5.9)$

$= 0.05$

$Var(V) = Var(X) + (\frac{1}{2})^2 Var Y$

$= 0.67 + 0.5^2(1.87)$

$= 1.1375$

$\mathrm{E}(D) = \mathrm{E}(Y) - 2\mathrm{E}(X)$

$= 5.9 - 2(3) = -0.1$

$\mathrm{Var}(D) = \mathrm{Var}(Y) + 2\mathrm{Var}(X)$

$= 1.87 + 2(0.67) = 3.21$

Example 8 [AJC8863Prelim/2009/Q10 part]

A fruit stall sells 3 types of durians: XO, Sultan and D24.
For each type of durian, the mass of a randomly chosen durian is normally distributed with mean and standard deviation as shown in the following table:

	Mean (kg)	Standard deviation (kg)	Price per kg
XO	2.5	0.8	\$7
Sultan	2.0	0.4	\$4
D24	2.0	0.5	\$4

Mr Lee picks two XO durians, and Mrs Lee picks one Sultan and three D24 durians at random.

(i) The probability that the total weight (in kg) of the four durians chosen by Mrs Lee lies in the range $(7.5, a)$ is 0.4. Find the value of a.

(ii) Find the probability that the total cost of the two XO durians that Mr Lee picks is more than the total cost of the four durians that Mrs Lee picks.

(iii) If the stallholder gives a discount of 50 cents for each durian, state the mean and the standard deviation of the total cost of the six durians after the discount.

Solution

Let X, S and D be the mass of a XO, Sultan and D24 durian in kg respectively.
Then $X \sim \mathrm{N}(2.5, 0.8^2)$, $S \sim \mathrm{N}(2, 0.4^2)$ and $D \sim \mathrm{N}(2, 0.5^2)$.

(i) Let $Y = S + D_1 + D_2 + D_3$

$\therefore\ Y \sim \mathrm{N}(8, 0.91)$

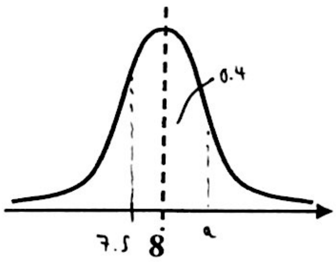
$P(7.5 < Y < a) = 0.4$

From the diagram,

$P(Y < a) = 0.4 + P(Y < 7.5)$

$= 0.70009$

$a = 8.50$ (3 s.f.)



(ii) let $T = 7(X_1 + X_2) - 4Y$

$T \sim N(3, 77.28)$

(iii) Let V be the total cost, in dollars, of the 6 durians after discount.

$V = 7(X_1 + X_2) + 4Y - 6(0.5)$

$E(V) = 7(2)E(X) + 4E(Y) - 3$

$= (14(2.5) + 4(8) - 3) = 64$

$\therefore$ Mean of total cost of the 6 durians after discount = \$64

$Var(V) = 7^2(2)Var(X) + 4^2(Var(Y))$

$= 98(0.8^2) + 16(0.91)$

$= 77.28$

Standard deviation of total cost of the 6 durians after discount

$= \sqrt{77.28} = \$8.79$

Remarks :

to use inv norm, use the lower tail

$E(Y) = E(S) + 3E(D)$
$= 2 + 3(2) = 8$

$Var(Y) = Var(S) + 3Var(D)$
$= 0.4^2 + 3(0.5^2) = 0.91$

$E(T) = E(7(X_1 + X_2) - 4(Y))$
$= 7(2E(X_1)) - 4E(Y))$
$= 7(2)(2.5) - 4(8)$
$= 3$

$Var(T) = Var(7(X_1 + X_2) - 4Y)$
$= 7^2[Var(X_1 + X_2)] + 4^2 Var(Y)$
$= 7^2[2Var(X)] + 4^2 Var(Y)$
$= 7^2(2)(0.8^2) + 4^2(0.91)$
$= 77.28$

Example 9 [NYJC9740Prelim/2007/02/Q11 (part)]

A soft drink dispenser delivers lemonade into a cup when a coin is inserted into the machine. The amount of lemonade delivered is normally distributed with mean 260 ml and standard deviation 10 ml. The nominal amount of lemonade in a cup is 250 ml and the capacity of the cup is 275 ml.

(i) What is the probability that the cup overflows?

(ii) On an occasion, five such cups of lemonade were purchased. Find the probability that not more than one cup contains less than 250 ml.

(iii) Some customers have complained that there is a high proportion of cups with less than the nominal amount of lemonade. The standard deviation of the amount of lemonade delivered per cup is fixed, but the mean can be altered. Find the least value of the new mean such that not more than 5% of the cups will contain less than 250 ml of lemonade.

Solution

Let X be the amount of lemonade delivered in a cup, in ml.
Then $X \sim \mathrm{N}(260, 10^2)$.

(i) P(cup overflows) $= \mathrm{P}(X > 275) = 0.0668$ (3 s.f.)

(ii) $\mathrm{P}(X < 250) = 0.15866$ (5 s.f.)

Let Y be the number of cups that contain less than 250 ml of lemonade each, out of 5.

$Y \sim B(5, 0.15866)$

$P(Y \leq 1) = 0.819$ (3 s.f.)

(iii) Let μ be the value of the new mean.
Then $X \sim \mathrm{N}(\mu, 10^2)$

We need $P(X < 250) \leq 0.05$

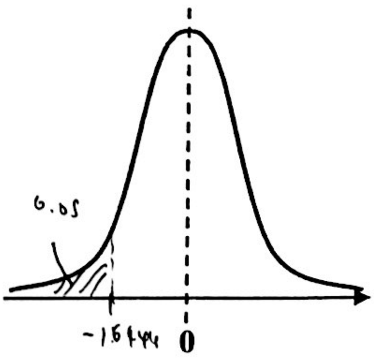
$P\left(Z < \frac{250-\mu}{10}\right) \leq 0.05$

From GC, $\mathrm{P}(Z < -1.6449) = 0.05$
From the diagram,

$\frac{250-\mu}{10} \leq -1.6449$

$\mu \geq 266.45$

The least value of the new mean is 267 ml (3 s.f.).



Example 10 [RI 9740 Prelim/2012/02/Q9]

Red and green grapes are sold in boxes by weight. The masses, in grams, of a box of red grapes and a box of green grapes are modeled as having independent normal distributions with means and standard deviations as shown in the table.

	Mean Mass	Standard Deviation
Box of red grapes	300	40
Box of green grapes	150	20

(i) A fruit-seller packs and weighs each box of red grapes. Find the probability that the 10th box of red grapes he weighs is the second box that is at least 310g.

(ii) Find the probability that the mass of 3 boxes of red grapes differs from 3 times the mass of a box of green grapes by at least 500 g.

(iii) Red grapes are sold at \$5 per kg and green grapes are sold at \$6 per kg. Find the probability that 10 boxes of red grapes cost more than 10 boxes of green grapes by at most \$7.

Solution

Let R and G be the weight of a box of red grapes and green grapes in grams respectively.
Then $R \sim \mathrm{N}(300, 40^2)$ and $G \sim \mathrm{N}(150, 20^2)$.

(i) Let X be the number of boxes of red grapes which weighs at least 310g, out of 9 boxes.
$X \sim \mathrm{B}(9, \mathrm{P}(R \geq 310))$
i.e. $X \sim \mathrm{B}(9, 0.40129)$

Required probability $= \mathrm{P}(X = 1)\,\mathrm{P}(R \geq 310) = 0.0239$ (3 s.f.)

(ii) Let $T = R_1 + R_2 + R_3 - 3G$
$T \sim \mathrm{N}(3(300)-3(150),\ 3(40^2)+3^2(20^2))$
$T \sim \mathrm{N}(450, 8400)$

$$\begin{aligned}\mathrm{P}(|T| \geq 500) &= \mathrm{P}(T \leq -500) + \mathrm{P}(T \geq 500) \\ &= 1 - \mathrm{P}(-500 < T < 500) \\ &= 0.293 \text{ (3 s.f.)}\end{aligned}$$

(iii) Let $W = 0.005(R_1 + \ldots + R_{10}) - 0.006(G_1 + \ldots + G_{10})$
$\mathrm{E}(W) = 0.005(10)\mathrm{E}(R) - 0.006(10)\mathrm{E}(G) = 6$

$\mathrm{Var}(W) = (0.005)^2(10)\mathrm{Var}(R) + (0.006)^2(10)\mathrm{Var}(G) = 0.544$
$W \sim \mathrm{N}(6, 0.544)$

Required probability $= \mathrm{P}(0 < W \leq 7) = 0.912$ (3 s.f.)

CONCLUSION

The normal distribution was introduced by the French mathematician Abraham De Moivre in 1733. He used this distribution to approximate probabilities connected with coin tossing, and called it the exponential bell-shaped curve. Its usefulness, however, became truly apparently only in 1809, when the famous German mathematician K. F. Gauss used it as an integral part of his approach to predict the location of astronomical entities. As a result, it became common after this time to call it the Gaussian distribution.

During the mid to late nineteenth century, however, most statisticians started to believe that the majority of data sets would have histograms conforming to the Gaussian bell-shaped form. Indeed, it came to be accepted that it was "normal" for any well-behaved data set to follow this curve. As a result, following the lead of the British statistician Karl Pearson, people began referring to the Gaussian curve by calling it simply the normal curve.

The normal distribution is used to model a wide range of real-world situations. Its applications are highly prevalent in pure sciences as well as in social sciences. You should be very familiar with its properties and be highly comfortable in solving related questions, as it is going to play a huge part in subsequent topics in Statistics.

APPENDIX 1

Probability Density Function of Continuous Random Variables

A continuous random variable X is defined by its probability density function which can be represented by a curve $y = \mathrm{f}(x)$. The probabilities are given by the area under the curve.

Properties of the probability density function:

(1) $\mathrm{f}(x) \geq 0$ for all values of x

(2) $\int_{-\infty}^{\infty} \mathrm{f}(x)\,\mathrm{d}x = 1$ **(total probability equals one)**

(3) $\mathrm{P}(a < X < b) = \mathrm{P}(a \leq X \leq b) = \mathrm{P}(a < X \leq b) = \mathrm{P}(a \leq X < b) = \int_{a}^{b} \mathrm{f}(x)\,\mathrm{d}x$

Remarks:

- $\mathrm{f}(a)$ is **not** a probability; it is not the probability at $x = a$.

- A continuous random variable has a probability of **zero** of assuming exactly any of its values. That is
  (a) $\mathrm{P}(X = x) = 0$ for all values of x, and
  (b) the probability of randomly selecting X to be exactly x (and not one of the infinitely many real numbers so close to x that you cannot humanly measure the difference) is extremely remote.

  Hence **(3)** follows.

Expectation and Variance of Continuous Random Variables

Let X be a continuous random variable.

Then

- the expectation of X is given by
$$\mathrm{E}(X) = \mu = \int_{-\infty}^{\infty} x\,\mathrm{f}(x)\,\mathrm{d}x$$

- the variance of X is given by
$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}\left[(X-\mu)^2\right] \\ &= \int_{-\infty}^{\infty} (x-\mu)^2\,\mathrm{f}(x)\,\mathrm{d}x\end{aligned}$$

  However, for most computational purposes, we prefer to use the equivalent form

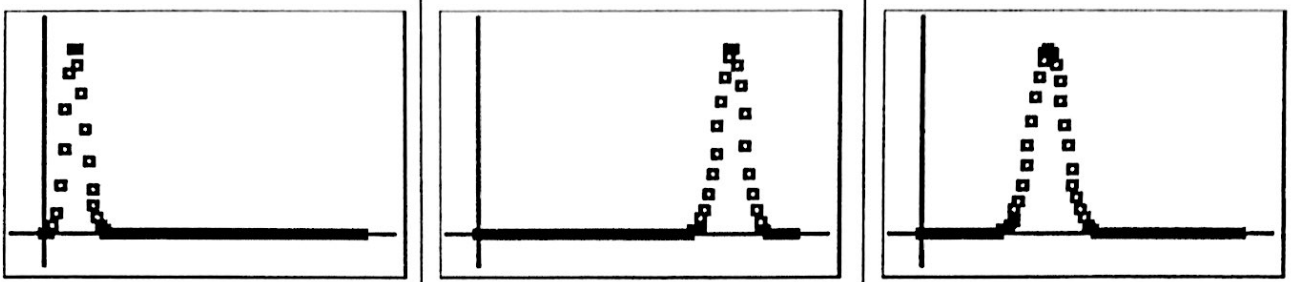
$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \left[\mathrm{E}(X)\right]^2, \text{ where } \mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2\,\mathrm{f}(x)\,\mathrm{d}x.$$

APPENDIX 2

Approximating a Binomial Distribution using a Normal Distribution

We may approximate a binomial distribution using a normal distribution when the number of trials n is large and the probability of success p is not too close to 0 or 1.

The closer p is to 0.5, the less skewed the binomial distribution, and the better the approximation using a normal distribution (which is symmetrical).



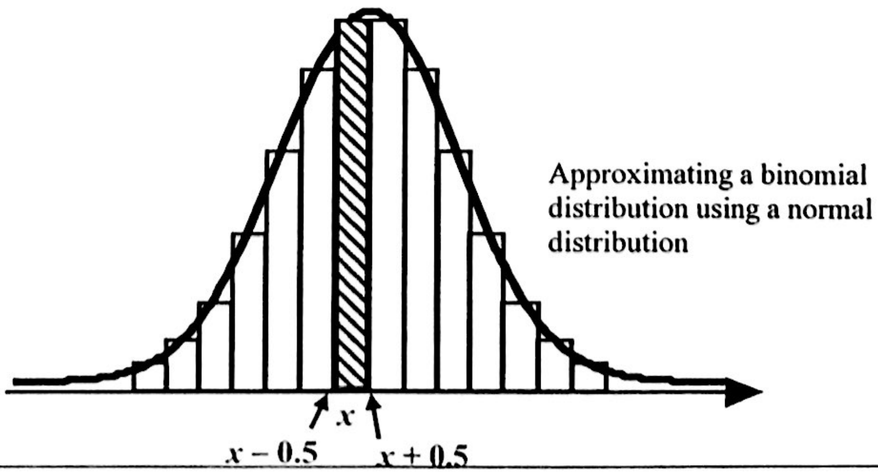
$n=100,\ p=0.1$	$n=100,\ p=0.8$	$n=100,\ p=0.4$

How large must n be?
Generally we require

n to be sufficiently large such that $np>5$ and $n(1-p)>5$.

Note however that a binomial random variable is discrete while a normal random variable is continuous. Hence a **continuity correction** must be applied.

We need to state "by continuity correction" when it is being applied, and it is always helpful to draw a diagram.



Let $X \sim \mathrm{B}(n,p)$.
If n is large such that $np>5$ and $n(1-p)>5$, then

$$X \sim \mathrm{N}(\ np,\ np(1-p)\ ) \text{ approximately.}$$

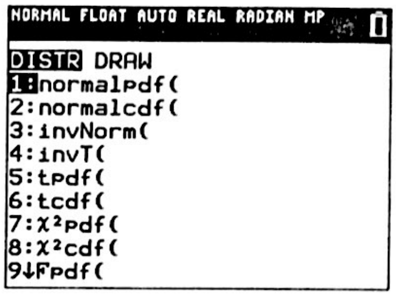
$\mathrm{P}(X=x)=\mathrm{P}(x-0.5<X<x+0.5)$ (by continuity correction)

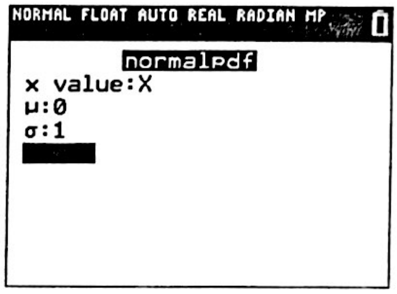
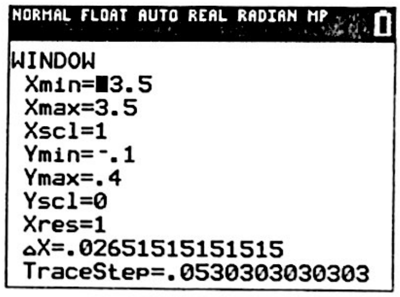
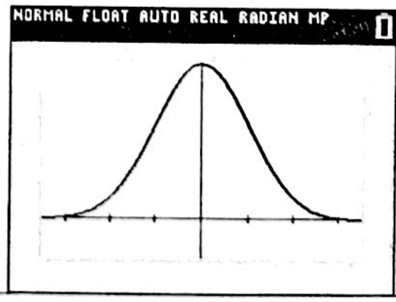
APPENDIX 3

Use of GC to sketch Normal Curves

<table>
<tr><td>1</td><td>Press [Y=]

Press [2nd][VARS] for [DISTR]

Select 1: normalpdf(</td><td>
</td></tr>
<tr><td>2</td><td>Enter X for x value
Leave μ as 0 and σ as 1

Highlight Paste and press [ENTER] twice</td><td>
</td></tr>
<tr><td>3</td><td>Press [WINDOW] and change the settings with the specifications as shown</td><td>
</td></tr>
<tr><td>4</td><td>Press [GRAPH] and to get the normal curve N(0,1)</td><td>
</td></tr>
</table>

Investigation

Sketch the normal curves N(1,1), N(2,1) and N(3,1) on the same diagram to see how the shape changes with different means.

Sketch the normal curves N(1,1), N(1,1.5) and N(1,2) on the same diagram to see how the shape changes with different standard deviations. (You may need to adjust the window settings)

APPENDIX 4

Mean and Standard Deviation of Standard Normal Random Variable

Let $X \sim \mathrm{N}(\mu, \sigma^2)$.

For $Z = \dfrac{X-\mu}{\sigma}$, $\mathrm{E}(Z) = 0$ and $\mathrm{Var}(Z) = 1$.

Proof: (using the properties of mean and variance in section 4.1)

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

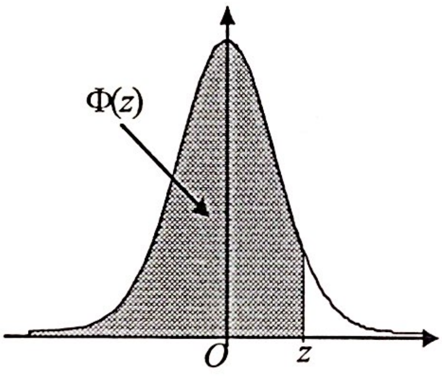
APPENDIX 5

Standard Normal Distribution Table

A table (List MF26 page 7) of values is provided for calculation of probability involving the standard normal random variable.

The following notation is used in the table:

$$\Phi(z) = \mathrm{P}(Z \le z).$$



The standard normal distribution table used by Cambridge gives the values of $\Phi(z)$ for values of z between 0.000 and 2.999 in step size of 0.001.

E.g. To find $\Phi(1.234)$, we look up the table as shown:
$\Phi(1.234) = 0.8907 + 0.0007 = 0.8914.$

z	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
															ADD				
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	4	8	12	16	20	24	28	32	36
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	4	8	12	16	20	24	28	32	36
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	4	8	12	15	19	23	27	31	35
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	7	11	15	19	22	26	30	34
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	4	7	11	14	18	22	25	29	32
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	3	7	10	14	17	20	24	27	31
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	3	7	10	13	16	19	23	26	29
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	3	6	9	12	15	18	21	24	27
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	3	5	8	11	14	16	19	22	25
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	3	5	8	10	13	15	18	20	23
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	2	5	7	9	12	14	16	19	21
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	2	4	6	8	10	12	14	16	18
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	2	4	6	7	9	11	13	15	17
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	2	3	5	6	8	10	11	13	14
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	1	3	4	6	7	8	10	11	13

Some Useful Properties:
(as table in MF26 gives values for positive values of z only)

- $\Phi(z) = \mathrm{P}(Z \le z)$
- $\mathrm{P}(Z > z) = 1 - \Phi(z)$
- $\Phi(-z) = 1 - \Phi(z)$ (by symmetry of the standard normal curve about $z = 0$)
- $\mathrm{P}(|Z| \le z) = 2\Phi(z) - 1$ (by symmetry of the standard normal curve about $z = 0$)
- $\mathrm{P}(|Z| > z) = 2[1 - \Phi(z)]$ (by symmetry of the standard normal curve about $z = 0$)

Summary

******THE END******