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 Random Variables

IB Maths - Revision Notes

4.6.1 Linear Combinations of Random Variables

Transformation of a Single Variable

What is Var(X)?

- Var(X) represents the variance of the random variable X
- Var(X) can be calculated by the formula
  - $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$
    - where $\mathrm{E}(X^2) = \sum x^2 \mathrm{P}(X = x)$
  - You will **not be required** to use this formula in the exam

What are the formulae for E($aX \pm b$) and Var($aX \pm b$)?

- If a and b are constants then the following formulae are true:
  - **E($aX \pm b$) = aE(X) ± b**
  - **Var($aX \pm b$) = a^2 Var(X)**
    - These are given in the **formula booklet**
- This is the same as linear transformations of data
  - The mean is affected by multiplication and addition/subtraction
  - The variance is affected by multiplication but not addition/subtraction
- Remember division can be written as a multiplication
  - $\frac{X}{a} = \frac{1}{a}X$

Worked example

X is a random variable such that $\mathrm{E}(X) = 5$ and $\mathrm{Var}(X) = 4$.

Find the value of:

(i) $\mathrm{E}(3X + 5)$

(ii) $\mathrm{Var}(3X + 5)$

(iii) $\mathrm{Var}(2 - X)$.

Formula booklet

Linear transformation of a single random variable	$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$ $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$

$E(3X + 5) = 3E(X) + 5 = 3(5) + 5$ $\boxed{E(3X + 5) = 20}$

$Var(3X + 5) = 3^2 Var(X) = 9(4)$ $\boxed{Var(3X + 5) = 36}$

$Var(2 - X) = (-1)^2 Var(X) = 1(4)$ $\boxed{Var(2 - X) = 4}$

Transformation of Multiple Variables

What is the mean and variance of $aX + bY$?

- Let X and Y be two random variables and let a and b be two constants
- $E(aX + bY) = aE(X) + bE(Y)$
  - This is true for **any random variables** X and Y
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$
  - This is true if X and Y are **independent**
- $E(aX - bY) = aE(X) - bE(Y)$
- $Var(aX - bY) = a^2 Var(X) + b^2 Var(Y)$
  - Notice that you still add the two terms together on the right hand side
    - This is because b^2 is positive even if b is negative
  - Therefore the variances of $aX + bY$ and $aX - bY$ are the same

What is the mean and variance of a linear combination of *n* random variables?

- Let $X_1, X_2, ..., X_n$ be n random variables and $a_1, a_2, ..., a_n$ be n constants

$$\mathrm{E}\left(a_1X_1 \pm a_2X_2 \pm \ ... \pm a_nX_n\right) = a_1\mathrm{E}\left(X_1\right) \pm a_2\mathrm{E}\left(X_2\right) \pm \ ... \pm a_n\mathrm{E}\left(X_n\right)$$

  - This is given in the **formula booklet**
  - This can be written as $\mathrm{E}\left(\sum a_iX_i\right) = \sum a_i\mathrm{E}\left(X_i\right)$
  - This is true for **any random variable**

$$\mathrm{Var}\left(a_1X_1 \pm a_2X_2 \pm \ ... \pm a_nX_n\right) = a_1^{\,2}\mathrm{Var}\left(X_1\right) + a_2^{\,2}\mathrm{Var}\left(X_2\right) + \ ... + a_n^{\,2}\mathrm{Var}\left(X_n\right)$$

  - This is given in the **formula booklet**
  - This can be written as $\mathrm{Var}\left(\sum a_iX_i\right) = \sum a_i^2\mathrm{Var}\left(X_i\right)$
  - This is true if the random variables are **independent**
    - Notice that the constants get squared so the terms on the right-hand side will always be positive

For a given random variable *X*, what is the difference between 2*X* and $X_1 + X_2$?

- $2X$ means **one observation** of X is taken and **then doubled**
- $X_1 + X_2$ means **two observations** of X are taken and then **added together**
- $2X$ and $X_1 + X_2$ have the **same expected values**
  - $E(2X) = 2E(X)$
  - $E(X_1 + X_2) = E(X_1) + E(X_2) = 2E(X)$
- $2X$ and $X_1 + X_2$ have **different variances**
  - $Var(2X) = 2^2 Var(X) = 4Var(X)$
  - $Var(X_1 + X_2) = Var(X_1) + Var(X_2) = 2Var(X)$

- To see the distinction:
  - Suppose X could take the values 0 and 1
    - $2X$ could then take the values 0 and 2
    - X_1+X_2 could then take the values 0, 1 and 2
- Questions are likely to describe the variables in context
  - For example: The mass of a carton containing 6 eggs is the mass of the carton plus the mass of the 6 **individual** eggs
  - This can be modelled by $M = C + E_1 + E_2 + E_3 + E_4 + E_5 + E_6$ where
    - C is the mass of a carton
    - E is the mass of an egg
  - It is **not** $C + 6E$ because the masses of the 6 eggs could be **different**

Exam Tip

- In an exam when dealing with multiple variables ask yourself which of the two cases is true
  - You are adding together **different observations** using the same variable: $X_1 + X_2 + \ldots + X_n$
  - You are taking a **single observation** of a variable and multiplying it by a constant: nX

Worked example

X and Y are independent random variables such that

$$\mathrm{E}(X) = 5 \text{ \& } \mathrm{Var}(X) = 3,$$

$$\mathrm{E}(Y) = -2 \text{ \& } \mathrm{Var}(Y) = 4.$$

Find the value of:

(i) $\mathrm{E}(2X+5Y)$,

(ii) $\mathrm{Var}(2X+5Y)$,

(iii) $\mathrm{Var}(4X-Y)$.

Formula booklet

Linear combinations of n independent random variables, $X_1, X_2, \ldots, X_n$	$\mathrm{E}(a_1X_1 \pm a_2X_2 \pm \ldots \pm a_nX_n) = a_1\mathrm{E}(X_1) \pm a_2\mathrm{E}(X_2) \pm \ldots \pm a_n\mathrm{E}(X_n)$ $\mathrm{Var}(a_1X_1 \pm a_2X_2 \pm \ldots \pm a_nX_n) = a_1^2\,\mathrm{Var}(X_1) + a_2^2\,\mathrm{Var}(X_2) + \ldots + a_n^2\,\mathrm{Var}(X_n)$

$\mathrm{E}(2X+5Y) = 2\mathrm{E}(X) + 5\mathrm{E}(Y) = 2(5) + 5(-2)$ $\boxed{\mathrm{E}(2X+5Y) = 0}$

$\mathrm{Var}(2X+5Y) = 2^2\mathrm{Var}(X) + 5^2\mathrm{Var}(Y) = 4(3) + 25(4)$ $\boxed{\mathrm{Var}(2X+5Y) = 112}$

$\mathrm{Var}(4X-Y) = 4^2\mathrm{Var}(X) + \mathrm{Var}(Y) = 16(3) + 4$ $\boxed{\mathrm{Var}(4X-Y) = 52}$

4.6.2 Unbiased Estimates

Unbiased Estimates

What is an unbiased estimator of a population parameter?

- An **estimator** is a **random variable** that is used to **estimate a population parameter**
  - An **estimate** is the value produced by the estimator when a sample is used
- An estimator is called unbiased if its expected value is equal to the population parameter
  - An estimate from an unbiased estimator is called an **unbiased estimate**
  - This means that the **mean** of the **unbiased estimates** will get **closer** to the **population parameter** as **more samples** are taken
- The **sample mean** is an **unbiased estimate** for the **population mean**
- The **sample variance** is **not an unbiased estimate** for the **population variance**
  - On average the sample variance will **underestimate** the population variance
  - As the **sample size increases** the sample variance gets **closer to the unbiased estimate**

What are the formulae for unbiased estimates of the mean and variance of a population?

- A sample of n data values ($x_1, x_2, \ldots$ etc) can be used to find unbiased estimates for the mean and variance of the population
- An unbiased estimate for the mean μ of a population can be calculated using
  - $\bar{x} = \frac{\sum x}{n}$
- An unbiased estimate for the variance σ^2 of a population can be calculated using
  - $s_{n-1}^2 = \frac{n}{n-1} s_n^2$
  - This is given in the **formula booklet**
  - s_n^2 is the variance of the sample data
    - $s_n^2 = \frac{\sum (x-\bar{x})^2}{n} = \frac{\sum x^2}{n} - (\bar{x})^2$
- Different calculators can use different notations for s_{n-1}^2
  - σ_{n-1}^2, s^2, $\hat{s}^2$ are notations you might see
  - You may also see the square roots of these

Is s_{n-1} an unbiased estimate for the standard deviation?

- Unfortunately s_{n-1} is not an unbiased estimate for the standard deviation of the population
- It is better to work with the unbiased variance rather than standard deviation
- There is not a formula for an unbiased estimate for the standard deviation that works for all populations
  - Therefore you will not be asked to find one in your exam

How do I show the sample mean is an unbiased estimate for the population mean?

- You **do not need to learn this proof**
  - It is simply here to help with your understanding
- Suppose the population of X has mean μ and variance σ^2
- Take a sample of n observations
  - $X_1, X_2, ..., X_n$
  - $E(X_i) = \mu$
- Using the formula for a linear combination of n independent variables:

$$\begin{aligned}
E(\overline{X}) &= E\left(\frac{X_1 + X_2 + \ ... + X_n}{n}\right) \\
&= \frac{E(X_1) + E(X_2) + \ ... + E(X_n)}{n} \\
&= \frac{\mu + \mu + \ ... + \mu}{n} \\
&= \frac{n\mu}{n} \\
&= \mu
\end{aligned}$$

- As $E(\overline{X}) = \mu$ this shows the formula will produce an **unbiased estimate** for the population mean

Why is there a divisor of n–1 in the unbiased estimate for the variance?

- You **do not need to learn this proof**
  - It is simply here to help with your understanding
- Suppose the population of X has mean μ and variance σ^2
- Take a sample of n observations
  - $X_1, X_2, ..., X_n$
  - $E(X_i) = \mu$
  - $\text{Var}(X_i) = \sigma^2$
- Using the formula for a linear combination of n independent variables:

$$\operatorname{Var}(\overline{X}) = \operatorname{Var}\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)$$
$$= \frac{\operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \ldots + \operatorname{Var}(X_n)}{n^2}$$
$$= \frac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n^2}$$
$$= \frac{n\sigma^2}{n^2}$$
$$= \frac{\sigma^2}{n}$$

- It can be shown that $\mathrm{E}(\overline{X}^2) = \mu^2 + \dfrac{\sigma^2}{n}$
  - This comes from rearranging $\operatorname{Var}(\overline{X}) = \mathrm{E}(\overline{X}^2) - [\mathrm{E}(\overline{X})]^2$
- It can be shown that $\mathrm{E}(X^2) = \mathrm{E}(X_i^{\ 2}) = \mu^2 + \sigma^2$
  - This comes from rearranging $\operatorname{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$
- Using the formula for a linear combination of n independent variables:

$$\mathrm{E}(S_n^2) = \mathrm{E}\left(\frac{\sum X_i^2}{n} - \overline{X}^2\right)$$
$$= \frac{\sum \mathrm{E}(X_i^2)}{n} - \mathrm{E}(\overline{X}^2)$$
$$= \frac{\sum(\mu^2 + \sigma^2)}{n} - \left(\mu^2 + \frac{\sigma^2}{n}\right)$$
$$= \frac{n(\mu^2 + \sigma^2)}{n} - \left(\mu^2 + \frac{\sigma^2}{n}\right)$$
$$= \mu^2 + \sigma^2 - \left(\mu^2 + \frac{\sigma^2}{n}\right)$$
$$= \sigma^2 - \frac{\sigma^2}{n}$$
$$= \frac{n\sigma^2 - \sigma^2}{n}$$
$$= \frac{n-1}{n}\sigma^2$$

- As $E\left(S_n^2\right) \neq \sigma^2$ this shows that the sample variance is not unbiased
  - You need to multiply by $\dfrac{n}{n-1}$
  - $E\left(S_{n-1}^2\right) = \sigma^2$

Exam Tip

- Check the wording of the exam question carefully to determine which of the following you are given:
  - The **population variance**: σ^2
  - The **sample variance**: s_n^2
  - An **unbiased estimate** for the **population variance**: s_{n-1}^2

Worked example

The times, X minutes, spent on daily revision of a random sample of 50 IB students from the UK are summarised as follows.

$$n = 50 \qquad \sum x = 6174 \qquad s_n^2 = 1384.3$$

Calculate unbiased estimates of the population mean and variance of the times spent on daily revision by IB students in the UK.

Unbiased estimate of population mean $\bar{x} = \dfrac{\Sigma x}{n}$

$\bar{x} = \dfrac{6174}{50} = 123.48$

$\boxed{\bar{x} = 123 \text{ minutes (3sf)}}$

Formula booklet

Unbiased estimate of population variance s_{n-1}^2	$s_{n-1}^2 = \dfrac{n}{n-1} s_n^2$

$s_{n-1}^2 = \dfrac{50}{49} \times 1384.3 = 1412.55\ldots$

$\boxed{s_{n-1}^2 = 1410 \text{ minutes}^2 \text{ (3sf)}}$