One question they posed in a survey was:

"All families of six children in a city were surveyed. In 72 families, the exact order of births of boys and girls was GBGBBG. What is your estimate of the number of families surveyed in which the exact order of births was BGBBBB?"

The median estimate was 30—suggesting the participants in the survey judged that GBGBBG was more than twice as likely an outcome as BGBBBB. Psychologists have studied this "representativeness fallacy" in research on subjective judgments and biases.

You will learn in this section the mathematics needed to model situations like this; a family of six children can be modelled as a sequence of six independent trials (births) in which the probability of a female birth is constant (0.5) over the six trials.

There are different ways to represent, experience and understand the processes behind quantifying the probabilities in this kind of experiment. Two examples are Pascal's triangle and the Galton board.

Investigation 16

In France in 1665, Blaise Pascal published the pattern partially shown.

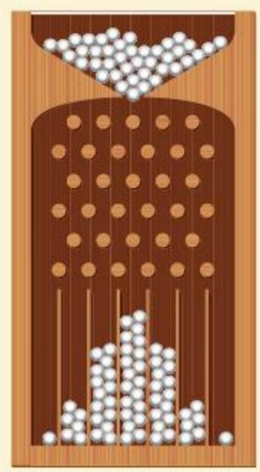
1

1 1

1 2 1

1 3 3 1

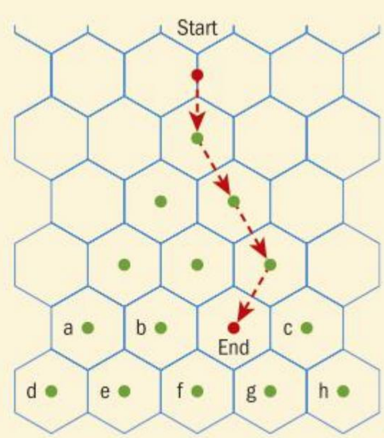
1 4 6 4 1



English polymath Sir Francis Galton invented a machine closely linked to Pascal's triangle.

In the **Galton board**, a ball is dropped into a vertical grid based on a tessellation of regular hexagons. Pegs force the ball to move down to the left or down to the right to reach the next level of the grid as shown.

At the end of the hexagonal grid the balls exit and stack up in vertical "bins". You may explore this process with technology here: https://www.mathsisfun.com/data/quincunx.html or here https://www.geogebra.org/m/ga8J6qDE.



1 At each peg, how many choices are there for the direction of a ball?

2 How many different paths can the ball take to finish in the same place as that shown?

3 Show that the number of ways that the ball can finish in positions a, b, c, ..., h follow the same pattern as Pascal's triangle.

4 Which positions in the Galton board is the ball least likely to finish in?

Continued on next page

5 Which positions in the Galton board is the ball most likely to finish in?

6 What do the numbers in Pascal's triangle count?

7 Why is Pascal's triangle symmetrical?

A convenient way to experience the possible gender outcomes of a family of various sizes is by flipping a coin six times.

The numbers in Pascal's triangle are **binomial coefficients.** You can calculate them with technology.

Make a prediction of the next two rows and their sums and check your answers with technology.

The binomial coefficients in row $(n+1)$ of the triangle are represented by the following notation: $C_0^n,\ C_1^n,\ C_2^n,\ \ldots C_r^n,\ \ldots,\ C_{n-1}^n,\ C_n^n$, where

$C_r^n = \dfrac{n!}{r!(n-r)!}$ and $n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$.

International-mindedness

The Dutch scientist Christian Huygens, a teacher of Leibniz, published the first book on probability in 1657.

For example, $C_2^5 = \dfrac{5!}{2!(5-2)!} = \dfrac{5\times4\times3\times2\times1}{(2\times1)\times3\times2\times1} = 10.$

Carry out 10 trials of flipping a coin six times, recording each trial as a sequence of H (heads) and T (tails). Compare your results with others in your class.

- How many of your trials resulted in outcomes like HTHTTH, which **appear** more random than THTTTT?
- Discuss this claim: "The probability of a total of three heads in six trials is more than the probability of a total of one head in six trials". Justify your answer.

Investigation 17

In this part, represent all probabilities as fractions. Do not simplify the fractions.
A fair coin is tossed twice. Let X be the discrete random variable equal to the number of heads tossed in two trials of the fair coin. Use a tree diagram to represent all the possibilities in the sample space and hence find the probabilities to complete the probability distribution table.

x	0	1	2
$\mathbf{P}(X=x)$			

Repeat the above for three trials and then four trials.

1 **Factual** What is the connection between your results and Pascal's triangle?

2 Predict the probability distribution table for five trials.

3 **Conceptual** Would the probability distribution tables for 0 trials and 1 trial be consistent with the pattern in Pascal's triangle?

You can use your probability distributions from questions **1**, **2** and **3** and the binomial coefficients definition to make general statements for the probability distribution function.

4 Complete: "Let X be the discrete random variable equal to the number of heads tossed in n trials of a fair coin.

Then $\mathrm{P}(X=x) =$ ______________ for $x \in \{0, 1, \ldots, \_\_\}$"

5 The experiment is changed so that the coin is not fair and it is thrown five times. $\mathrm{P}(H) = p$, $\mathrm{P}(T) = 1 - p$.

Complete the general statement "$\mathrm{P}(X=x) =$ ______ for $x \in \{0, 1, \ldots, \_\_\}$"

This investigation leads to the formal definition of the **binomial distribution:**

In a sequence of n independent trials of an experiment in which there are exactly two outcomes "success" and "failure" with constant probabilities $\mathrm{P}(\text{success}) = p$, $\mathrm{P}(\text{failure}) = 1 - p$, if X denotes the discrete random variable equal to the number of successes in n trials, then the probability distribution function of X is

$$\mathrm{P}(X=x) = \mathrm{C}_x^n \, p^x (1-p)^{n-x}, \; x \in \{0, 1, 2, \ldots, n\}$$

These facts are summarized in words as "X is distributed binomially with parameters n and p" and in symbols as $X \sim \mathrm{B}(n, p)$.

You can use the binomial distribution to reflect on the questions at the start of this section.

Reflect Use the binomial distribution to find the probability of having exactly **three** boys in a family of six. Compare and contrast your answer with the probability of the outcome GBGBBG.

Which part of the formula for the binomial distribution counts the number of successes in n trials?

What is true about the number of trials and the probability of success in each binomial experiment?

In examinations, binomial probabilities will be found using technology.

Example 22

A multiple choice quiz has six questions each of which has four equally likely options A, B, C and D to choose from. If answers are guessed and if X represents the number of correct answers, find:

a $P(X = 2)$ **b** $P(X \leq 2)$ **c** $P(X < 2)$ **d** $P(3 \leq X < 6)$.

Assuming the guesses are equally likely and made independently of each other, X can be modelled by a binomial distribution with parameters 6 and $\frac{1}{4}$. Hence $X \sim B(6, \frac{1}{4})$.	Examine the context to see whether it fits the requirements of the binomial distribution. If it does, state the distribution and its parameters clearly before answering the question.
a $P(X = 2) = 0.297$	Use technology to find the probability required. $P(X = x) = C_x^6 0.25^x (0.75)^{6-x}$ Express your answer to three significant figures.
b $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$ $P(X \leq 2) = 0.831$	This event can be expressed as a sum of three mutually exclusive events. This sum can be found faster using technology.
c $P(X < 2) = P(X = 0) + P(X = 1)$ $P(X < 2) = 0.534$	Take care to apply the strict inequality correctly.
d $P(3 \leq X < 6) = P(3 \leq X \leq 5)$ $P(3 \leq X < 6) = 0.169$	Write the inequality in the form needed for applying technology.

It is essential to examine the context of a problem in order to understand whether the binomial distribution is an appropriate model to apply. To determine whether a context can be modelled by a binomial distribution, ask:

- What is the random variable **counting?**

  These **trials** should be independent of each other, which means that knowing what happens in one trial does not change the probabilities in any other trial.
- How many **trials** are there? This is the parameter ***n***.
- What is the **probability of success** in each trial? It should always be the same. This is the parameter ***p***.

TOK

Play the game of the St Petersburg Paradox and decide how much would you pay to play the game.

Example 23

For each situation state if the random variable is distributed binomially. If so, find the probability asked for.

a A coin is biased so that the probability of a head is 0.74. The coin is tossed seven times. A is the number of tails. Find $P(A = 5)$.

b A bag contains 12 white chocolates and 7 dark chocolates. A chocolate is selected at random and its type noted and then eaten. This is repeated five times. B is the number of dark chocolates eaten. Find $P(B = 4)$.

c A bag contains 10 red dice, 1 blue die and 7 yellow dice. A die is selected at random and its colour noted and replaced. This is repeated 12 times. C is the number of yellow dice recorded. Find $P(C \le 6)$.

d Ciaran plays a lottery in which the probability of buying a winning ticket is 0.001. E is the number of tickets Ciaran buys until he wins a prize. Find $P(E < 7)$.

a Each toss of the coin is independent of the others. There are exactly two outcomes and a fixed number of trials. Therefore $A \sim B(7, 0.26)$.	Write down the distribution. This clarifies your thoughts and you will receive method marks in the examination since you have demonstrated your knowledge and understanding.
$P(A = 5) = 0.0137$	Use technology to find the binomial probability. Write the answer to three significant figures.
b Since the probability of selecting a dark chocolate is dependent on what was selected in previous trials, the trials are not independent. Hence the binomial distribution is not an appropriate model for B.	
c Since the die is replaced at each trial, the probability of success is constant and equal to $\frac{7}{18}$. Therefore $C \sim B(12, \frac{7}{18})$.	Use technology to find the binomial probability.
$P(C \le 6) = 0.861$	Write the answer to three significant figures.
d There is not a fixed number of trials nor are the trials independent so the binomial distribution is not an appropriate model for E.	

The binomial distribution can be applied in problem-solving situations.

Example 24

Solve the problems, stating any assumptions and interpretations you make.

a In a family of six children, find

  i the probability that there are exactly three girls

  ii the probability that exactly three consecutive girls are born.

b A study shows that 0.9% of a population of over 4 000 000 carries a virus. Find the smallest sample from the population so that the probability of the sample having no carriers is less than 0.4.

Continued on next page

a i Assuming that boy and girl are the only two outcomes, then the probability of each is 0.5 if the gender of each child is independent of the other. Let G represent the number of girls, then $G \sim \mathrm{B}(6, 0.5)$ and	Understanding the context is the first step in solving the problem. State the distribution.
$\mathrm{P}(G = 3) = 0.3125$	Write down the event and find the probability with technology.
ii Three consecutive girls are born: GGGBBB, BGGGBB, BBGGGB, BBBGGG	Use a diagram to represent the entire sample space.
Each of these four outcomes has probability $\left(\frac{1}{2}\right)^6$, so	
$\mathrm{P}(\text{Three consecutive girls are born}) = 4 \times \left(\frac{1}{2}\right)^6 = \frac{1}{16} = 0.0625$	Both answers can be expressed exactly in this case.
This is considerably smaller than 0.3125 since there are many more combinations of three girls that are not consecutive than are.	Examine your result critically and check that it is feasible.
b In a population of this size the binomial distribution is an appropriate model since sampling without replacement does not alter the probability of choosing a carrier significantly. If C is the number of carriers chosen in a sample of size n then	State the assumptions.
$C \sim B(n, 0.009)$.	Write down the distribution.
Find the smallest value of n so that $\mathrm{P}(C = 0) < 0.4$.	Translate the problem into an inequality.
$n = 102$ is the minimum value required.	Solve using technology.
	Interpret the output from technology. The sequence is decreasing and 102 is the first value of n for which the probability is less than 0.4.

Exercise 7K

1 For each context, determine whether the binomial model can be applied. If it can, write down the distribution in the form $X \sim \mathrm{B}(n, p)$. If it cannot, state why.

a A fair die numbered 1, 2, 3 and 4 is thrown seven times. Find the probability that an even number is obtained exactly three times.

b A box contains six green dice, three red dice and three white dice.

A die is selected at random from the box.

The die's colour is noted and it is not replaced. This experiment is repeated four times. Find the probability that a green die is selected fewer than three times.

c A die is selected at random from the same box, its colour noted and the die replaced. This experiment is repeated four times. Find the probability that a green die is selected fewer than three times.

d Jasmine is packing her bag for a four-day music festival. The weather forecast for each of the days states that the probability of rain is a constant value of 30%. Find the probability that it will rain on at least two of the four days.

e In a soccer match, seven substitute players are available for selection during the match. FIFA rules state that a maximum of three substitutions can be made in a match. Find the probability that exactly three substitutions are made in a match.

2 A fair coin is flipped six times. Calculate the probability of obtaining:

a exactly three heads

b no more than four heads

c at least three heads and fewer than six.

3 Silk scarves are produced in a factory. Quality control investigations find that 0.5% of the scarves produced have flaws, which reduce their value. A sample of 30 scarves is selected from the factory. Calculate the probability that the sample has:

a exactly one flawed scarf

b no flawed scarves

c more than three flawed scarves.

4 Paul works Monday to Friday for a car recovery service in Mathcity. His records show that on any given day, the probability that he is called to recover a car in an area outside the city limits is 0.17. Calculate the probability that:

a In one working week, Paul has to recover a car outside the city limits more than three times.

b Paul has to recover a car outside the city limits more than three times on two consecutive weeks.

5 A weather station in a remote area is powered by 10 solar panels, which operate independently of each other. The manufacturers state that each solar panel has a probability of failure of 0.085% in any given week. The weather station needs at least six of the solar panels to be working in order to function.

a Find the probability that the weather station will not fail in a week.

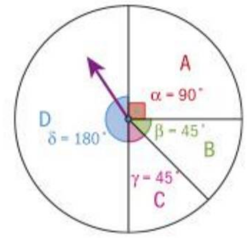
b The weather station is inspected and any necessary maintenance is carried out every six weeks. Find the probability that the weather station will still be functioning when inspected.

6 Alexandre uses his spinning arrow to design another game. Only if the spinner stops in regions A or C is a prize is awarded. If the game is played eight times, find the probability that:

a exactly five prizes are awarded

b fewer than five prizes are awarded

c five prizes or fewer are awarded.



7 Zeke is exploring a biased coin. He tells Francesco that the probability of throwing a head on a coin he has designed is 0.964. However, the probability of throwing exactly four heads with this coin in five trials is approximately the same as the probability of the same event but with a fair coin. Francesco does not believe this is true. Demonstrate that Zeke is correct.

In Section 7.5 you learned about the expected value of a discrete random variable X: $\mathrm{E}(X) = \mu = \sum_x x\mathrm{P}(X = x)$.

You can use this to find the expected value when X is distributed binomially with parameters n and p. In the following investigation there are two other approaches.

Investigation 18

A A subjective perspective

Iana tosses a fair coin 20 times and counts the number of heads.

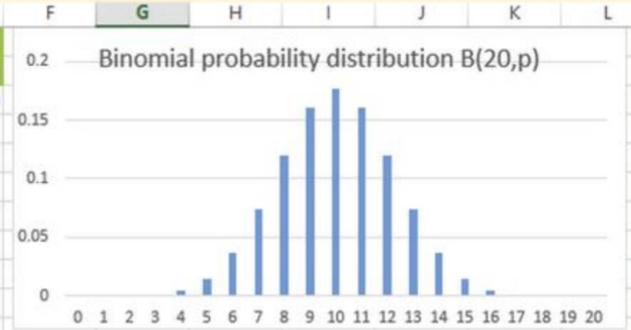
1 **Factual** On average, what is the number of heads she would expect? What if she tossed the coin 32 times? Discuss in a group.

2 **Factual** What if she tosses a biased coin with $\mathrm{P(Head)} = 0.7$ fifty times? What number of heads would she expect?

3 **Conceptual** Can you generalize your answers to make a conjecture on the expected number of heads for n tosses if $\mathrm{P(Head)} = p$? Discuss your justification in a group.

B An experimental perspective

	A	B	C	D	E
1	0	9.53674E-07	0	p=	n=
2	1	1.90735E-05	1.91E-05	0.5	20
3	2	0.000181198	0.000362	mu=	10
4	3	0.001087189	0.003262		
5	4	0.004620552	0.018482		
6	5	0.014785767	0.073929		
7	6	0.036964417	0.221786		
8	7	0.073928833	0.517502		
9	8	0.120134354	0.961075		
10	9	0.160179138	1.441612		
11	10	0.176197052	1.761971		



4 Use a spreadsheet to represent $X \sim \mathrm{B}(n, p)$ as a probability distribution table and a bar chart by following these steps. Save your spreadsheet for use in the next investigation!

- In column A type the numbers 0, 1, 2, …, 20 down to cell A21 to show the possible outcomes of the experiment.
- Fill in cells D1, D2, E1 and E2 as shown above to set up the parameters of the distribution.
- Type "= BINOM.DIST(A1,E2,D2,FALSE)" in cell B1. This is the probability of the event $\mathrm{P}(X=0)$ where $X \sim \mathrm{B}(20, p)$.
- Type "= A1*B1" in cell C1.
- Copy and drag cells B1 and C1 down to row 21 to complete the probability distribution table.
- Type "= sum(C1:C21)" in cell E3 and "mu =" in cell D3.
- Add a chart to display the domain of X on the x-axis and the corresponding probabilities on the y-axis.

5 Which cell displays $E(X)$?

6 Change the values of p and of n to test your conjecture from part (a) **a**.

7 **Factual** How do the parameters of the binomial distribution affect the expected value of the binomial random variable?

8 **Conceptual** How does the expected value of the binomial random variable relate to the formula for the expected number of occurrences?

In statistics you learned about the measures of central tendency (mean, median and mode), and measures of dispersion (range, interquartile range and standard deviation).

You have learned that the mean of the binomial distribution $X \sim B(n, p)$ is $E(X) = np$. Now you will explore its variance. The variance is the square of the standard deviation, and the standard deviation compares each data point with the mean.

TOK

During the mid-1600s, mathematicians Blaise Pascal, Pierre de Fermat and Antoine Gombaud puzzled over this simple gambling problem:

Which is more likely: rolling at least one six on four throws of one dice or rolling at least one double six on 24 throws with two dice?

Investigation 19

A Making visual comparisons of the spread: empirical evidence

With your spreadsheet from the previous investigation, alter the values of p and of n. Hence explore the effect these parameters have on the spread of the distribution. You may find it useful to collaborate with another student to compare and contrast, or just use Ctrl+Z when a parameter has been changed.

- Compare and contrast the spread of $X \sim B(20, 0.15)$ with that of $X \sim B(20, 0.5)$.
- Compare and contrast the spread of $X \sim B(20, 0.15)$ with that of $X \sim B(20, 0.85)$.
- Compare and contrast the spread of $X \sim B(5, 0.85)$ with that of $X \sim B(20, 0.85)$.

1 **Factual** By changing the probability of success, how can you maximize the spread for a fixed number of trials? Describe the reasons why your answer has this effect.

2 **Factual** By changing the number of trials, how can you make the spread greater for a fixed probability of success? Describe the reasons why your answer has this effect.

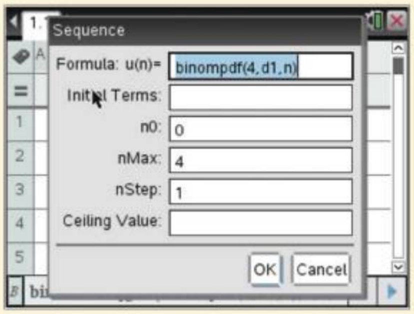
3 **Factual** Which parameters of $X \sim B(n, p)$ affect variance?

B Looking for patterns and making a conjecture: inductive reasoning

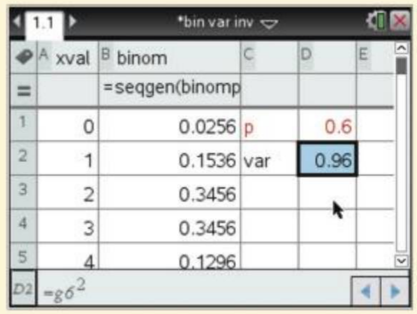
You can investigate the variance with technology to see the effect of changing the probability and keeping the number of trials fixed as follows:

Continued on next page

Statistics and probability



D1 is the cell where the probability parameter can be changed.



Use the steps menu->statistics->stat calculations->one variable statistics. The cell reference G6 is where σx is found.

4 Use technology to collect data to complete the following table. You may repeat for other values of n.

5 Write down a conjecture for the variance of $X \sim \mathrm{B}(4,\ p)$.

6 Write down a conjecture for the variance of $X \sim \mathrm{B}(n,\ p)$.

7 **Factual** Is your conjecture consistent with your answers for **1**, **2** and **3**?

8 **Conceptual** How do the binomial distribution parameters model the spread and central tendency of the binomial distribution?

$X \sim \mathrm{B}(4, p)$	
p	Variance of X
0	
0.1	
0.2	
…	
0.9	
1	

If $X \sim \mathrm{B}(n, p)$ then $\mathrm{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.

Example 25

Two machines in a lightbulb factory are being inspected because quality control raised concerns. Managers have found that the probability that the first machine produces a defective lightbulb is 0.3, and that the probability that the second machine produces a defective bulb is 0.2.

Inspectors take a sample of six bulbs from the first machine and five from the second, and use $X \sim \mathrm{B}(6,\ 0.3)$ and $Y \sim \mathrm{B}(5,\ 0.2)$ to model the number of defective lightbulbs in the samples from the first and second machines respectively.

Compare and contrast the central tendency and spread of these binomially distributed random variables.

$\mathrm{E}(X) = 6 \times 0.3 = 1.8$ and $\mathrm{E}(Y) = 5 \times 0.2 = 1$	Apply the formula $\mathrm{E}(X) = np$.
X has the higher expected value: on average almost twice as many defective bulbs are predicted to appear in the sample of 6 than in the sample of 5.	Interpret your results.

Variance of $X = 6 \times 0.3 \times 0.7 = 1.26$ and Variance of $Y = 5 \times 0.2 \times 0.8 = 0.8$	Apply the formula variance of $X = np(1 - p)$.
There is more spread predicted in the values of the number of defective bulbs from the first machine.	Interpret your results.

Exercise 7L

1 Given $X \sim \mathrm{B}(6, 0.29)$ find the probabilities:

a $\mathrm{P}(X = 4)$ **b** $\mathrm{P}(X \le 4)$ **c** $\mathrm{P}(1 \le X < 4)$

d $\mathrm{P}(X \ge 2)$ **e** $\mathrm{P}(X \le 4 \mid X \ge 2)$.

f Use your answers to determine whether $X \le 4$ and $X \ge 2$ are independent events.

g Find $\mathrm{E}(X)$. **h** Find the variance of X.

2 A fair octahedral die numbered 1, 2, …, 8 is thrown seven times.

Let Q denote the number of prime numbers thrown. Find:

a the probability that at least three prime numbers are thrown

b $\mathrm{E}(Q)$

c the variance of Q.

3 a A biased coin is coloured red on one side and black on the other. The probability of throwing red is 0.78. In 10 throws of the coin, find the probability that:

i exactly three blacks are thrown

ii the number of reds thrown is more than three but fewer than seven.

Let A represent the event "fewer than three blacks are thrown" and B "more than seven blacks are thrown".

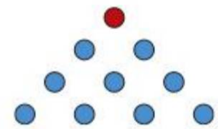
b Find the probabilities $\mathrm{P}(A)$, $\mathrm{P}(B)$, $\mathrm{P}(A \mid B)$.

Hence determine if the events A and B are:

c independent

d mutually exclusive.

4 David plays a game at a fair. He throws a ball towards a pattern of 10 holes in this formation. The aim of the game is to have the ball fall into the red hole to win a point.

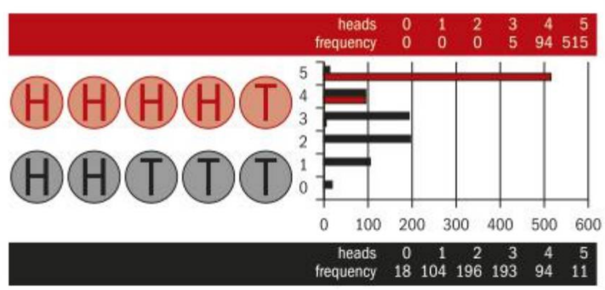


One game consists of throwing 10 balls. Assume David has no skill whatsoever at aiming and that a ball thrown must fall through one of the holes.

a Find the probability that David scores at least five points in a game.

b David plays six games. Find the probability that he scores no points in at least two games.

5 Zeke explores his biased coin in more detail. He designs a spreadsheet to simulate five throws of his biased (red) coin. The probability of a head is 0.964. He also throws a fair (black) coin five times. He collects data for 614 trials as shown below. The image also shows the outcome of the 614th trial.



Let R and B represent the number of heads in five throws of the red coin and five throws of the black coin respectively.

a Find $\mathrm{E}(R)$ and $\mathrm{E}(B)$ and interpret your results.

b Find the variance of R and the variance of B and interpret your results.

6 In a mathematics competition, students try to find the correct answer from five options in a multiple choice exam of 25 questions. Alex decides his best strategy is to guess all the answers.

a State an appropriate model for the random variable A = the number of questions Alex gets correct.

Find the probability that the number of questions that Alex gets correct is:

b at most five **c** at least seven

d no more than three.

e Write down $\mathrm{E}(A)$ and interpret this value.

f Find the probability that Alex scores more than expected.

g In the test, a correct answer is awarded 4 points. An incorrect answer incurs a penalty of 1 point. If Alex guesses all questions, find the expected value of his total points for the examination.

h Four students in total decide to guess all their answers. Find the probability that at least two of the four students will get seven or more questions correct.

7 Calcair buys a new passenger jet with 538 seats. For the first flight of the new jet all 538 tickets are sold. Assume that the probability that an individual passenger turns up to the airport in time to take their seat on the jet is 0.91.

a Write down the distribution of the random variable T = the number of passengers that arrive on time to take their seats, stating any assumptions you make.

b Find $\mathrm{P}(T = 538)$ and interpret your answer.

c Find $\mathrm{P}(T \geq 510)$ and interpret your answer.

d Calcair knows that it is highly likely that there will be some empty seats on any flight unless it sells more tickets than seats. Find the smallest possible number of tickets sold so that $\mathrm{P}(T \geq 510)$ is at least 0.1.

e Determine the number of tickets Calcair should sell so that the expected number of passengers turning up on time is as close to 538 as possible.

f For this number of tickets sold, find $\mathrm{P}(T = 538)$ and $\mathrm{P}(T > 538)$. Interpret your answers.

8 You are given $X \sim \mathrm{B}(n, p)$. Analyse the variance of X as a function of p to find the value of p that gives the most dispersion (spread) of the probability distribution.

Developing inquiry skills

Look back at the opening scenarios.

Can you solve one of the questions in the opening scenario with the binomial distribution? If so, what assumptions would you have to make?

7.7 Modelling measurements that are distributed randomly

In this section you will model an example of a **continuous** random variable.

For example, consider the height Y metres of an adult human chosen at random. Recall from Section 7.5 that Y is a measurement and would therefore be an example of **continuous** data.

We use the following terminology for random variables that are found by measuring: