Edexcel International AS Maths: Statistics 1

Your notes

Discrete Random Variables

Contents

- Discrete Probability Distributions
- E(X) & Var(X) (Discrete)
- aX+b
- Discrete Uniform Distribution

Discrete Random Variables

What is a discrete random variable?

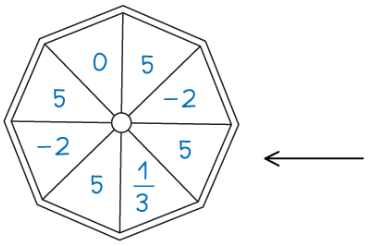
- A **random variable** is a variable whose value depends on the outcome of a **random event**
  - The value of the random variable is not known until the event is carried out (this is what is meant by 'random' in this case)
- **Random variables** are denoted using **upper case letters** (X, Y, etc)
- **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 on a bingo ball when one is drawn at random (this has a finite number of outcomes: 1,2,3…,90)

Probability Distributions (Discrete)

What is a probability distribution?

- 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** (similar to GCSE)
  - Or it can be given as a **function** (called a probability mass function)
  - 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**

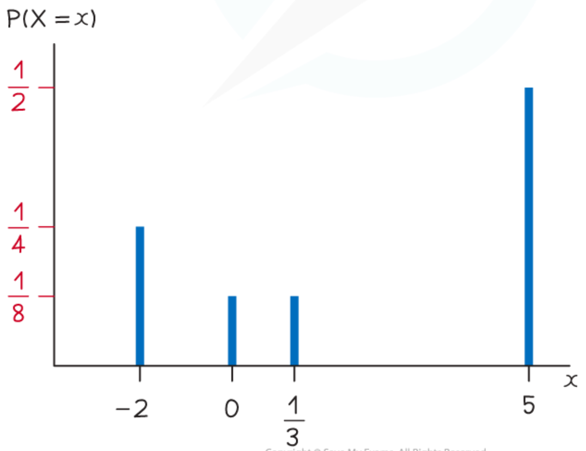
- This is usually written $\Sigma P(X = x) = 1$
- A **discrete uniform distribution** is one where the random variable takes a finite number of values each with an **equal probability**
    - If there are n values then the probability of each one is $\frac{1}{n}$



LET x BE THE NUMBER THAT THE SPINNER LANDS ON

x	-2	0	$\frac{1}{3}$	5
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{2}$

$$P(X = x) = \begin{cases} \frac{1}{8} & x = 0, \frac{1}{3} \\ \frac{1}{4} & x = -2 \\ \frac{1}{2} & x = 5 \\ 0 & \text{OTHERWISE} \end{cases}$$



Cumulative Probabilities (Discrete)

How do I calculate probabilities using a discrete probability distribution?

- For probability distributions that take a small number of values start by **drawing a table** to represent the probability distribution
    - If the distribution is given as a function then find each probability
    - If any probabilities are unknown then use algebra to represent them

Your notes

- **Form an equation** using $\sum P(X = x) = 1$

- To find $P(X = k)$
  - 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 $P(X = k) = 0$

What is the cumulative distribution function?

- The cumulative distribution function, denoted $F(x)$, is the probability that the random variable takes a value less than or equal to x.
  - $F(x) = P(X \le x)$
- You may be asked to draw a table for the cumulative distribution function
  - This will be similar to a probability distribution function but instead the bottom row will be F(x) instead of P(*X* = *x*)

How do I calculate cumulative probabilities?

- To find $P(X \le x)$ (equivalently F(x))
  - Identify all possible values, x_i, that *X* can take which satisfy $x_i \le k$
  - Add together all their corresponding probabilities
  - $P(X \le k) = \sum_{x_i \le k} P(X = x_i)$
- Using a similar method you can find $P(X < k)$, $P(X \ge k)$ and $P(X > k)$
- As all the probabilities add up to 1 you can form the following equivalent equations:
  - $P(X < k) + P(X = k) + P(X > k) = 1$
  - $P(X > k) = 1 - P(X \le k)$
  - $P(X \ge k) = 1 - P(X < k)$
- To calculate more complicated probabilities such as $P(X^2 < 4)$
  - Identify which values of the random variable satisfy the inequality or event in the brackets
  - Add together the corresponding probabilities

How do I know which inequality to use?

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

Your notes

- Fewer than k
- $\mathrm{P}(X \geq k)$ would be used for phrases such as:
  - At least k , no fewer than k, etc
- $\mathrm{P}(X > k)$ would be used for phrases such as:
  - Greater than k, etc

Your notes

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

(b) Calculate $\mathrm{F}(3)$.

(c) Calculate $\mathrm{P}(X^2 < 5)$.

Answer:

a) Construct a table

x	-3	-1	2	4
$\mathrm{P}(X = x)$	$9k$	k	$4k$	$16k$

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

The probabilities add up to 1

$$9k + k + 4k + 16k = 1$$
$$30k = 1$$
$$k = \frac{1}{30}$$

b) Substitute k into the probabilities

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

F(3) means $P(X \leqslant 3)$

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

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

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

$$F(3) = \frac{7}{15}$$

Note: $1 - \frac{8}{15} = \frac{7}{15}$

c) $X^2 < 5$ for $X = -1, 2$

$$P(X^2 < 5) = P(X = -1) + P(X = 2)$$

$$= \frac{1}{30} + \frac{2}{15}$$

$$P(X^2 < 5) = \frac{1}{6}$$

Examiner Tips and Tricks

- Try to draw a table if there are a finite number of values that the discrete random variable can take
- When finding a probability, it will sometimes be quicker to subtract the probabilities of the unwanted values from 1 rather than adding together the probabilities of the wanted values
- Always make sure that the probabilities are between 0 and 1, and that they add up to 1!

Your notes

E(X) & Var(X) (Discrete)

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**
- For a **discrete** random variable, it is calculated by:
  - **Multiplying each value** of X with its corresponding **probability**
  - **Adding** all these terms together

$$\Sigma x\mathrm{P}(X = x)$$

- 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 do I calculate E(X²)?

- **E(X²)** means the **expected value** or the **mean** of a **random variable** defined as **X²**
- 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

$$\Sigma x^2\mathrm{P}(X = x)$$

- In a similar way E(f(x)) can be calculated for a discrete random variable by:
  - **Applying the function f** to each value of to get the values of f(X)
  - **Multiplying each value** of f(X) with its corresponding **probability**
  - **Adding** all these terms together

$$\Sigma f(x)\,\mathrm{P}(X = x)$$

Your notes

x	a	b	c
$P(X = x)$	p	q	r

$$E(X) = ap + bq + cr$$

$$E(X^2) = a^2p + b^2q + c^2r$$

$$E\left(\frac{1}{X}\right) = \frac{p}{a} + \frac{q}{b} + \frac{r}{c}$$

x	-2	3	5
$P(X = x)$	0.5	0.1	0.4

$$E(X) = (-2)(0.5) + (3)(0.1) + (5)(0.4)$$

$$E(X^2) = (-2)^2(0.5) + (3^2)(0.1) + (5^2)(0.4)$$

$$E\left(\frac{1}{X}\right) = \left(-\frac{1}{2}\right)(0.5) + \left(\frac{1}{3}\right)(0.1) + \left(\frac{1}{5}\right)(0.4)$$

Is $E(X^2)$ equal to $(E(X))^2$?

- **Definitely not!**
  - They are only equal if X can take only one value with probability 1
    - if this was the case it would no longer be a random variable
- $E(X^2)$ is the **mean** of the values of $\mathbf{X^2}$
- $(E(X))^2$ is the **square** of the **mean** of the values of **X**
- To see the difference
  - Imagine a random variable X that can only take the values 1 and -1 with equal chance
  - The mean would be 0 so the square of the mean would also be 0
  - The square values would be 1 and 1 so the mean of the squares would also be 1
- In general E(f(X)) **does not equal** f(E(X)) where f is a function
  - So if you wanted to find something like $E\left(\frac{1}{X}\right)$ then you would have to use the definition and calculate:

$$\sum \frac{1}{x} \mathrm{P}(X=x)$$

Your notes

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

- **Var(X)** means the **variance** of a **random variable X**
- For **any** random variable this can be calculated using the formula

$$E(X^2)-(E(X))^2$$

- This is the **mean of the squares of *X*** minus the **square of the mean of *X***
  - Compare this to the definition of the **variance of a set of data**
- Var(X) is always positive
- The **standard deviation** of a random variable X is the **square root** of **Var(X)**

Worked Example

The discrete random variable X has the probability distribution shown in the following table:

x	2	3	5	7
$\mathrm{P}(X=x)$	0.1	0.3	0.2	0.4

(a) Find the value of $\mathrm{E}(X)$.

(b) Find the value of $E(X^2)$.

(c) Find the value of $\mathrm{Var}(X)$.

Answer:

a) Use the formula "$E(x) = \Sigma x P(X = x)$"

$E(x) = (2)(0.1) + (3)(0.3) + (5)(0.2) + (7)(0.4)$

$E(x) = 4.9$

b) Use the formula "$E(x^2) = \Sigma x^2 P(X = x)$"

$E(x^2) = (2^2)(0.1) + (3^2)(0.3) + (5^2)(0.2) + (7^2)(0.4)$

$E(x^2) = 27.7$

c) Use the formula "$Var(x) = E(x^2) - (E(x))^2$"

$Var(x) = 27.7 - 4.9^2$

$Var(x) = 3.69$

Your notes

Examiner Tips and Tricks

- Check if your answer makes sense. The mean should fit within the range of the values of X.

aX+b

How are the mean and variance of X related to the mean and variance of aX + b?

- If *a* and *b* are **constants** then the following results are true
  - $E(aX + b) = aE(X) + b$
  - $Var(aX + b) = a^2 Var(X)$
- Note that the mean is affected by multiplication and addition whereas **addition does not change the variance**
- The factor of a^2 includes the squared because the values of X are squared in the calculation
  - You could try and use the first result and the formula for variance to verify the second result
- Remember a subtraction can be written as an addition
  - $X - b$ can be written as $X + (-b)$
- And division can be written as a multiplication
  - $\frac{x}{a}$ can be written as $\frac{1}{a}X$

Worked Example

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

Find the value of:

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

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

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

Answer:

Your notes

(i) Use "$E(aX + b) = aE(X) + b$"

$E(3X + 5) = 3E(X) + 5$

$= 3(5) + 5$

$E(3X + 5) = 20$

(ii) Use "$Var(aX + b) = a^2Var(X)$"

$Var(3X + 5) = 3^2Var(X)$

$= 9(4)$

$Var(3X - 5) = 36$

(iii) Rewrite in form $aX + b$

$2 - X = (-1)X + 2$

$Var((-1)X + 2) = (-1)^2Var(X)$

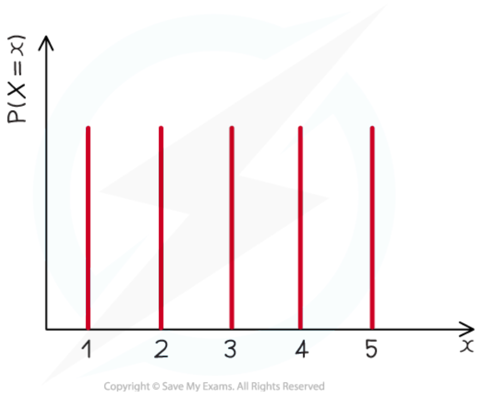
$= 1(4)$

$Var(2 - X) = 4$

Discrete Uniform Distribution

What is a discrete uniform distribution?

- A discrete uniform distribution is a discrete probability distribution
- The **discrete random variable** *X* follows a **discrete uniform distribution** if
  - There are a **finite number** of distinct outcomes (*n*)
  - Each **outcome** is **equally likely**
- If there are n distinct outcomes, $P(X = x) = \frac{1}{n}$
- In many cases the outcomes of *X* are the integers 1, 2, 3, .., *n*
  - $P(X = x) = \frac{1}{n}$ for $n = 1, 2, 3, \ldots, n$
  - 0 for any other value of X
- The distribution can be represented visually using a **vertical line graph** where the lines have **equal heights**



What is the mean and variance of a discrete uniform distribution?

- If the outcomes of *X* are the integers 1, 2, 3, ..., n
  - The **expected value (mean)** is $\frac{n+1}{2}$
  - The **variance** is $\frac{n^2-1}{12}$
    - Square root to get the standard deviation

Your notes

- The discrete uniform distribution is **symmetrical** so the median is the same as the mean
  - There is no mode as each value is equally likely

Do the outcomes have to be 1 to n?

- The numbers can be anything as long as they are equally likely
- The formulae for the mean and variance only apply when the values are the integers 1 to n
- If the outcomes form an **arithmetic sequence** then the distribution can be transformed to the distribution with the values 1 to n
- If X is the discrete uniform distribution using 1 to n and Y is a discrete uniform distribution whose outcomes form an arithmetic sequence then:
  - $Y = aX + b$
- You can then use this formula to find the **mean** and **variance**
  - $E(Y) = aE(X) + b$
  - $Var(Y) = a^2 Var(X)$
- For example: $Y = 2, 5, 8, 11$ can be transformed to $X = 1, 2, 3, 4$ using $Y = 3X - 1$

What can be modelled using a discrete uniform distribution?

- Anything which satisfies the **two conditions**
  - finite distinct outcomes and all equally likely
- For example, let R be the second digit of a number given by a random number generator
  - There are 10 distinct outcomes: 0, 1, 2, ..., 9
  - As it is a random number then each value is equally likely to be the second digit

What can not be modelled using a discrete uniform distribution?

- Anything where the number of outcomes is **infinite**
  - The number obtained when a person is asked to write down any integer
- Anything where the outcomes are **not equally likely**
  - The number obtained when one of the first 5 Fibonacci numbers is randomly selected
    - 1, 1, 2, 3, 5
    - 1 appears twice so is more likely to be picked than the rest

Worked Example

Each odd number from 1 to 99 is written on an individual tile and one is chosen at random. The random variable T represents the number on the chosen tile.

(a) Find $E(T)$.

(b) Find $\mathrm{Var}(T)$.

Answer:

Let X be the discrete uniform distribution with outcomes 1 to n

Find a formula linking X and T

X				T
1	$\longrightarrow$ ×2	$\longrightarrow$ −1	$\longrightarrow$	1
2	$\longrightarrow$ ×2	$\longrightarrow$ −1	$\longrightarrow$	3
3	$\longrightarrow$ ×2	$\longrightarrow$ −1	$\longrightarrow$	5
⋮				⋮
n				99

You could also use $U_n = a + (n-1)d$ for the arithmetic sequence

$T = 2X - 1$

Find the value of n

T = 99 X = n

$99 = 2n - 1$

$n = 50$

a) Use $E(aX + b) = aE(X) + b$

$E(T) = 2E(X) - 1$

Find $E(X)$ using $\dfrac{n+1}{2}$

$$= 2\left(\frac{50+1}{2}\right) - 1$$

$E(T) = 50$

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

$\mathrm{Var}(T) = 2^2\,\mathrm{Var}(X)$

Find $\mathrm{Var}(X)$ using $\dfrac{n^2-1}{12}$

$$= 4\left(\frac{50^2-1}{12}\right)$$

$\mathrm{Var}(T) = 833$

Examiner Tips and Tricks

- Always check your mean and variance makes sense. If the numbers go from 1 to 100 then a mean of 101 is not possible!

Your notes