DP IB Maths: AA HL

Your notes

4.4 Probability Distributions

Contents

4.4.1 Discrete Probability Distributions

Your notes

Discrete Probability Distributions

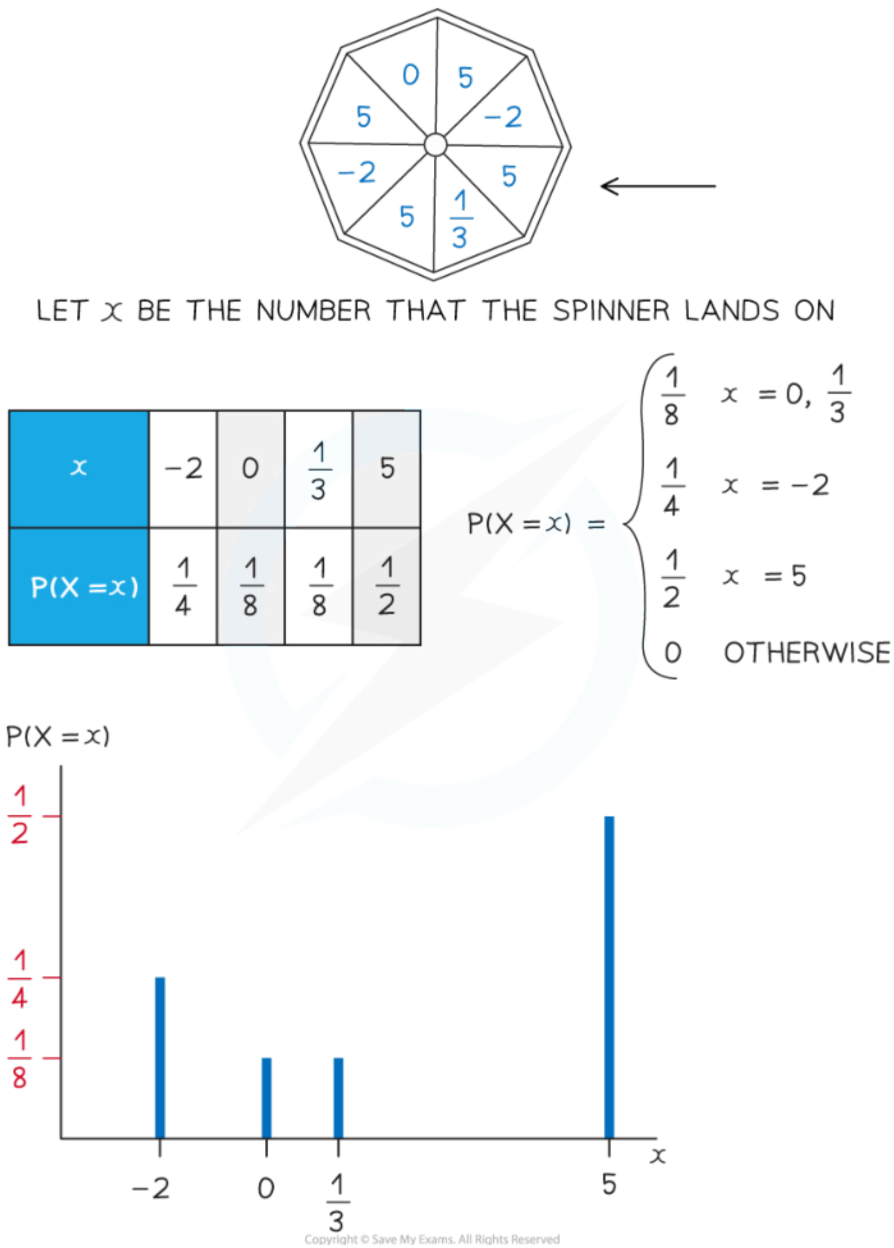
What is a discrete random variable?

- A **random variable** is a variable whose value depends on the outcome of a **random event**
  - The value of the random variable is not known until the event is carried out (this is what is meant by 'random' in this case)
- **Random variables** are denoted using **upper case letters** (X, Y, etc)
- **Particular outcomes** of the event are denoted using **lower case letters** (x, y, etc)
- $\mathrm{P}(X = x)$ means "the probability of the random variable X taking the value x"
- A **discrete** random variable (often abbreviated to DRV) can only take **certain values** within a set
  - Discrete random variables **usually count** something
  - Discrete random variables usually can only take a finite number of values but it is possible that it can take an infinite number of values (see the examples below)
- **Examples** of discrete random variables include:
  - The number of times a coin lands on heads when flipped 20 times
    - this has a finite number of outcomes: {0,1,2,...,20}
  - The number of emails a manager receives within an hour
    - this has an infinite number of outcomes: {1,2,3,...}
  - The number of times a dice is rolled until it lands on a 6
    - this has an infinite number of outcomes: {1,2,3,...}
  - The number that a dice lands on when rolled once
    - this has a finite number of outcomes: {1,2,3,4,5,6}

What is a probability distribution of a discrete random variable?

- A **discrete probability distribution** fully describes **all the values** that a discrete random variable can take along with their **associated probabilities**
  - This can be given in a **table**
  - Or it can be given as a **function** (called a discrete probability distribution function or "pdf")
  - They can be represented by **vertical line graphs** (the possible values for along the horizontal axis and the probability on the vertical axis)
- The **sum of the probabilities** of **all the values** of a discrete random variable is **1**
  - This is usually written $\sum \mathrm{P}(X = x) = 1$
- A **discrete uniform distribution** is one where the random variable takes a finite number of values each with an **equal probability**
  - If there are n values then the probability of each one is $\frac{1}{n}$

Your notes



How do I calculate probabilities using a discrete probability distribution?

- First **draw a table** to represent the probability distribution
  - If it is given as a function then find each probability
  - If any probabilities are unknown then use algebra to represent them
- **Form an equation** using $\sum \mathrm{P}(X = x) = 1$
  - Add together all the probabilities and make the sum equal to 1
- To find $\mathrm{P}(X = k)$

Your notes

  - If k is a possible value of the random variable X then $P(X = k)$ will be given in the table
  - If k is not a possible value then $P(X = k) = 0$
- To find $P(X \le k)$
  - Identify all possible values, x_i, that X can take which satisfy $x_i \le k$
  - Add together all their corresponding probabilities
  - $P(X \le k) = \sum_{x_i \le k} P(X = x_i)$
  - Some mathematicians use the notation $F(x)$ to represent the cumulative distribution
    - $F(x) = P(X \le x)$
- Using a similar method you can find $P(X < k)$, $P(X > k)$ and $P(X \ge k)$
- As all the probabilities add up to 1 you can form the following equivalent equations:
  - $P(X < k) + P(X = k) + P(X > k) = 1$
  - $P(X > k) = 1 - P(X \le k)$
  - $P(X \ge k) = 1 - P(X < k)$

How do I know which inequality to use?

- $P(X \le k)$ would be used for phrases such as:
  - At most , no greater than , etc
- $P(X < k)$ would be used for phrases such as:
  - Fewer than
- $P(X \ge k)$ would be used for phrases such as:
  - At least , no fewer than , etc
- $P(X > k)$ would be used for phrases such as:
  - Greater than , etc

Worked example

The probability distribution of the discrete random variable X is given by the function

$$P(X=x)=\begin{cases} kx^2 & x=-3,\ -1,\ 2,\ 4 \\ 0 & \text{otherwise.} \end{cases}$$

a) Show that $k=\dfrac{1}{30}$.

Construct a table

x	-3	-1	2	4
$P(X=x)$	$9k$	k	$4k$	$16k$

Substitute in the values of x
e.g. $P(X=-3)=k(-3)^2=9k$

The probabilities add up to 1

$9k+k+4k+16k=1$

$30k=1$

$k=\dfrac{1}{30}$

b) Calculate $P(X\le 3)$.

Substitute k into the probabilities

x	-3	-1	2	4
$P(X=x)$	$\frac{3}{10}$	$\frac{1}{30}$	$\frac{2}{15}$	$\frac{8}{15}$

$X\le 3$: $X=-3,-1,2$

$P(X\le 3)=P(X=-3)+P(X=-1)+P(X=2)$

$=\dfrac{3}{10}+\dfrac{1}{30}+\dfrac{2}{15}$

$P(X\le 3)=\dfrac{7}{15}$

Your notes

Your notes

4.4.2 Mean & Variance

Expected Values E(X)

What does E(X) mean and how do I calculate E(X)?

- **E(X)** means the **expected value** or the **mean** of a **random variable X**
  - The expected value does not need to be an obtainable value of *X*
  - For example: the expected value number of times a coin will land on tails when flipped 5 times is 2.5
- For a **discrete** random variable, it is calculated by:
  - **Multiplying each value** of X with its corresponding **probability**
  - **Adding** all these terms together

$$\mathrm{E}(X) = \sum x\mathrm{P}(X = x)$$

    - This is given in the **formula booklet**
- Look out for **symmetrical** distributions (where the values of X are symmetrical and their probabilities are symmetrical) as the mean of these is the same as the median
  - For example: if X can take the values 1, 5, 9 with probabilities 0.3, 0.4, 0.3 respectively then by symmetry the mean would be 5

How can I decide if a game is fair?

- Let *X* be the random variable that represents the **gain/loss** of a player in a game
  - *X* will be **negative** if there is a **loss**
- Normally the expected gain or loss is calculated by **subtracting** the **cost to play** the game from the **expected value** of the **prize**
- If E(*X*) is **positive** then it means the player can **expect to make a gain**
- If E(*X*) is **negative** then it means the player can **expect to make a loss**
- The game is called **fair** if the **expected gain is 0**
  - E(*X*) = 0

Your notes

Worked example

Daphne pays $15 to play a game where she wins a prize of $1, $5, $10 or $100. The random variable W represents the amount she wins and has the probability distribution shown in the following table:

w	1	5	10	100
$\mathrm{P}(W = w)$	0.35	0.5	0.05	0.1

a) Calculate the expected value of Daphne's prize.

Formula booklet

Expected value of a discrete random variable X	$\mathrm{E}(X) = \sum x\,\mathrm{P}(X = x)$

$$\mathrm{E}(W) = \sum w\,\mathrm{P}(W = w)$$

$$= 1 \times 0.35 + 5 \times 0.5 + 10 \times 0.05 + 100 \times 0.1$$

Expected value = $13.35

b) Determine whether the game is fair.

A game is fair is expected gain/loss is 0

Prize − cost

$13.35 - 15 = -1.65$

Expected loss is $1.65 so game is not fair

Your notes

Variance Var(X)

What does Var(X) mean and how do I calculate Var(X)?

- **Var(X)** means the variance of a **random variable X**
  - The **standard deviation** is the **square root** of the variance
    - This provides a **measure of the spread** of the outcomes of *X*
  - The variance and standard deviation can **never be negative**
- The variance of X is the **mean of the squared difference** between X and the mean

$$\mathrm{Var}(X) = \mathrm{E}(X - \mu)^2$$

  - This is given in the **formula booklet**
- This formula can be rearranged into the more useful form:

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$$

  - This is given in the **formula booklet**
    - Compare this formula to the formula for the **variance of a set of data**
  - This formula works for both **discrete** and **continuous** X

How do I calculate E(X²) for discrete X?

- **$E(X^2)$** means the **expected value** or the **mean** of the **random variable** defined as **X^2**
- For a **discrete** random variable, it is calculated by:
  - **Squaring** each value of X to get the values of X^2
  - **Multiplying each value** of X^2 with its corresponding **probability**
  - **Adding** all these terms together
    - $\mathrm{E}(X^2) = \sum x^2 \mathrm{P}(X = x)$
    - This is given in the formula booklet as part of the formula for Var(*X*)
    - $\mathrm{Var}(X) = \sum x^2 \mathrm{P}(X = x) - \mu^2$
- **E(f(X))** can be found in a similar way

Is E(X²) equal to E(X)²?

- **Definitely not!**
  - They are only equal if X can only take one value
- $E(X^2)$ is the **mean of the values of X^2**
- $E(X)^2$ is the **square of the mean of the values of X**
- To see the difference
  - Imagine a random variable X that can only take 1 and -1 with equal chance
  - E(X) = 0 so **$E(X)^2 = 0$**
  - The square values are 1 and 1 so **$E(X^2) = 1$**

Your notes

Examiner Tip

- In an exam you can enter the probability distribution into your GDC using the statistics mode
  - Enter the possible values as the data
  - Enter the probabilities as the frequencies
- You can then calculate the mean and variance just like you would with data

Worked example

The score on a game is represented by the random variable S defined below.

s	0	1	2	10
$\mathrm{P}(S=s)$	0.4	0.3	0.25	0.05

Calculate $\mathrm{Var}(S)$.

Calculate E(S)

Formula booklet

Expected value of a discrete random variable X	$\mathrm{E}(X)=\sum x\,\mathrm{P}(X=x)$

$$E(S)=\sum s\,P(S=s)=0\times0.4+1\times0.3+2\times0.25+10\times0.05=1.3$$

Calculate $E(S^2)$

$$E(S^2)=\sum s^2P(S=s)=0^2\times0.4+1^2\times0.3+2^2\times0.25+10^2\times0.05=6.3$$

Calculate Var(S)

Formula booklet

Variance	$\mathrm{Var}(X)=\mathrm{E}(X-\mu)^2=\mathrm{E}(X^2)-[\mathrm{E}(X)]^2$

$$\mathrm{Var}(S)=E(S^2)-[E(S)]^2=6.3-1.3^2$$

$$\mathrm{Var}(S)=4.61$$

Your notes

Transformation of a Single Variable

How do I calculate the expected value and variance of a transformation of X?

- Suppose *X* is **transformed** by the function *f* to form a new variable *T* = *f*(*X*)
  - This means the function *f* is applied to all possible values of *X*
- Create a **new probability distribution table**
  - The top row contains the values $t_i = f(x_i)$
  - The bottom row still contains the values $\mathrm{P}(X = x_i)$ which are unchanged as:
    - $\mathrm{P}(X = x_i) = \mathrm{P}(f(X) = f(x_i)) = \mathrm{P}(T = t_i)$
    - Some values of *T* may be equal so you can add their probabilities together
- The **mean** is calculated in the same way
  - $\mathrm{E}(T) = \sum t\mathrm{P}(X = x)$
- The **variance** is calculated using the same formula
  - $\mathrm{Var}(T) = \mathrm{E}(T^2) - [\mathrm{E}(T)]^2$

Are there any shortcuts?

- There are formulae which can be used if the transformation is **linear**
  - $T = aX + b$ where *a* and *b* are constants
- If the transformation is **not linear** then there are **no shortcuts**
  - You will have to first find the probability distribution of *T*

What are the formulae for E(*aX* + *b*) and Var(*aX* + *b*)?

- If *a* and *b* are constants then the following formulae are true:
  - $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$
  - $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$
    - 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
  - $\frac{X}{a} = \frac{1}{a}X$

Your notes

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	$\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$ $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$

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

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

$\mathrm{Var}(2-X) = (-1)^2\,\mathrm{Var}(X) = 1(4)$ **$\mathrm{Var}(2-X) = 4$**