Statistics Sector 1 : Discrete Probability Distributions and Binomial Distribution

Aims:

- Understand and use a discrete probability table / function
- To find the probability of events occurring using a binomial probability function.
- To be able to calculate binomial mean and variance.
- Recognise when it is suitable to model a situation using a binomial distribution

Discrete Probability Distributions

A *Probability Distribution* shows all the values of a variable (x) and their probabilities, $P(X = x)$:

Example 1

X: total score when 2 fair dice are rolled

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Always check that

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

A probability distribution can also be given as a function of x:

Example 2

A discrete random variable, X, has a probability density function defined by:

$$P(X = x) = kx^2 \qquad \text{for } x = 1,2,3,4,5$$

a) Write down the probability distribution in a table.
b) Find the value of *k*.
c) Find $P(X < 3)$.

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

a)

x	1	2	3	4	5
$P(X=x)$	k	$4k$	$9k$	$16k$	$25k$

b) $k + 4k + 9k + 16k + 25k = 1$

$55k = 1$

$k = \frac{1}{55}$

c) $P(X<3) = k + 4k$

$= \frac{1}{11}$

Example 3

The probability distribution of X, the number of boxes of 'Tiddles' purchased per customer, is given by:

$$P(X = x) = \begin{cases} \left(\frac{2}{3}\right)^{x+2} + \left(\frac{1}{3}\right)^{x+3} & x = 1,2,3 \\ k & x = 4 \\ 0 & otherwise \end{cases}$$

a) Find the exact value of the constant k
b) Find $P(3 \le X \le 4)$

a)

x	1	2	3	4
$P(X=x)$	$\frac{25}{81}$	$\frac{49}{243}$	$\frac{97}{729}$	K

$$\frac{25}{81} + \frac{49}{243} + \frac{97}{729} + K = 1$$

$$K = \frac{260}{729}$$

b) $P(3 \le X \le 4) = \frac{97}{729} + \frac{260}{729} = \frac{119}{243}$

Binomial Distribution

In the binomial distribution, the random variable has two possible outcomes. A tree diagram can be used to find binomial probabilities but only in very simple cases otherwise this is too difficult to construct and time consuming.

Random variables are denoted by italic capitals and the value of the random in lower case italics. For example the probability that the random variable X has the value x is 0.6 is written as $P(X = x) = 0.6$.

For a situation to be modelled by the binomial distribution it must satisfy the following conditions:

- There are a fixed number of trials, n.
- Each trial results in one of two outcomes, success or failure.
- The probability of success, p, is constant for each trial (this means the probability of failure is always $q = (1 - p)$).
- The trials are independent.

The letters n and p are the parameters of the binomial distribution. We write this as:

$$X \sim B(n, p)$$

It means that the random variable X has a binomial distribution with parameters n and p (number of trial n and probability of success p).

For example the probability that Rob is late for college is 0.2; we can calculate the probability Rob is late for college a certain number of times a week using a binomial distribution. So for Rob's lateness we would write $X \sim B(5, 0.2)$ then X is the number of times Rob is late out of 5 days.

Example 5

For each of the following random variables, state whether or not a binomial distribution is a suitable model. If it is state the values of n and p. If it is not give a reason why.

a) The number of throws of a fair die required to roll a six.
b) The number of heads observed when tossing a coin 30 times.
c) The number of red counters selected in a sample of 5, without replacement, from a bag containing 20 red and 15 blue counters.

a) Not suitable – the number of throws (n) is not fixed.

b) Suitable – $X \sim B(30, 0.5)$

c) Not suitable – the probability is not constant

Binomial Probability Function

If $X \sim B(n, p)$ then

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \ldots \ldots, n$$

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

This is given in the formula booklet. BPd gives this on graphical calculator.

Example 6

Rebecca has 11 meetings booked with clients on a particular day. The probability that a client fails to turn up for a meeting is 0.12. Find the probability that:

a) All her clients attend their meetings
b) Exactly 8 clients arrive for their meetings
c) At least 10 clients arrive for their meetings

$X \sim B(11, 0.12)$ – fails to turn up

a) $P(x=0) = 0.24508\ldots$
$= 0.245$ (3sf)

b) $P(x=3) = 0.102538\ldots$
$= 0.103$

c) $P(x=0) + P(x=1)$
$= 0.24508 + 0.36762$
$= 0.613$ (3sf)

Binomial Cumulative Distribution

Cumulative binomial probability tables or graphical calculators can be used to find probabilities, $P(X \leq x)$, for certain values of n and p. These tables are included in the formula booklet for when it is not appropriate to use the formula. As the binomial tables or graphical calculators only give probabilities for $P(X \leq x)$ you will need to rearrange your inequalities.

Hint: Draw a number line!

Example 7

A fair ten-sided die is rolled 15 times. Find the probability of:

a) Exactly 3 sixes
b) At least one six
c) Between 3 and 8, inclusive, sixes

$X \sim B(15, 0.1)$

a) $P(x=3) = 0.12850\ldots$
$= 0.129$ (3sf)

b) $0 \; \boxed{1 \; 2 \; 3 \ldots 15}$

$P(x \leq 15) - P(x \leq 0)$
$= 1 - 0.20589..$
$= 0.794$ (3sf)

c) $0 .. 2 \; \boxed{3 \; 4 \ldots 7 \; 8} \; 9 .. 15$

$P(x \leq 8) - P(x \leq 2)$
$= 0.999997 - 0.81593$
$= 0.184$ (3sf)

Example 8

Louise works as a receptionist in a dentist surgery. The probability that she is late for work is 0.15 and is independent from day to day.

a) Find the probability that, she is late:
   i) On at most 4 days during a period of two weeks (10 days)
   ii) On exactly 3 days during a week (5 days)
   iii) More than 10 but fewer than 20 days during an 8 week period.

ai) $X \sim B(10, 0.15)$

$\boxed{0 \; 1 \; 2 \; 3 \; 4} \; 5 \ldots 10$

$P(x \leq 4) = 0.99012..$
$= 0.990$ (3sf)

ii) $X \sim B(5, 0.15)$

$P(x=3) = 0.02438\ldots$
$= 0.0244$ (3sf)

iii) $X \sim B(40, 0.15)$

$0 \ldots 9 \; 10 \; \boxed{11 \ldots 19} \; 20 \; 21 \ldots 40$

$P(x \leq 19) - P(x \leq 10)$
$= 0.9999\ldots - 0.970079$
$= 0.0299$ (3sf)

b) Katy also works at the dentist surgery the probability she is late is 0.7. Calculate the probability that Katy is late on at least 8 days in a period of 4 weeks.

$X \sim B(20, 0.7)$

$0 .. 7 \; \boxed{8 \; 9 \ldots 20}$

$P(x \leq 20) - P(x \leq 7)$
$= 1 - 0.0012789$
$= 0.999$ (3sf)

Binomial Mean and Variance

If $X \sim B(n, p)$ then:

- Mean of X (or μ) $= np$
- Variance of X $(or\ \sigma^2) = np(1-p)$
- Standard deviation of X $(or\ \sigma) = \sqrt{np(1-p)}$

The formula for mean and variance can be found in the formula booklet!

Example 9

$X \sim B(45, 0.12)$

The records at a bank show that on average 12% of credit card applications are completed incorrectly.

a) Calculate the mean and standard deviation for the number of credit card applications that are filled in incorrectly in a random sample of 45 applications.

Mean $= np = 45 \times 0.12 = 5.4$

$SD = \sqrt{np(1-p)} = \sqrt{5.4(1-0.12)} = 2.1799..$

$= 2.18$ (3sf)

b) Mrs Andrews processes 45 applications each day. For a period of two weeks (10 days) she records the number of applications that are completed incorrectly, her results are given below:

4 7 2 3 10 5 4 5 3 8

By calculating the mean and standard deviation of these values, comment with reason, on the suitability of the model $X \sim B(45, 0.12)$ for the number of incorrect applications processed by Mrs Andrews each day.

$\bar{x} = 5.1$

$s = 2.51$ (3sf)

Exam Questions

Every morning before breakfast Laura and Mike play a game of chess. The probability that Laura wins is 0.7. The outcome of any particular game is independent of the outcome of other games. Calculate the probability that, in the next 20 games,

$X \sim B(20, 0.7)$

(i) Laura wins exactly 14 games. [2]

(ii) Laura wins at least 14 games. [2]

i) $P(X = 14) = 0.19163..$

$= 0.192$ (3sf)

ii) $0 \ldots 13 \boxed{14\ 15 \ldots 20}$

$1 - P(X \le 13)$

$= 1 - 0.39199$

$= 0.608$ (3sf)

A biased spinner can only land on one of the numbers 1, 2, 3 or 4. The random variable X represents the number that the spinner lands on after a single spin and $P(X = r) = P(X = r + 2)$ for $r = 1, 2$

Given that $P(X = 2) = 0.35$

(a) find the complete probability distribution of X.

Ambroh spins the spinner 60 times.

(b) Find the probability that more than half of the spins land on the number 4
Give your answer to 3 significant figures.

The random variable $Y = \frac{12}{X}$

(c) Find $P(Y - X \leqslant 4)$

$\frac{1 - 0.35 - 0.35}{2}$

$r = 1$
$P(x=1) = P(x=3)$
$r = 2$
$P(x=2) = P(x=4)$

a)

x	1	2	3	4
$P(x=x)$	0.15	0.35	0.15	0.35

b) $X \sim B(60, 0.35)$ $\qquad$ 0 .. 29 30 $\boxed{31 \ldots 60}$

$P(x > 30) = 1 - P(x \leq 30)$
$= 1 - 0.99411$
$= 0.00589$ (3sf)

c)

x	1	2	3	4
y	12	6	4	3
$y-x$	11	4	1	-1
$P(y-x)$	0.15	0.35	0.15	0.35

$P(Y - X \leq 4) = 0.35 + 0.15 + 0.35$
$= 0.85$

6 Plastic clothes pegs are made in various colours.

The number of red pegs may be modelled by a binomial distribution with parameter p equal to 0.2.

The contents of packets of 50 pegs of mixed colours may be considered to be random samples.

(a) Determine the probability that a packet contains:

(i) less than or equal to 15 red pegs; *(2 marks)*

(ii) exactly 10 red pegs; *(2 marks)*

(iii) more than 5 but fewer than 15 red pegs. *(3 marks)*

(b) Sly, a student, claims to have counted the number of red pegs in each of 100 packets of 50 pegs. From his results the following values are calculated.

Mean number of red pegs per packet = 10.5

Variance of number of red pegs per packet = 20.41

Comment on the validity of Sly's claim. *(4 marks)*

a) $X \sim B(50, 0.2)$

i) $P(x \leq 15) = 0.969196\ldots$
$= 0.969$ (3sf)

ii) $P(x = 10) = 0.1398\ldots$
$= 0.140$ (3sf)

iii) $0 \ldots 4\ 5\ \boxed{6 \ldots 14}\ 15\ 16 \ldots 50$

$P(X \leq 14) - P(x \leq 5)$
$= 0.93928 - 0.04803$
$= 0.891$ (3sf)

b) mean $= np = 50 \times 0.2 = 10$
Variance $= np(1-p) = 10(1-0.2) = 8$

The means are similar. The variances are different.
Sly's claim is not valid.

In an experiment a group of children each repeatedly throw a dart at a target.
For each child, the random variable H represents the number of times the dart hits the target in the first 10 throws.

① The probability of a dart hitting the target is consta

Peta models H as B(10, 0.1)

② Each throw of the darts are independent.

(a) State two assumptions Peta needs to make to use her model.

(b) Using Peta's model, find $P(H \geqslant 4)$

$0 \ldots 3 \, \boxed{4 \; 5 \ldots 10}$

$1 - P(H \leq 3)$

$= 1 - 0.987204 = 0.0128$ (3sf)

For each child the random variable F represents the number of the throw on which the dart first hits the target.

Using Peta's assumptions about this experiment,

(c) find $P(F = 5)$ $= \underbrace{0.9 \times 0.9 \times 0.9 \times 0.9}_{\text{miss 4 times}} \times 0.1 = 0.06561$

Thomas assumes that in this experiment no child will need more than 10 throws for the dart to hit the target for the first time. He models $P(F = n)$ as

$$P(F = n) = 0.01 + (n - 1) \times \alpha$$

where α is a constant.

(d) Find the value of α

(e) Using Thomas' model, find $P(F = 5)$ $= 0.01 + 4\alpha = 0.01 + 0.08 = 0.09$

(f) Explain how Peta's and Thomas' models differ in describing the probability that a dart hits the target in this experiment.

d)

n	1	2	3	4	5	6	7	8	9	10
	0.01	0.01 $+\alpha$	0.01 $+2\alpha$	0.01 $+3\alpha$	0.01 $+4\alpha$	0.01 $+5\alpha$	0.01 $+6\alpha$	0.01 $+7\alpha$	0.01 $+8\alpha$	0.01 $+9\alpha$

$10(0.01) + \alpha + 2\alpha + 3\alpha + \ldots + 9\alpha = 1$

$0.1 + 45\alpha = 1$

$45\alpha = 0.9$

$\alpha = 0.02$

f) Peta's assumes the probability of hitting the target is constant.
Thomas' assumes the probability of hitting the target increases with each attempt.