EXAM PAPERS PRACTICE

Boost your performance and confidence with these topic-based exam questions

Practice questions created by actual examiners and assessment experts

Detailed mark scheme

Suitable for all boards

Designed to test your ability and thoroughly prepare you

4.6 Normal Distribution

IB Maths - Revision Notes

AA HL

4.6.1 The Normal Distribution

Properties of Normal Distribution

The binomial distribution is an example of a discrete probability distribution. The normal distribution is an example of a **continuous** probability distribution.

What is a continuous random variable?

- A continuous random variable (often abbreviated to CRV) is a random variable that can take **any value** within a range of infinite values
  - Continuous random variables **usually measure** something
  - For example, height, weight, time, etc

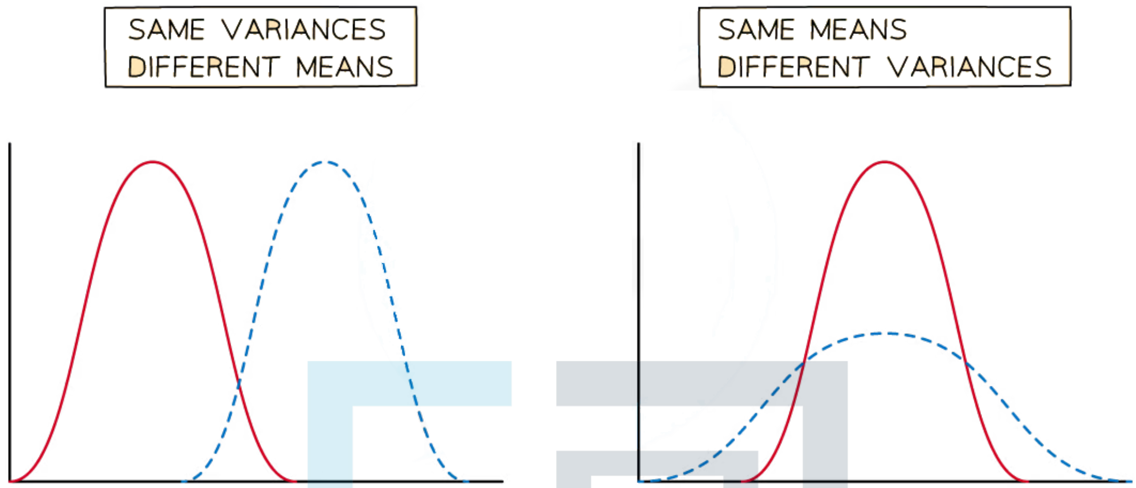
What is a continuous probability distribution?

- A continuous probability distribution is a probability distribution in which the random variable X is continuous
- The probability of X being a **particular value is always zero**
  - $\mathrm{P}(X=k)=0$ for any value k
  - Instead we define the **probability density function** $\mathrm{f}(x)$ for a specific value
    - This is a function that describes the **relative likelihood** that the random variable would be close to that value
  - We talk about the **probability** of X being within a **certain range**
- A continuous probability distribution can be represented by a continuous graph (the values for X along the horizontal axis and probability **density** on the vertical axis)
- The **area under the graph** between the points $x=a$ and $x=b$ is equal to $\mathrm{P}(a \leq X \leq b)$
  - The **total area under the graph equals 1**
- As $\mathrm{P}(X=k)=0$ for any value k, it does not matter if we use strict or weak inequalities
  - $\mathrm{P}(X \leq k)=\mathrm{P}(X<k)$ for any value k when X is a **continuous random variable**

What is a normal distribution?

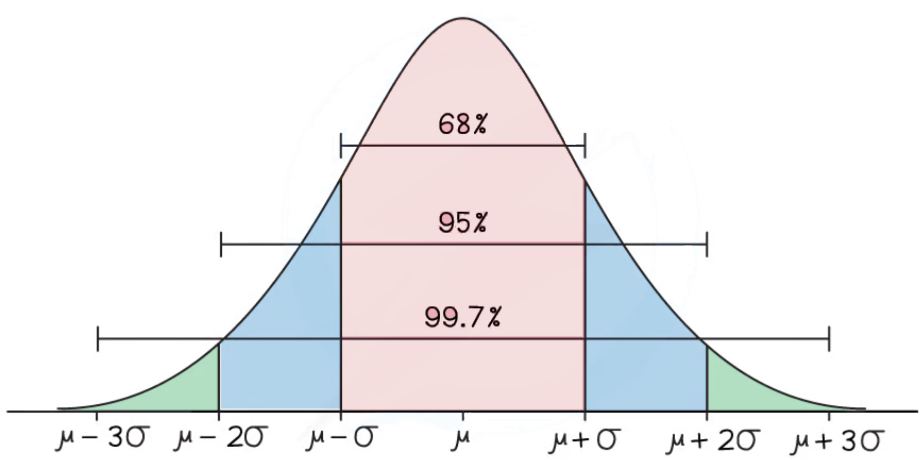
- A normal distribution is a **continuous probability distribution**
- The **continuous random variable** X can follow a normal distribution if:
  - The distribution is **symmetrical**
  - The distribution is **bell-shaped**
- If X follows a normal distribution then it is denoted $X \sim \mathrm{N}(\mu, \sigma^2)$
  - μ is the **mean**
  - σ^2 is the **variance**

- σ is the **standard deviation**
- If the **mean** changes then the graph is **translated horizontally**
- If the **variance** increases then the graph is **widened horizontally** and **made taller vertically** to maintain the same area
  - A **small variance** leads to a **tall** curve with a **narrow** centre
  - A **large variance** leads to a **short** curve with a **wide** centre



What are the important properties of a normal distribution?

- The **mean** is μ
- The **variance** is σ^2
  - If you need the **standard deviation** remember to square root this
- The normal distribution is symmetrical about
  - Mean = Median = Mode = μ
- There are the results:
  - Approximately **two-thirds (68%)** of the data lies within **one standard deviation** of the mean ($\mu \pm \sigma$)
  - Approximately **95%** of the data lies within **two standard deviations** of the mean ($\mu \pm 2\sigma$)
  - Nearly **all of the data (99.7%)** lies within **three standard deviations** of the mean ($\mu \pm 3\sigma$)



Modelling with Normal Distribution

What can be modelled using a normal distribution?

- A lot of real-life continuous variables can be modelled by a normal distribution provided that the population is large enough and that the variable is **symmetrical** with **one mode**
- For a normal distribution X can take any real value, however values far from the mean (more than 4 standard deviations away from the mean) have a probability density of **practically zero**
    - This fact allows us to model variables that are not defined for all real values such as height and weight

What can not be modelled using a normal distribution?

- Variables which have **more than one mode** or **no mode**
    - For example: the number given by a random number generator
- Variables which are **not symmetrical**
    - For example: how long a human lives for

Exam Tip

- An exam question might involve different types of distributions so make it clear which distribution is being used for each variable

Worked example

The random variable S represents the speeds (mph) of a certain species of cheetahs when they run. The variable is modelled using $\mathrm{N}(40,\ 100)$.

a) Write down the mean and standard deviation of the running speeds of cheetahs.

$\mu = 40$ and $\sigma^2 = 100$

Square root to get standard deviation

Mean $\mu = 40$
Standard deviation $\sigma = 10$

b) State two assumptions that have been made in order to use this model.

We assume that the distribution of the speeds is
- symmetrical
- bell-shaped

4.6.2 Calculations with Normal Distribution

Calculating Normal Probabilities

Throughout this section we will use the random variable $X \sim \mathrm{N}(\mu, \sigma^2)$. For X distributed normally, X can take any real number. Therefore any values mentioned in this section will be assumed to be real numbers.

How do I find probabilities using a normal distribution?

- The **area under a normal curve** between the points $x = a$ and $x = b$ is equal to the **probability** $\mathrm{P}(a < X < b)$
  - Remember for a normal distribution you do not need to worry about whether the inequality is strict (< or >) or weak (≤ or ≥)
    - $\mathrm{P}(a < X < b) = \mathrm{P}(a \le X \le b)$
- You will be **expected to use** distribution functions on your **GDC** to find the probabilities when working with a normal distribution

How do I calculate P(*X* = *x*): the probability of a single value for a normal distribution?

- The probability of a **single value** is **always zero** for a normal distribution
  - You can picture this as the area of a single line is zero
- $\mathrm{P}(X = x) = 0$
- Your GDC is likely to have a "**Normal Probability Density**" function
  - This is sometimes shortened to NPD, Normal PD or Normal Pdf
  - **IGNORE THIS FUNCTION** for this course!
  - This calculates the **probability density function** at a point **NOT the probability**

How do I calculate P(*a* < *X* < *b*): the probability of a range of values for a normal distribution?

- You need a **GDC** that can calculate **cumulative normal probabilities**
- You want to use the "**Normal Cumulative Distribution**" function
  - This is sometimes shortened to NCD, Normal CD or Normal Cdf
- You will need to enter:
  - The 'lower bound' - this is the value *a*
  - The 'upper bound' - this is the value *b*
  - The 'μ' value - this is the mean
  - The 'σ' value - this is the standard deviation
- **Check the order carefully** as some calculators ask for standard deviation before mean
  - Remember it is the standard deviation

- so if you have the **variance** then **square root it**
- **Always sketch** a quick diagram to visualise which area you are looking for

How do I calculate P(*X* > *a*) or P(*X* < *b*) for a normal distribution?

- You will still use the "**Normal Cumulative Distribution"** function
- $\mathrm{P}(X > a)$ can be estimated using an **upper bound that is sufficiently bigger** than the **mean**
  - Using a value that is more than 4 standard deviations **bigger than the mean** is quite accurate
  - Or an easier option is just to input lots of 9's for the upper bound (**99999999...** or **10^{99}**)
- $\mathrm{P}(X < b)$ can be estimated using a **lower bound that is sufficiently smaller** than the **mean**
  - Using a value that is more than 4 standard deviations **smaller than the mean** is quite accurate
  - Or an easier option is just to input lots of 9's for the lower bound with a negative sign (**-99999999...** or **-10^{99}**)

Are there any useful identities?

- $\mathrm{P}(X < \mu) = \mathrm{P}(X > \mu) = 0.5$
- As $\mathrm{P}(X = a) = 0$ you can use:
  - $\mathrm{P}(X < a) + \mathrm{P}(X > a) = 1$
  - $\mathrm{P}(X > a) = 1 - \mathrm{P}(X < a)$
  - $\mathrm{P}(a < X < b) = \mathrm{P}(X < b) - \mathrm{P}(X < a)$
- These are useful when:
  - The mean and/or standard deviation are unknown
  - You only have a diagram
  - You are working with the **inverse distribution**

Exam Tip

- Check carefully whether you have entered the standard deviation or variance into your GDC

Worked example

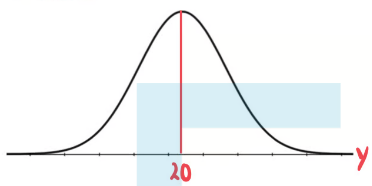
The random variable $Y \sim \mathrm{N}(20, 5^2)$. Calculate:

i) $\mathrm{P}(Y = 20)$.

Identify μ and σ

$\mu = 20 \quad \sigma^2 = 5^2$ so $\sigma = 5$

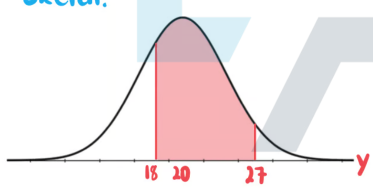
Sketch!



$\boxed{P(Y=20)=0}$

ii) $\mathrm{P}(18 \le Y < 27)$.

Sketch!



Using GDC

Lower $= 18$

Upper $= 27$

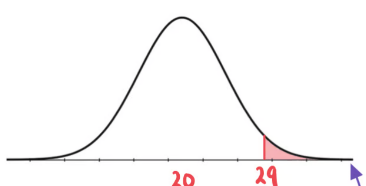
We can use $\le$ or $<$

$P(18 < Y < 27) = 0.574665...$

$\boxed{0.575 \text{ (3sf)}}$

iii) $\mathrm{P}(Y > 29)$

Sketch!



Using GDC

Lower $= 29$

Upper $= 99999$

$P(Y > 29) = 0.035930...$

$\boxed{0.0359 \text{ (3sf)}}$

No upper bound so choose a big number

Inverse Normal Distribution

Given the value of P(*X* < *a*) how do I find the value of *a*?

- Your **GDC** will have a function called "**Inverse Normal Distribution**"
  - Some calculators call this InvN
- Given that $\mathrm{P}(X < a) = p$ you will need to enter:
  - The 'area' - this is the value p
    - Some calculators might ask for the 'tail' - this is the left tail as you know the area to the left of a
  - The 'μ' value - this is the mean
  - The 'σ' value - this is the standard deviation

Given the value of P(*X* > *a*) how do I find the value of *a*?

- If your calculator **does** have the **tail option** (left, right or centre) then you can use the "Inverse Normal Distribution" function straightaway by:
  - Selecting 'right' for the tail
  - Entering the area as 'p'
- If your calculator **does not** have the **tail option** (left, right or centre) then:
  - Given $\mathrm{P}(X > a) = p$
  - Use $\mathrm{P}(X < a) = 1 - \mathrm{P}(X > a)$ to rewrite this as
    - $\mathrm{P}(X < a) = 1 - p$
  - Then use the **method for P(*X* < *a*)** to find a

Exam Tip

- Always check your **answer makes sense**
  - If P($X < a$) is **less than 0.5** then a should be **smaller than the mean**
  - If P($X < a$) is **more than 0.5** then a should be **bigger than the mean**
  - A sketch will help you see this

Worked example

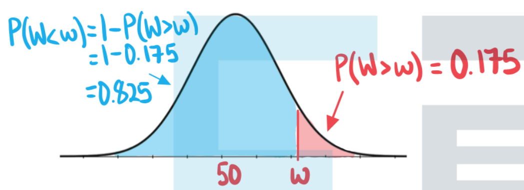
The random variable $W \sim \mathrm{N}(50,\ 36)$.

Find the value of w such that $\mathrm{P}(W > w) = 0.175$.

Identify μ and σ

$\mu = 50 \quad \sigma^2 = 36$ so $\sigma = 6$

Sketch!



$\mathrm{P}(W > w)$ is less than 0.5 so w is bigger than the mean

Area from left is 0.825

Use Inverse Normal Distribution function on GDC

$w = 55.6075...$

$\boxed{w = 55.6 \text{ (3sf)}}$

4.6.3 Standardisation of Normal Variables

Standard Normal Distribution

What is the standard normal distribution?

- The **standard normal distribution** is a normal distribution where the **mean is 0** and the **standard deviation is 1**
  - It is denoted by Z
  - $Z \sim \mathrm{N}(0,\ 1^2)$

Why is the standard normal distribution important?

- Any **normal distribution curve** can be transformed to the standard normal distribution curve by a **horizontal translation** and a **horizontal stretch**
- Therefore we have the relationship:
  - $Z = \dfrac{X - \mu}{\sigma}$
  - Where $X \sim \mathrm{N}(\mu,\ \sigma^2)$ and $Z \sim \mathrm{N}(0,\ 1^2)$
- Probabilities are related by:
  - $\mathrm{P}(a < X < b) = \mathrm{P}\left(\dfrac{a-\mu}{\sigma} < Z < \dfrac{b-\mu}{\sigma}\right)$
  - This will be useful when the mean or variance is unknown
- Some mathematicians use the function $\Phi(z)$ to represent $\mathrm{P}(Z < z)$

z-values

What are *z*-values (standardised values)?

- For a normal distribution $X \sim \mathrm{N}(\mu,\ \sigma^2)$ the *z*-value (standardised value) of an *x*-value tells you how many standard deviations it is away from the mean
  - If $z = 1$ then that means the *x*-value is 1 standard deviation bigger than the mean
  - If $z = -1$ then that means the *x*-value is 1 standard deviation smaller than the mean
- If the *x*-value is **more than the mean** then its corresponding *z*-value will be **positive**
- If the *x*-value is **less than the mean** then its corresponding *z*-value will be **negative**
- The *z*-value can be calculated using the formula:
  - $z = \dfrac{x - \mu}{\sigma}$
  - This is given in the **formula booklet**
- *z*-values can be used to compare values from different distributions

Exam Papers Practice

Finding Sigma and Mu

How do I find the mean (μ) or the standard deviation (σ) if one of them is unknown?

- If the **mean** or **standard deviation** of $X \sim N(\mu, \sigma^2)$ is **unknown** then you will need to use the **standard normal distribution**
- You will need to use the formula
  - $z = \frac{x-\mu}{\sigma}$ or its rearranged form $x = \mu + \sigma z$
- You will be given a **probability for a specific value** of
  - $P(X < x) = p$ or $P(X > x) = p$
- To find the unknown parameter:
- **STEP 1**: **Sketch** the normal curve
  - Label the known value and the mean
- **STEP 2**: **Find** the ***z*-value** for the given value of ***x***
  - Use the **Inverse Normal Distribution** to find the value of z such that $P(Z < z) = p$ or $P(Z > z) = p$
  - Make sure the direction of the inequality for Z is consistent with the inequality for X
  - Try to **use lots of decimal places** for the *z*-value or **store your answer** to **avoid rounding errors**
    - You should use at least one extra decimal place within your working than your intended degree of accuracy for your answer
- **STEP 3**: **Substitute** the known values into $z = \frac{x-\mu}{\sigma}$ or $x = \mu + \sigma z$
  - You will be given and one of the parameters (μ or σ) in the question
  - You will have calculated z in STEP 2
- **STEP 4**: **Solve** the equation

How do I find the mean (μ) and the standard deviation (σ) if both of them are unknown?

- If **both** of them are **unknown** then you will be given two probabilities for two specific values of ***x***
- The process is the same as above
  - You will now be able to **calculate two *z*-values**
  - You can form **two equations** (rearranging to the form $x = \mu + \sigma z$ is helpful)
  - You now have to **solve the two equations simultaneously** (you can use your calculator to do this)
  - Be careful not to mix up which z-value goes with which value of *x*

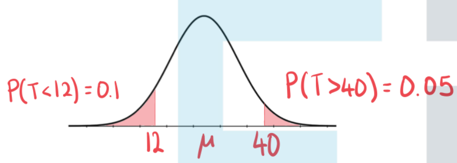
Worked example

It is known that the times, in minutes, taken by students at a school to eat their lunch can be modelled using a normal distribution with mean μ minutes and standard deviation σ minutes.

Given that 10% of students at the school take less than 12 minutes to eat their lunch and 5% of the students take more than 40 minutes to eat their lunch, find the mean and standard deviation of the time taken by the students at the school.

Let $T \sim N(\mu, \sigma^2)$ be the time taken to eat lunch

STEP 1
Sketch the information



STEP 2
Find the corresponding z-values using inverse normal on GDC

$Z \sim N(0, 1^2)$

$P(Z < z_1) = 0.1 \Rightarrow z_1 = -1.2815...$

$P(Z > z_2) = 0.05 \Rightarrow P(Z < z_2) = 0.95 \Rightarrow z_2 = 1.6448...$

STEP 3
Form equations using $z = \frac{x - \mu}{\sigma}$ or $x = \mu + \sigma z$

$12 = \mu - (1.2815...)\sigma$

$40 = \mu + (1.6448...)\sigma$

STEP 4
Solve equations using GDC

$\mu = 24.26...$ $\sigma = 9.568...$

Mean = 24.3 mins (3sf)
Standard deviation = 9.57 mins (3sf)