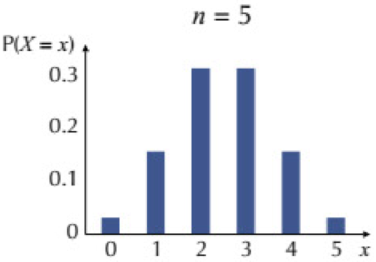
Normal Approximation to B(n, p)

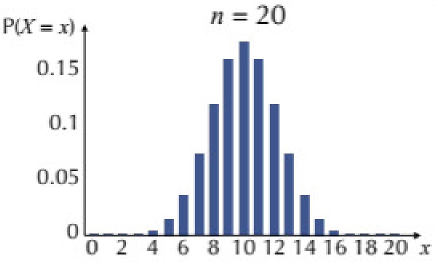
If n is big, a binomial distribution (B(n, p)) can be tricky to work with. However, in this situation, you can often approximate a binomial distribution with a normal distribution. Another great Maths life hack (sort of).

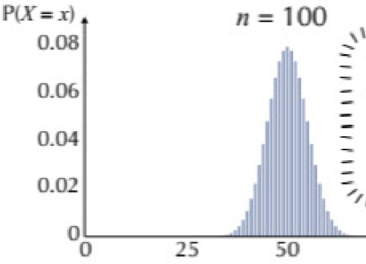
*You can **Approximate** a **Binomial Distribution** with a **Normal Distribution***

The number of times a fair coin lands on heads when it is tossed n times is a classic **binomial random variable — call it** X, so $X \sim B(n, 0.5)$. The graphs below show the **probability distribution** (see p.158) of X, for a few different values of n:

You can find these probabilities using the formula from p.162.







These probabilities aren't actually zero — the bars are just too small to see on this scale.

As n gets **bigger**, this starts to look more and more like a **normal distribution**.
So, for large n, you can use a normally distributed variable to **approximate** a binomial one.
For the approximation to work well, you need the following conditions to be true:

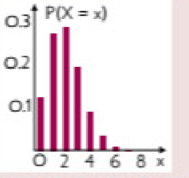
Normal Approximation to the Binomial

Suppose the random variable X follows a binomial distribution, i.e. $X \sim B(n, p)$.

If (i) $p \approx 0.5$,
and (ii) n is large,

then X can be approximated by the normal distribution $Y \sim N(np, npq)$ (where $q = 1 - p$).

If p **isn't** close to 0.5, then the binomial distribution will be **skewed** to one side.
E.g. here's the graph of $X \sim B(20, 0.1)$. This is **not** a good fit for the normal distribution, which is always **symmetrical**.



Example: If $X \sim B(800, 0.4)$, use a suitable approximation to find:
a) $P(X < 350)$, b) $P(X \geq 300)$, c) $P(320 < X \leq 350)$.

You need to make sure first that the normal approximation is **suitable**:

n is **large**, and p is **not far** from 0.5, so the normal approximation is valid.

Next, work out np and npq: $np = 800 \times 0.4 = 320$ and
$npq = 800 \times 0.4 \times (1 - 0.4) = 192$

So the approximation you need is: $Y \sim N(320, 192)$. ← The standard deviation is $\sqrt{192}$.

Now use your **calculator** to find the probabilities.

a) $P(X < 350) \approx P(Y < 350) = \mathbf{0.985}$ (3 s.f.)

b) $P(X \geq 300) \approx P(Y \geq 300) = \mathbf{0.926}$ (3 s.f.)

c) $P(320 < X \leq 350) \approx P(320 < Y \leq 350) = \mathbf{0.485}$ (3 s.f.)

If your calculator can only work out probabilities for the standard normal distribution, you'll need to convert the values of X to values of Z first, using the method on p.166.

Example: Each piglet born on a farm is equally likely to be male or female. 250 piglets are born. Use a suitable normal approximation to estimate the probability that there will be more than 130 male piglets born.

$n = 250$, and p (the probability that the piglet is male) is 0.5.
So if X represents the number of male piglets born, then $X \sim B(250, 0.5)$.

Since n is large and p is 0.5, a normal approximation is appropriate
— X can be approximated by a normal random variable $Y \sim N(\mu, \sigma^2)$:

$\mu = np = 250 \times 0.5 = 125$ and $\sigma^2 = npq = 250 \times 0.5 \times 0.5 = 62.5$

So the approximation you need is: $Y \sim N(125, 62.5)$.

So, using your approximation, $P(X > 130) \approx P(Y > 130)$.

From your calculator, $P(Y > 130) = \mathbf{0.264}$ (3 s.f.)

The Piggins family didn't take kindly to being called "approximately normal".

Normal Approximation to B(n, p)

The approximation also works as long as np and nq are Bigger Than 5

Even if p **isn't** all that close to 0.5, the normal approximation usually works fine as long as np and nq are both **bigger than 5**.

Example: a) On average, only 23% of robin chicks survive to adulthood. If 200 robin chicks are randomly selected, use a suitable approximation to find the probability that at least 25% of them survive to adulthood.

b) A student takes a sample of 11 robin chicks from her garden. Explain whether or not a normal approximation would still be appropriate to estimate the probability of at least 25% of them surviving to adulthood.

a) If X represents the number of survivors, then $X \sim B(200, 0.23)$.

Here, p **isn't** particularly close to 0.5, but n is **large**, so calculate np and nq:

$np = 200 \times 0.23 = \mathbf{46}$ and $nq = 200 \times (1 - 0.23) = \mathbf{154}$.

Both np and nq are greater than 5, so a **normal approximation** should be okay to use — $Y \sim N(46, 35.42)$.

Variance = npq
= 200 × 0.23 × 0.77
= 35.42

25% of 200 = 50, so you want to find $P(X \geq 50) \approx P(Y \geq 50)$.

Using your calculator, $P(Y \geq 50) = \mathbf{0.251}$ (3 s.f.).

Using the original binomial distribution gives an answer of 0.27497... for part a), so this is a reasonable approximation.

b) This time, $X \sim B(11, 0.23)$, which means $np = 11 \times 0.23 = \mathbf{2.53}$, which is **less than 5**.

n is small and p isn't close to 0.5, so a normal approximation is **not appropriate**.

Practice Questions

Q1 The random variable X follows a binomial distribution: $X \sim B(100, 0.45)$.
Using a normal approximation, find:
a) $P(X > 50)$, b) $P(X \leq 42)$, c) $P(40 < X \leq 47)$.

Q2 The number of people working in the post office is constantly adjusted depending on how busy it is, with the result that the probability of any person being served within 1 minute can be modelled as having a constant value of 0.7. Exactly 200 people come to the post office on a particular day. Using a normal approximation to a binomial distribution, estimate the probability that fewer than 60% of them are served within a minute.

Exam Questions

Q1 The random variable X is binomially distributed with $X \sim B(100, 0.6)$.

a) State the conditions needed for X to be well approximated by a normal distribution. [2 marks]

b) Using a suitable approximation, find:
(i) $P(X \geq 65)$ [2 marks]
(ii) $P(50 < X < 62)$ [1 mark]

Q2 On average, 55% of customers in a cafe order a cup of coffee. One day, the cafe served 150 customers. At the start of the day, the manager calculated that they had enough coffee to make 75 cups of coffee. By using a suitable approximation, find the probability that more then 75 customers try to order coffee. [3 marks]

Q3 The random variable X follows a binomial distribution: $X \sim B(n, p)$. X is approximated by the normally distributed random variable Y. Using this normal approximation, $P(X \leq 153) = 0.9332$ and $P(X > 127) = 0.9977$ (4 d.p.).

a) Find the mean and standard deviation of the normal approximation. [6 marks]

b) Hence, estimate n and p. [4 marks]

Admit it — the normal distribution is the most amazing thing ever...

So the normal approximation can work pretty well, even when p isn't really all that close to 0.5. But even so, you should always show that your approximation is 'suitable'. In fact, the question might even ask you to show it.