

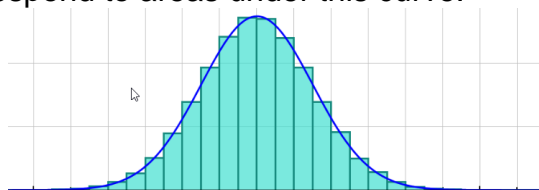
## Probability distributions

<b>N2</b>	Understand and use the Normal distribution as a model; find probabilities using the Normal distribution Link to histograms, mean, standard deviation, points of inflection and the binomial distribution
<b>N3</b>	Select an appropriate probability distribution for a context, with appropriate reasoning, including recognising when the binomial or Normal model may not be appropriate

### Commentary

The world is a lot more complex than we typically think. We are used to discrete, constant, ideas like the colour red, or an animal like a rabbit being fixed concepts. However the colour red is part of a continuous spectrum of colour, and no two rabbits are identical and modern rabbits radically different from their distant evolutionary ancestors. The same is true of data which is on a continuous scale, such as time – it makes no sense to talk about exact values as they cannot even be measured exactly. Usain Bolt ran the 100 metres in Berlin in 2009 in a time measured as 9.58 seconds but we will never know the exact time because we are limited by the accuracy of the measuring instrument – the best we can say is that the exact time lies in an interval from 9.575 to 9.585 seconds.

Students will appreciate that many real-world continuous data sets are likely to produce symmetrical unimodal histograms, such as weights of babies. In these situations it is reasonable to use the Normal distribution curve to estimate proportions which correspond to areas under this curve.



The key quantity for any calculation is  $z$ , the number of standard deviations from the mean. The curve has inflection points when  $z = \pm 1$ . Up to a transformation of variables, all normal curves are the same.

The equation of the Normal distribution curve is a transformation of  $y = e^{-x^2}$ . This curve cannot be integrated using standard methods; calculators use accurate numerical methods to determine areas (and therefore probabilities).

It is possible to use the normal distribution  $N(np, np(1 - p))$  to estimate probabilities from the binomial distribution  $B(n, p)$  as long as  $n$  is large and  $p$  is close to  $\frac{1}{2}$ . In Autograph you can show  $B(1000, 0.5)$  and  $N(500, 250)$  on the same axes – the effect of this approximation is negligible in this instance.

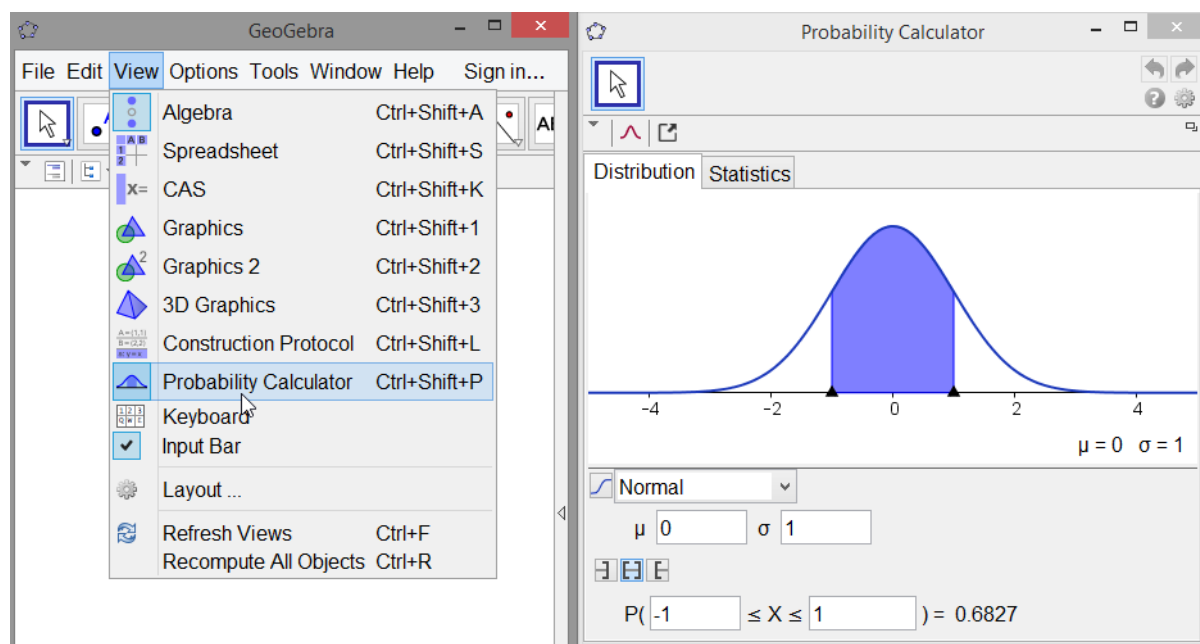
## Sample MEI resource

'Normal curves' (which can be found at <http://integralmaths.org/sow-resources.php>) is designed to help students make strong visual links between probabilities, areas under normal curves, and the cumulative normal probability function.

$2 - 2\Phi(2)$	$2\Phi(1) - 1$	$1 - \Phi(2)$
$\Phi(1) - \frac{1}{2}$	$\Phi(1) - \frac{1}{2}$	$2\Phi(2) - 1$
$\Phi(2) - \frac{1}{2}$	$\frac{3}{2} - \Phi(2)$	$\Phi(2) - \Phi(1)$
$\Phi(2.5) - \Phi(2)$	$\Phi(2.5) - \Phi(0.5)$	$1 + \Phi(0.5) - \Phi(2.5)$

## Effective use of technology

'Probability Calculator' (which can be found at <http://www.mei.org.uk/integrating-technology>) is a built in feature in the free software GeoGebra. It is possible to easily change the parameters to visualise normal probabilities. The best thing is that the curve never changes shape; instead it is the axes which adjust.



## Probability distributions

Time allocation:

### Pre-requisites

- GCSE/AS: Histograms and cumulative frequency
- AS: The Binomial Distribution
- 

### Links with other topics

- Hypothesis Testing: In the Hypothesis testing unit we will be able to question assertions about the population mean  $\mu$ .
- Differentiation: Finding points of inflection using calculus.
- 

### Questions and prompts for mathematical thinking

- Make up three questions that show you understand how the Normal Distribution can be used in context.
- Change one number in the following statement to make it true:  
If  $X \sim N(100, 15^2)$  then  $P(X < 90) = 0.1$
- 

### Applications and modelling

- How do the proportions of the normal distribution correspond to measurement of outliers, such as the  $2xSD$  rule or  $1.5xIQR$ ?
- A teacher sets a test and plans to give grades  $\{A, B, C, D, E, U\}$  based on a normal curve with the mean and standard deviation of the students' results. Suggest how the teacher should construct the grade boundaries. Is this a fair way to grade a test?
- 

### Common errors

- Using the height of the normal curve ( $Npd$ ) to find  $P(X = a)$ .
- Not dealing clearly with Normal data that produces negative z-scores e.g. 5% were less than, say, 10.76 so  $z = 1.96$  rather than  $-1.96$ .
- Confusing probabilities with z-scores.
-