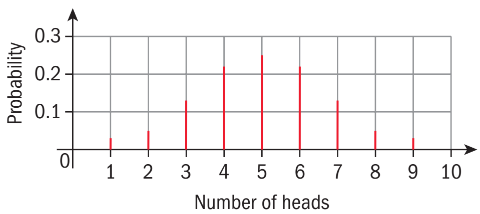
15 Normal distribution as an approximation to a binomial distribution

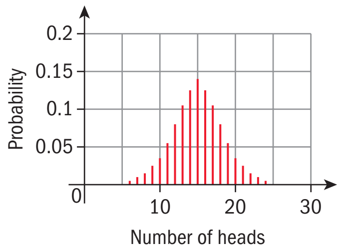
Try this worksheet after you have completed Exercise 15L.

Binomial distribution

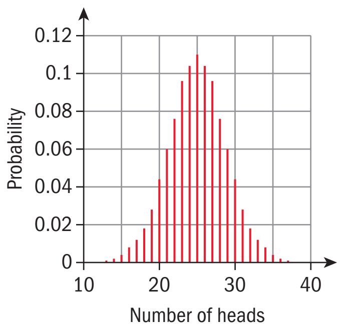
If you throw a coin a number of times and record the number of heads you obtain the results can be modeled using a binomial distribution with $p = 0.5$ and n = number of trials.
These graphs show the distribution for various values of n.



$n = 10$



$n = 30$

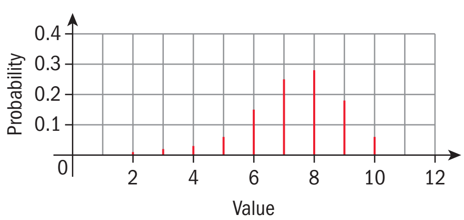


$n = 40$

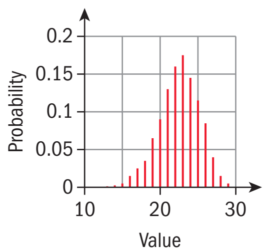
As the number of trials increases the distribution takes on the characteristic bell-shape of a normal distribution. When $p = 0.5$ this will occur quite quickly as the binomial distribution is symmetrical for this value.

When $p \neq 0.5$ it can also be seen as the number of trials increases the distribution will also tend to the bell-shaped normal distribution.

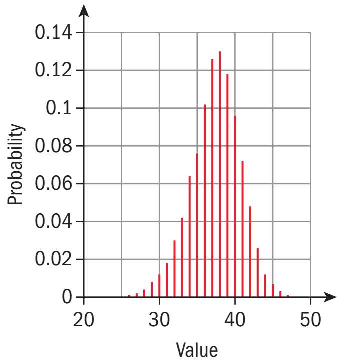
For $p = 0.75$, $n = 10$, 30 and 50 the graphs are:



$n = 10$



$n = 30$



$n = 50$

In some cases you can use the normal distribution as an approximation to the binomial distribution and this makes the calculation of probabilities much easier. The normal distribution is a continuous distribution whereas the binomial distribution is a discrete distribution.

→ When you use the normal distribution as an approximation to the binomial distribution you have to introduce a **continuity correction**.

For the coin throwing example, if you want to calculate the probability that at most 6 heads are obtained you calculate $\mathrm{P}(X \le 6)$.

In a continuous distribution $X \le 6$ will equate to $X < 6.5$

Similarly $X < 6$ will equate to $X < 5.5$

EXAMPLE 1

It is known that a particular variety of flower seed has a probability of 0.9 of successfully producing a flower. 300 of these seeds are planted.
Find the probability that the number of flowers produced will be

a less than 275

b between 280 and 290 inclusive.

Answers

a If X is the random variable 'the number of seeds that produce flowers' then

$X \sim \mathrm{B}(300, 0.9)$

Define the binomial distribution.

Use the normal distribution with random variable W to represent the approximate distribution of variable X then since $W \sim \mathrm{N}(\mu, \sigma^2)$ then here $W \sim \mathrm{N}(np, npq)$

Define the related normal distribution
$\mu = np = 300 \times 0.9$
and $\sigma^2 = npq$ *(where* $q = 1 - p$*)*
$= 300 \times 0.9 \times 0.1$

So $W \sim \mathrm{N}(270, 27)$ and $\mathrm{P}(X < 275)$ transforms to $\mathrm{P}(W < 274.5)$

$\mathrm{P}(W < 274.5) = 0.807$

A number between 274.5 and 275 rounds up to 275 so to find $X < 275$ *the continuity correction is* $W < 274.5$

Using GDC:
normcdf$(-1000, 274.5, 270, \sqrt{27})$

b $\mathrm{P}(280 \le X \le 290)$ transforms to $\mathrm{P}(279.5 < W < 290.5)$

$\mathrm{P}(279.5 < W < 290.5) = 0.0337$

If you use the binomial distribution these answers are

a 0.805

b 0.0287

These answers are very close to the approximate values using the normal distribution.

Exercise

1 It is given that $X \sim \mathrm{B}(25,0.6)$.
Using a normal approximation, determine $\mathrm{P}(X \le 16)$.

2 A fair tetrahedral dice is thrown 100 times. What is the probability of it landing on 3 between 23 and 30 times inclusively?

3 It is known that in a sack of mixed grass seed 35% of seeds are ryegrass. Use the normal approximation to the binomial distribution to find the probability that in a sample of 400 seeds there are

a fewer than 120 ryegrass seeds
b between 120 and 150 ryegrass seeds (inclusive)
c more than 160 ryegrass seeds.

4 A fair coin is tossed 1000 times. The number of heads obtained is denoted by X. Using a normal approximation, determine the largest value of x for which $\mathrm{P}(X \le x) < 0.95$.

Chapter 15 extension worked solutions

1 $X \sim B(25, 0.6)$, $np = 15$, $npq = 6$
$X \sim B(25, 0.6) \rightarrow W \sim N(15, 6)$
$P(X \le 16) \rightarrow P(W < 16.5) = 0.730$ (3 sf)

2 Let X be the random variable of the number of times the dice lands on a 3.
So $X \sim B(100, 0.25)$, $np = 25$, $npq = 18.75$
$X \sim B(100, 0.25) \rightarrow W \sim N(25, 18.75)$
$P(23 \le X \le 30) \rightarrow P(22.5 < W < 30.5) = 0.616$ (3 sf)

3 Let X be the random variable of the number of ryegrass seeds.
So $X \sim B(400, 0.35)$, $np = 140$, $npq = 91$
$X \sim B(400, 0.35) \rightarrow W \sim N(140, 91)$
a $P(X \le 120) \rightarrow P(W < 119.5) = 0.0158$ (3 sf)
b $P(120 \le X \le 150) \rightarrow P(119.5 < W < 150.5) = 0.849$ (3 sf)
c $P(X > 160) \rightarrow P(W > 160.5) = 0.0158$ (3 sf)

4 $X \sim B(1000, 0.5)$, $np = 500$, $npq = 250$
$X \sim B(1000, 0.5) \rightarrow W \sim N(500, 250)$

$$P(X \le x) \rightarrow P\left(W < \frac{\left(x + \frac{1}{2}\right) - 500}{\sqrt{250}}\right) = 0.95$$

Now $P(Z < 1.645) = 0.95$

So $\dfrac{\left(x + \frac{1}{2}\right) - 500}{\sqrt{250}} = 1.645$

Rearranging gives $499.5 + (1.645 \times \sqrt{250}) = 525.51$
Therefore the largest value of x satisfying $P(X \le x) < 0.95$ is 525