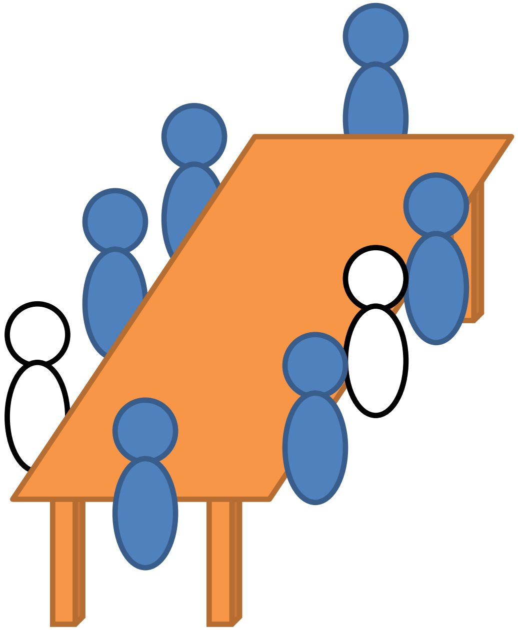
S1 Chapter 6: Statistical Distributions

Binomial Distribution

On holiday in Hawaii visiting the family of a friend, we noticed that at the dinner table that out of the **8 of us, 6 of us were left-handed**.
One of them commented, “The chances of that must be very low”.
“What are the odds?”.



Leftie Example

Let's simplify the problem by using just 3 people:

The probability a randomly chosen person is left-handed is 0.1. If there is a group of 3 people, what is the probability that:

a) All 3 are left-handed.
b) 0 are left-handed.
c) 1 person is left-handed.
d) 2 people are left-handed.

d ?

a ?

b ?

c ?

Let's try to generalise!
If there were x 'lefties' out of 3, then we can see, using the examples, that the probability of a single matching outcome is $0.1^x \times 0.9^{3-x}$. How many rows did we have each time? In a sequence of three L's and R's, there are "3 choose x", i.e. $\binom{3}{x}$ ways of choosing x of the 3 letters to be L's. Therefore the probability of x out of 3 people being left handed is:

$$\binom{3}{x} 0.1^x \, 0.9^{3-x}$$

Leftie Example

Let's simplify the problem by using just 3 people:

The probability a randomly chosen person is left-handed is 0.1. If there is a group of 3 people, what is the probability that:

a) All 3 are left-handed.
b) 0 are left-handed.
c) 1 person is left-handed.
d) 2 people are left-handed.

a $0.1^3 = 0.001$

b $0.9^3 = 0.729$

c As we would do at GCSE, we could list the possibilities than find the probability of each before adding:

LRR: $0.1 \times 0.9 \times 0.9 = 0.081$
RLR: $0.9 \times 0.1 \times 0.9 = 0.081$
RRL: $0.9 \times 0.9 \times 0.1 = 0.081$
$0.081 \times 3 = 0.243$

d Again, listing the possibilities:

LLR: $0.1 \times 0.1 \times 0.9 = 0.009$
RLL: $0.9 \times 0.1 \times 0.1 = 0.009$
LRL: $0.9 \times 0.1 \times 0.9 = 0.009$
$0.009 \times 3 = 0.027$

Let's try to generalise!
If there were x 'lefties' out of 3, then we can see, using the examples, that the probability of a single matching outcome is $0.1^x \times 0.9^{3-x}$. How many rows did we have each time? In a sequence of three L's and R's, there are "3 choose x", i.e. $\binom{3}{x}$ ways of choosing x of the 3 letters to be L's. Therefore the probability of x out of 3 people being left handed is:

$$\binom{3}{x} 0.1^x \, 0.9^{3-x}$$

The Binomial Distribution

✎ You can model a random variable X with a binomial distribution $B(n,p)$ if

- there are a fixed number of trials, n,
- there are two possible outcomes: 'success' and 'failure',
- there is a fixed probability of success, p
- the trials are independent of each other

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

$$P(X=r) = \binom{n}{r} p^r (1-p)^{n-r}$$

In our example, 'success' was 'leftie'.

r is the number of successes out of n.

"$\sim$" means "has the distribution"

On a table of 8 people, 6 people are left handed.

a) Suggest a suitable model for a random variable X: the number of left-handed people in a group of 8, where the probability of being left-handed is 0.1.

b) Find the probability 6 people are left handed.

c) Suggest why the chosen model may not have been appropriate.

a ?

b ?

c ?

In general, choosing a well-known model, such as a Binomial distribution, makes certain **simplifying assumptions**. Such assumptions simplifies the maths involved, but potentially at the expense of not adequately modelling the situation.

The Binomial Distribution

✎ You can model a random variable X with a binomial distribution $B(n,p)$ if

- there are a fixed number of trials, n,
- there are two possible outcomes: 'success' and 'failure',
- there is a fixed probability of success, p
- the trials are independent of each other

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

$$P(X=r) = \binom{n}{r} p^r (1-p)^{n-r}$$

In our example, 'success' was 'leftie'.

r is the number of successes out of n.

"$\sim$" means "has the distribution"

On a table of 8 people, 6 people are left handed.

a) Suggest a suitable model for a random variable X: the number of left-handed people in a group of 8, where the probability of being left-handed is 0.1.

b) Find the probability 6 people are left handed.

c) Suggest why the chosen model may not have been appropriate.

a $X \sim B(8, 0.1)$

b $P(X=6) = \binom{8}{6} 0.1^6\, 0.9^2 = 0.00002268$

c In using a Binomial distribution, we assumed that each person being left handed is independent of each other. However, left-handedness is partially genetic and many people on the table were from the same family.

In general, choosing a well-known model, such as a Binomial distribution, makes certain **simplifying assumptions**. Such assumptions simplifies the maths involved, but potentially at the expense of not adequately modelling the situation.

Further Examples

The random variable $X \sim B\left(12, \frac{1}{6}\right)$. Find:

a) $P(X = 2)$
b) $P(X = 9)$
c) $P(X \leq 1)$

a ?

b ?

c ?

Fro Mental Tip: The two powers add up to n.

Fro Tip: Remember the two 'edge cases':

$$P(X = 0) = (1 - p)^n$$
$$P(X = n) = p^n$$

Edexcel S2 June 2010 Q6

A company claims that a quarter of the bolts sent to them are faulty. To test this claim the number of faulty bolts in a random sample of 50 is recorded.
(a) Give two reasons why a binomial distribution may be a suitable model for the number of faulty bolts in the sample. (2)

?

Further Examples

The random variable $X \sim B\left(12, \frac{1}{6}\right)$. Find:

a) $P(X = 2)$
b) $P(X = 9)$
c) $P(X \leq 1)$

a

$$P(X = 2) = \binom{12}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^{10} = 0.296$$

b

$$P(X = 9) = \binom{12}{9} \left(\frac{1}{6}\right)^9 \left(\frac{5}{6}\right)^3 = 0.0000126$$

c

$$P(X \leq 1) = P(X = 0) + P(X = 1)$$
$$= \left(\frac{5}{6}\right)^{12} + \binom{12}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{11} = 0.381$$

Fro Mental Tip: The two powers add up to n.

Fro Tip: Remember the two 'edge cases':

$$P(X = 0) = (1 - p)^n$$
$$P(X = n) = p^n$$

Edexcel S2 June 2010 Q6

A company claims that a quarter of the bolts sent to them are faulty. To test this claim the number of faulty bolts in a random sample of 50 is recorded.

(a) Give two reasons why a binomial distribution may be a suitable model for the number of faulty bolts in the sample. (2)

(a)	2 outcomes/faulty or not faulty/success or fail A constant probability Independence Fixed number of trials (fixed n)

Test Your Understanding

1

$X \sim B(6, 0.2)$
What is $P(X = 2)$?

?

What is $P(X \geq 5)$?

?

2

I have a bag of 2 red and 8 white balls. X represents the number of red balls I chose after 5 selections (with replacement).

a

How is X distributed?

?

b

Determine the probability that I chose 3 red balls.

?

Test Your Understanding

1

$X \sim B(6, 0.2)$
What is $P(X = 2)$?

$$\boldsymbol{P(X = 2) = \binom{6}{2} 0.2^2\, 0.8^4 = 0.24576}$$

What is $P(X \geq 5)$?

$$\boldsymbol{P(X \geq 5) = P(X = 5) + P(X = 6)}$$
$$\boldsymbol{= \binom{6}{5} 0.2^5 0.8^1 + 0.2^6}$$
$$\boldsymbol{= 0.0016}$$

2

I have a bag of 2 red and 8 white balls. X represents the number of red balls I chose after 5 selections (with replacement).

a

How is X distributed?

$$\boldsymbol{X \sim B(5, 0.2)}$$

b

Determine the probability that I chose 3 red balls.

$$\boldsymbol{P(X = 3) = \binom{5}{3} 0.2^3 0.8^2 = 0.0512}$$

Exercise 6.2

Pearson Applied Year 1/AS
Pages 40-41

Homework Exercise

1 The random variable $X \sim \mathrm{B}\left(8, \frac{1}{3}\right)$. Find:

a $\mathrm{P}(X = 2)$ **b** $\mathrm{P}(X = 5)$ **c** $\mathrm{P}(X \leqslant 1)$

2 The random variable $T \sim \mathrm{B}\left(15, \frac{2}{3}\right)$. Find:

a $\mathrm{P}(T = 5)$ **b** $\mathrm{P}(T = 10)$ **c** $\mathrm{P}(3 \leqslant T \leqslant 4)$

3 A student suggests using a binomial distribution to model the following situations. Give a description of the random variable, state any assumptions that must be made and give possible values for n and p.

a A sample of 20 bolts from a large batch is checked for defects. The production process should produce 1% of defective bolts.

b Some traffic lights have three phases: stop 48% of the time, wait or get ready 4% of the time, and go 48% of the time. Assuming that you only cross a traffic light when it is in the go position, model the number of times that you have to wait or stop on a journey passing through 6 sets of traffic lights.

c When Stephanie plays tennis with Timothy, on average one in eight of her serves is an 'ace'. How many 'aces' does Stephanie serve in the next 30 serves against Timothy?

4 State which of the following can be modelled with a binomial distribution and which cannot. Give reasons for your answers.

a Given that 15% of people have blood that is Rhesus negative (Rh^-), model the number of pupils in a statistics class of 14 who are Rh^-.

b You are given a fair coin and told to keep tossing it until you obtain 4 heads in succession. Model the number of tosses you need.

c A certain car manufacturer produces 12% of new cars in the colour red, 8% in blue, 15% in white and the rest in other colours. You make a note of the colour of the first 15 new cars of this make. Model the number of red cars you observe.

Homework Exercise

5 A balloon manufacturer claims that 95% of his balloons will not burst when blown up. If you have 20 of these balloons to blow up for a birthday party:

a What is the probability that none of them burst when blown up?

b Find the probability that exactly 2 balloons burst.

6 The probability of a switch being faulty is 0.08. A random sample of 10 switches is taken from the production line.

a Define a suitable distribution to model the number of faulty switches in this sample, and justify your choice. **(2 marks)**

b Find the probability that the sample contains 4 faulty switches. **(2 marks)**

7 A particular genetic marker is present in 4% of the population.

a State any assumptions that are required to model the number of people with this genetic marker in a sample of size n as a binomial distribution. **(2 marks)**

b Using this model, find the probability of exactly 6 people having this marker in a sample of size 50. **(2 marks)**

8 A dice is biased so that the probability of it landing on a six is 0.3. Hannah rolls the dice 15 times.

a State any assumptions that are required to model the number of sixes as a binomial distribution. State the distribution. **(2 marks)**

b Find the probability that Hannah rolls exactly 4 sixes. **(2 marks)**

c Find the probability that she rolls two or fewer sixes. **(3 marks)**

Homework Answers

1 **a** 0.273 **b** 0.0683 **c** 0.195

2 **a** 0.00670 **b** 0.214 **c** 0.00178

3 **a** $X \sim B(20, 0.01)$, $n = 20$, $p = 0.01$
Assume bolts being defective are independent of each other.

b $X \sim B(6, 0.52)$, $n = 6$, $p = 0.52$
Assume the lights operate independently and the time lights are on/off is constant.

c $X \sim B(30, \frac{1}{8})$, $n = 30$, $p = \frac{1}{8}$
Assume serves are independent and probability of an ace is constant.

4 **a** $X \sim B(14, 0.15)$ is OK if we assume the children in the class being Rh^- is independent from child to child (so no siblings/twins).

b This is not binomial since the number of tosses is not fixed. The probability of a head at each toss is constant ($p = 0.5$) but there is no value of n.

c Assuming the colours of the cars are independent (which should be reasonable).
X = number of red cars out of 15
$X \sim B(15, 0.12)$

5 **a** 0.358 **b** 0.189

6 **a** The random variable can take two values, faulty or not faulty.
There are a fixed number of trials, 10, and fixed probability of success: 0.08.
Assuming each member in the sample is independent, a suitable model is $X \sim B(10, 0.08)$

b 0.00522

7 **a** Assumptions: There is a fixed sample size, there are only two outcomes for the genetic marker (i.e. fully present or not present), there is a fixed probability of people having the marker.

b 0.0108

8 **a** The random variable can take two values, 6 or not 6. There are a fixed number of trials (15) and a fixed probability of success (0.3), Each roll of the dice is independent. A suitable distribution is
$X \sim B(15, 0.3)$

b 0.219 **c** 0.127