# Probability Distributions

*You need to know about a couple of specific statistical distributions for A-Level Maths. But first of all, you need to know what statistical distributions are...*

## ***Random Variables** have **Probability Distributions***

This first bit isn't particularly interesting. But understanding the difference between $X$ and $x$ (bear with me) might make the later stuff a bit less confusing.

1) $X$ (upper case) is just the **name** of a **random variable**. So $X$ could be 'score on a fair, six-sided dice'.
2) A **random variable** doesn't have a **fixed** value. Like with the dice score — the value on any 'roll' is all down to **chance**.
3) $x$ (lower case) is a **particular value** that $X$ can take. So for one roll of the dice, $x$ could be 1, 2, 3, 4, 5 or 6.
4) **Discrete** random variables only have a **certain number** of possible values. Often these values are whole numbers, but they don't have to be. Usually there are only a few possible values (e.g. the possible scores with one roll of a dice).
5) A **probability distribution** is a **table showing the possible values** of $x$, and the **probability** for each one.
6) A **probability function** is a formula that generates the probabilities for different values of $x$.

## *All the probabilities **Add Up To 1***

For a discrete random variable $X$:

$$\sum_{\text{all } x} P(X = x) = 1$$

This says that if you add up the probabilities of all the possible values of X, you get 1.

**Example:** The random variable $X$, where $X$ can only take values 1, 2, 3, has probability function $P(X = x) = kx$ for $x = 1, 2, 3$. Find the value of $k$.

$X$ has three possible values ($x = 1$, 2 and 3), and the probability of each is $kx$ (where you need to find $k$).

It's easier to understand with a table:

| $x$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(X = x)$ | $k \times 1 = \mathbf{k}$ | $k \times 2 = \mathbf{2k}$ | $k \times 3 = \mathbf{3k}$ |

Now just use the formula: $\sum_{\text{all } x} P(X = x) = 1$

Here, this means: $k + 2k + 3k = 6k = 1$, so $k = \frac{1}{6}$

For a discrete random variable where every value of X is equally likely, you get a discrete uniform distribution — e.g. rolling a normal unbiased dice.

The **mode** is the **most likely** value — so it's the value with the **biggest probability**.

**Example:** The discrete random variable $X$, where $X$ can only take values 0, 1, 2, 3, 4, has the probability distribution shown below.

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(X = x)$ | 0.1 | 0.2 | 0.3 | 0.2 | $a$ |

Find: a) the value of $a$, b) $P(2 \le X < 4)$, c) the mode.

a) Use the formula $\sum_{\text{all } x} P(X = x) = 1$ again.

From the table: $0.1 + 0.2 + 0.3 + 0.2 + a = 1$
$0.8 + a = 1$
$\mathbf{a = 0.2}$

Careful with the inequality signs — you need to include x = 2 but not x = 4.

b) This is asking for the probability that '$X$ is greater than or equal to 2, but less than 4'. Easy — just add up the probabilities.

$P(2 \le X < 4) = P(X = 2) + P(X = 3) = 0.3 + 0.2 = \mathbf{0.5}$

c) The mode is the value of $x$ with the biggest probability — so **mode = 2**.

# Probability Distributions

## *Draw a **Diagram** showing **All Possible Outcomes***

**Example:** An unbiased six-sided dice has faces marked 1, 1, 1, 2, 2, 3.
The dice is rolled twice. Let $X$ be the random variable "sum of the two scores on the dice".
a) Show that $P(X = 4) = \frac{5}{18}$. b) Find the probability distribution of $X$.

a) Make a table showing the 36 possible outcomes.

Score on roll 1 (columns); Score on roll 2 (rows)

| + | 1 | 1 | 1 | 2 | 2 | 3 |
|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 2 | 3 | 3 | 4 |
| 1 | 2 | 2 | 2 | 3 | 3 | 4 |
| 1 | 2 | 2 | 2 | 3 | 3 | 4 |
| 2 | 3 | 3 | 3 | 4 | 4 | 5 |
| 2 | 3 | 3 | 3 | 4 | 4 | 5 |
| 3 | 4 | 4 | 4 | 5 | 5 | 6 |

You can see from the table that 10 of these have the outcome $X = 4$, so $P(X = 4) = \frac{10}{36} = \mathbf{\frac{5}{18}}$

b) Use the table to work out the probabilities for the other outcomes and then fill in a table summarising the probability distribution:

$\frac{9}{36}$ of the outcomes are a score of 2
$\frac{12}{36}$ of the outcomes are a score of 3
$\frac{4}{36}$ of the outcomes are a score of 5
$\frac{1}{36}$ of the outcomes are a score of 6

| $x$ | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $P(X = x)$ | $\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{5}{18}$ | $\frac{1}{9}$ | $\frac{1}{36}$ |

## *Do complicated questions **Bit By Bit***

**Example:** A game involves rolling two fair, six-sided dice. If the sum of the scores is greater than 10 then the player wins 50p. If the sum is between 8 and 10 (inclusive) then they win 20p. Otherwise they get nothing. If $X$ is the random variable "amount player wins in pence", find the probability distribution of $X$.

There are **3 possible values** for $X$ (0, 20 and 50) and you need the **probability** of each.
To work these out, you need the probability of getting various totals on the dice.

1) You need to know **$P(8 \le \text{score} \le 10)$** — the probability that the score is between 8 and 10 **inclusive** (i.e. including 8 and 10) and **$P(11 \le \text{score} \le 12)$** — the probability that the score is **greater than** 10.
Use a table:

Score on dice 1 (columns); Score on dice 2 (rows)

| + | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

The table: making things easier to understand since 3000 BC (DISCLAIMER: CGP takes no responsibility for the historical accuracy of this 'fact').

2) There are **36 possible outcomes**:

**12 of these** have a total of **8, 9 or 10** so $P(8 \le \text{score} \le 10) = \frac{12}{36} = \frac{1}{3}$

**3 of these** have a total of **11 or 12** so $P(11 \le \text{score} \le 12) = \frac{3}{36} = \frac{1}{12}$

3) Use these to find the probabilities you need:

$P(X = 20\text{p}) = P(8 \le \text{score} \le 10) = \frac{1}{3}$

$P(X = 50\text{p}) = P(11 \le \text{score} \le 12) = \frac{1}{12}$

To find $P(X = 0)$ take the total of the two probabilities above from 1 (since $X = 0$ is the only other possibility).

$$P(X = 0) = 1 - \left[\frac{12}{36} + \frac{3}{36}\right] = 1 - \frac{15}{36} = \frac{21}{36} = \frac{7}{12}$$

4) Now just stick all this info in a table (and check that the probabilities all add up to 1):

| $x$ | 0 | 20 | 50 |
|---|---|---|---|
| $P(X = x)$ | $\frac{7}{12}$ | $\frac{1}{3}$ | $\frac{1}{12}$ |

Check: $\frac{7}{12} + \frac{1}{3} + \frac{1}{12} = \frac{7}{12} + \frac{4}{12} + \frac{1}{12} = 1$ ✓

# Probability Distributions

## *The **Cumulative Distribution Function** is a **Running Total** of probabilities*

The **cumulative distribution function** $F(x)$ gives the probability that $X$ will be **less than or equal to** a particular value.

$$F(x_0) = P(X \le x_0) = \sum_{x \le x_0} p(x)$$

$p(x) = P(X = x)$

**Example:** The probability distribution of the discrete random variable $H$, where $H$ can only take values 0.1, 0.2, 0.3, 0.4, is shown in the table. Draw up a table to show the cumulative distribution of $H$.

| $h$ | 0.1 | 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|
| $P(H = h)$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{1}{6}$ |

There are 4 values of $h$, so you have to find the probability that $H$ is **less than or equal to** each of them in turn. It sounds trickier than it actually is — you only have to add up a few probabilities...

$F(0.1) = P(H \le 0.1)$ — this is the same as $P(H = 0.1)$, since $H$ can't be less than 0.1. So $\mathbf{F(0.1) = \frac{1}{4}}$

$F(0.2) = P(H \le 0.2)$ — this is the probability that $H = 0.1$ or $H = 0.2$.

$\mathbf{F(0.2)} = P(H = 0.1) + P(H = 0.2) = \frac{1}{4} + \frac{1}{4} = \mathbf{\frac{1}{2}}$

$\mathbf{F(0.3)} = P(H \le 0.3) = P(H \le 0.2) + P(H = 0.3) = \frac{1}{2} + \frac{1}{3} = \mathbf{\frac{5}{6}}$

$\mathbf{F(0.4)} = P(H \le 0.4) = P(H \le 0.3) + P(H = 0.4) = \frac{5}{6} + \frac{1}{6} = \mathbf{1}$

*Here you're just adding one more probability to the previous cumulative probability.*

Finally, put these values in a table, and you're done:

| $h$ | 0.1 | 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|
| $F(h) = P(H \le h)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{5}{6}$ | 1 |

*$P(X \le$ largest value of $x)$ is always 1.*

**Example:** For a discrete random variable $X$, where $X$ can only take values 1, 2, 3, 4, the cumulative distribution function $F(x) = kx$, for $x = 1, 2, 3$ and 4. Find $k$, and the probability function.

1) First find $k$. You know that $X$ has to be 4 or less — so $\mathbf{P(X \le 4) = 1}$.
   Put $x = 4$ into the cumulative distribution function: $F(4) = P(X \le 4) = 4k = 1$, so $k = \frac{1}{4}$.
2) Now you can work out the probabilities of $X$ being less than or equal to 1, 2, 3 and 4.
   $\mathbf{F(1)} = P(X \le 1) = 1 \times k = \frac{1}{4}$, $\mathbf{F(2)} = P(X \le 2) = 2 \times k = \frac{1}{2}$, $\mathbf{F(3)} = P(X \le 3) = 3 \times k = \frac{3}{4}$, $\mathbf{F(4)} = P(X \le 4) = 1$
3) Then $\mathbf{P(X = 4)} = P(X \le 4) - P(X \le 3) = 1 - \frac{3}{4} = \frac{1}{4}$, $\mathbf{P(X = 3)} = P(X \le 3) - P(X \le 2) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$,
   *Think about it — if it's less than or equal to 4, but it's **not** less than or equal to 3, then it has to be 4.*
   $\mathbf{P(X = 2)} = P(X \le 2) - P(X \le 1) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$ and $\mathbf{P(X = 1)} = P(X \le 1) = \frac{1}{4}$
   *Because x doesn't take any values less than 1.*
4) Finish it all off by making a table. The probability distribution of $X$ is:

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

So the probability function is: $\mathbf{P(X = x) = \frac{1}{4}}$ **for $x$ = 1, 2, 3, 4**

This is a uniform distribution (see p.158).

## *Practice Question*

Q1 The random variable $X$, where $X$ takes values 1, 2, 3, 4, has probability function $P(X = x) = kx$ for $x = 1, 2, 3, 4$.

a) Find the value of $k$.
b) Find $P(X > 2)$
c) Find $P(1 \le X \le 3)$
d) Draw a table to show: (i) the probability distribution, (ii) the cumulative distribution.

**Exam Question**

Q1 The probability function for the discrete random variable $X$ is given by $P(X = x) = \frac{1}{k}x^2$ for $x = 1, 2, 3, 4$.
Find the value of $k$ and $P(X \le 2)$. [4 marks]

## *Fact: the probability of this coming up in the exam is less than or equal to one...*

*If you've got a probability distribution, you can work out the table for the cumulative distribution function and vice versa. Don't forget, all the stuff so far is for discrete variables — these can only take a certain number of values.*