Table of Contents

Subtopic 2.3: Binomial distribution

Key questions and key concepts	Considerations for developing teaching and learning strategies
What is a binomial distribution and how is it related to Bernoulli trials? • When a Bernoulli trial is repeated, the number of successes is classed as a binomial random variable • The possible values for the different numbers of successes and their probabilities make up a binomial distribution	Students use tree diagrams to build simple binomial distributions such as the number of heads resulting when a coin is tossed three times.
What is the mean and standard deviation of the binomial distribution? • The mean of the binomial distribution is *np*, and the standard deviation is given by $\sqrt{np(1-p)}$, where *p* is the probability of success in a Bernoulli trial and *n* is the number of trials	Both of these expressions can be built by considering a binomial variable as the sum of *n* Bernoulli variables. The mean and standard deviation of the Bernoulli distribution are used to obtain these expressions. In this subject, the principal purpose of these values is to determine the mean and standard deviation for the distribution of sample proportions in Topic 6: Sampling and confidence intervals.
When can a situation be modelled using the binomial distribution? • A binomial distribution is suitable when the number of trials is fixed in advance, the trials are independent, and each trial has the same probability of success	Students classify scenarios as ones that can or cannot be modelled with a binomial distribution. In a context where sampling is done without replacement (e.g. an opinion poll), the distribution is not strictly binomial. When the population is large in comparison with the sample, the binomial distribution provides an excellent approximation of the probability of success.
How can binomial probabilities be calculated? • The probability of *k* successes from *n* trials is given by $\Pr(X=k)=C_k^n p^k (1-p)^{n-k}$, where *p* is the probability of success in the single Bernoulli trial	The tree diagram used to determine the probability of obtaining a 6 on a dice twice when the dice is rolled three times, is considered to show that the calculation of binomial probabilities involves: • calculation of the number of ways that *k* successes can occur within *n* events (using the concepts in Topic 4: Counting and statistics in Stage 1 Mathematics) • probability of success to the power of the required number of successes • probability of failure to the power of the required number of failures.

Key questions and key concepts	Considerations for developing teaching and learning strategies
How can electronic technology be used to calculate binomial probabilities?	Although an understanding of the binomial probability formula is a requirement of this subject, calculations of binomial probabilities in problem-solving should be done using electronic technology. Students, given the probability of success, calculate probabilities such as: • exactly *k* successes out of *n* trials • at least *k* successes out of *n* trials • between k_1 and k_2 successes out of *n* trials
What happens to the binomial distribution as *n* gets larger and larger? • The binomial distribution for large values of *n* has a symmetrical shape that many students will recognize	Students will not have been formally introduced to the normal distribution yet, but it may be familiar to many. Probability bar charts for binomial distributions of $n = 10$, $n = 100$, $n = 1000$ are built using a spreadsheet (for a specific value of *p*). Comparing these graphs shows the shape approaching a normal distribution. Discussion on adding a smooth curve to emphasize the shape leads to the concept of continuous random variables.

1 Binomial distribution

When a Bernoulli trial is repeated a number of times, we have a Binomial distribution. A Binomial distribution is characterised by the following rules

The binomial random variable X is the total number of successes in n trials.

- It is made up of n Bernoulli trials or n identical trials.
- Each trial is an independent trial.
- There are two possible outcomes for each trial, a success, p, and a failure, 1 – p.

If there are 'n' independent trials, then the probability that there are 'k' successes and '(n – k)' failures is

$$\boxed{Pr(X = k) = C_x^n p^k (1-p)^{n-k}}$$

$$\boxed{Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}}$$

where

$Pr(X = k)$ is the binomial probability distribution function

$C_k^n = \binom{n}{k}$ Number of ways the 'k' successes can be ordered amongst the n trials

$$\boxed{C_k^n = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}}$$

$p^k(1-p)^{n-k}$ Probability of obtaining 'k' succeses and (n - k) failures in a particular order.

If X is the random variable of a binomial experiment with parameter n and p, then we write X ~ B(n,p)

Note

If the order is specified for a particular scenario, then the binomial probability distribution rule cannot be used. The probabilities need to be multiplied in the given order.

Example 1

(a) Expand $\left(\frac{2}{3}+\frac{1}{3}\right)^4$

$$\left(\frac{2}{3}+\frac{1}{3}\right)^4=\underbrace{\left(\frac{2}{3}\right)^4}_{4\,sb}+\underbrace{4\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^1}_{3\,sb+1\,al}+\underbrace{6\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2}_{2\,sb+2\,al}+\underbrace{4\left(\frac{2}{3}\right)^1\left(\frac{1}{3}\right)^3}_{1\,sb+3\,al}+\underbrace{\left(\frac{1}{3}\right)^4}_{4\,al}$$

(b) A box of choclates contain strawberry crems and almond centres in the ratio 2:1. Four chocolates are selected at random, with repalcement. Find the probability of getting

(i) all strawberry crems

Getting a strawberry is considered a success here, and p = 2/3, and failure is 1/3, and n = 4.

Pr(X = 4) $= \binom{4}{4}(\frac{2}{3})^4(1/3)^0$

$= (2/3)^4$ $= 16/81$

(ii) two of each type

Pr(X = 2) $= \binom{4}{2}(\frac{2}{3})^2(1/3)^2$

$= 6\,(2/3)^2\,(1/3)^2$ $= 8/27$

(iii) atleast two strawberry

Pr(X ≥ 2) = Pr(X = 2) + Pr(X = 3) + Pr(X = 4)

$= \binom{4}{2}(\frac{2}{3})^2(1/3)^2 + \binom{4}{3}(\frac{2}{3})^3(1/3)^1 + \binom{4}{4}(\frac{2}{3})^4(1/3)^0$

= 8/27 + 4(8/27)(1/3) + 16/81

= 8/27 + 32/81 + 16/81

= 72/81 = 8/9

Example 2

Find the probability of obtaining exactly three heads when a fair coin is tossed seven times, correct to four decimal places.

Obtaining a head is considered a success here, and the probability of success on each of the seven independent trials is 0.5.

Let X be the number of heads obtained. The parameters are p = 0.5, (1- p) = 0.5, and n = 7.

If X is the random variable of a binomial experiment, then we write X ~ B(7, 0.5)

Pr(X = 3) $= \binom{7}{3}0.5^3(0.5)^{7-3}$

$= 35\,(0.5)^7$ $= 0.2734$

Example 3

The probability that a person currently in prison has ever been imprisoned before is 0.72. Find the probability that of five prisoners chosen at random at least three have been imprisoned before, correct to four decimal places.

Considering a person prisoned is success here, and the probability is 0.72.

Let X be the number of persons prisoner. The parameters are p = 0.72, (1- p) = 0.28, and n = 5.

$\Pr(X \geq 3) \quad = \Pr(X = 3) + \Pr(X = 4) + \Pr(X = 5)$

$= \binom{5}{3}0.72^3(0.28)^2 + \binom{5}{4}0.72^4(0.28)^1 + \binom{5}{5}0.72^5(0.28)^0 \quad = 0.8624$

Example 4

Records show that x% of people will pass their driver's license on the first attempt. If six students attempt their driver's license, write down in terms of x the probability that:

(a) all six students pass

Let X be the number of people passed. The parameters are p = x/100, (1- p) = 1 - x/100, and n = 6.

$\Pr(X = 6) \quad = \binom{6}{6}\left(\frac{x}{100}\right)^6(1 - x/100)^0 \quad = \frac{x^6}{100^6}$

(b) only one fails

$\Pr(X = 5) \quad = \binom{6}{5}\left(\frac{x}{100}\right)^5(1 - x/100)^1$

$= \frac{6x^5}{100^5}\left(\frac{100-x}{100}\right)$

$= \frac{6x^5(100-x)}{100^6}$

(c) no more than two fail.

$\Pr(X \geq 4) \quad = \Pr(X = 4) + \Pr(X = 5) + \Pr(X = 6)$

$$= \frac{x^6}{100^6} + \frac{6x^5(100 - x)}{100^6} + \frac{15x^4(100 - x)^2}{100^6}$$

Note

We can also use the binomial distribution to solve problems involving conditional probabilities.

Example 5

The probability of a netballer scoring a goal is 0.3. Find the probability that out of six attempts the netballer scores a goal:

(a) four times

Let X be the number of goals scored.

Then X has a binomial distribution with n = 6 and p = 0.3.

Pr(X = 4) $= \binom{6}{4}0.3^4(0.7)^2$

= 15 * 0.0081 * 0.49

= 0.059535

(b) four times, given that she scores at least one goal.

Pr(X = 4|X ≥ 1) = Pr(X = 4 ∩ X ≥ 1) / Pr(X ≥ 1)

= Pr(X = 4) / Pr(X ≥ 1)

= 0.059535 / (1 - 0.7^6) [Since Pr(X ≥ 1) = 1 - Pr(X = 0)]

= 0.0675

Example 6

The manager of a shop knows from experience that 60% of her customers will use a credit card to pay for their purchases. Find the probability that:

(a) the next three customers will use a credit card, and the three after that will not

Let X be the number of customers using credit card.

Then X has a binomial distribution with n = 3 and p = 0.6 and (1 - p) = 0.4

Pr(X = 3) = $\binom{3}{3}0.6^3(0.4)^3$

= 1 * 0.216 * 0.064 = 0.0138

(b) three of the next six customers will use a credit card

Let X be the number of customers using credit card.

Then X has a binomial distribution with n = 6 and p = 0.6 and (1 - p) = 0.4

Pr(X = 3) = $\binom{6}{3}0.6^3(0.4)^3$

= 20 * 0.216 * 0.064 = 0.2765

(c) at least three of the next six customers will use a credit card

Pr(X ≥ 3) = Pr(X = 3) + Pr(X = 4) + Pr(X = 5) + Pr(X = 5)

= $\binom{6}{3}0.6^3(0.4)^3 + \binom{6}{4}0.6^4(0.4)^2 + \binom{6}{5}0.6^5(0.4)^1 + \binom{6}{6}0.6^6(0.4)^0$

= 0.8208

(d) exactly three of the next six customers will use a credit card, given that at least three of the next six customers use a credit card.

Pr(X = 3|X ≥ 3) = Pr(X = 3 ∩ X ≥ 3) / Pr(X ≥ 3)

= Pr(X = 3) / Pr(X ≥ 3)

= 0.2765 / 0.8202

= 0.3368

Example 7

It is known that 52% of the population participates in sport on a regular basis. Five random individuals are interviewed and asked whether they participate in sport on a regular basis. Let X be the number of people who regularly participate in sport.

(a). Construct a probability distribution table for X.

$X \sim Bi(5, 0.52)$

$P(X = k) = {}^{n}C_{k}\, p^{k}\, (1 - p)^{n-k}$

$P(X = 0) = {}^{5}C_{0}(0.52)^{0}(0.48)^{5} \quad = 0.02548$

$P(X = 1) = {}^{5}C_{1}(0.52)^{1}(0.48)^{4} \quad = 0.13802$

$P(X = 2) = {}^{5}C_{2}(0.52)^{2}(0.48)^{3} \quad = 0.29904$

$P(X = 3) = {}^{5}C_{3}(0.52)^{3}(0.48)^{2} \quad = 0.32396$

$P(X = 4) = {}^{5}C_{4}(0.52)^{4}(0.48)^{1} \quad = 0.17548$

$P(X = 5) = {}^{5}C_{5}(0.52)^{5}(0.48)^{0} \quad = 0.03802$

x	P(X = k)
0	0.02548
1	0.13802
2	0.29904
3	0.32396
4	0.17548
5	0.03802
ΣP(X = k)	1

(b). Determine the probability that 3 people or less play sport.

$P(X \le 3) \quad = Pr(X = 0) + Pr(X = 1) + Pr(X = 2) + Pr(X = 3)$

$= 0.02548 + 0.13802 + 0.29904 + 0.32396$

$= 0.7865$

(c). Determine the probability that at least one person plays sport, given that no more than 3 people play sport.

$P(X \ge 1 | X \le 3) \quad = Pr(X \ge 1 \cap X \le 3) / Pr(X \le 3)$

$= Pr(X = 1) + Pr(X = 2) + Pr(X = 3) / Pr(X \le 3)$

$= 0.13802 + 0.29904 + 0.32396 / 0.7865$

$= 0.9676$

(d). Determine the probability that the first person interviewed plays sport but the next 2 do not.

Order has been specified for this question. Therefore, the binomial probability distribution rule cannot be used. The probabilities must be multiplied together in order.

Let S = plays sport, N = does not t play sport

$P(SNN) \quad = P(S) \times P(N) \times P(N)$

$= 0.52 \times 0.48 \times 0.48$

$= 0.1198$

Example 8

The probability of an Olympic archer hitting the centre of the target is 0.7. What is the smallest number of arrows he must shoot to ensure that the probability he hits the centre at least once is more than 0.9?

$X \sim Bi(n, 0.7)$

$P(X \geq 1) > 0.9$

$P(X \geq 1) = 1 - P(X = 0)$

The upper limit of successes is unknown because 'n' is unknown. Therefore, $P(X \geq 1)$ cannot be found by adding up the probabilities. However, the required probability can be found by subtracting from 1 the only probability not included in $P(X \geq 1)$.

$1 - P(X = 0) > 0.9$

$1 - {}^{n}C_{0}(0.7)^{0}(0.3)^{n} > 0.9$

$1 - 1 * 1 * (0.3)^{n} > 0.9$ $\qquad {}^{n}C_{0} = n!/[(n - 0)! * 0!] = 1$

$1 - (0.3)^{n} > 0.9$

$1 - 0.9 > (0.3)^{n}$

$\log_{10} 0.1 > \log_{10}(0.3)^{n}$

$\log_{10} 0.1 > n \log_{10}(0.3)$

$n > \log_{10}0.1/\log_{10}(0.3)$

$n > 1.9124$

n = 2 (as n must be an integer) The smallest number of arrows the archer needs to shoot in order to guarantee a probability of 0.9 of hitting the centre is 2.

2 The mean and variance of the binomial distribution

For a binomial distribution X~Bi (n, p), the mean is calculated simply from the following equation:

$$\boxed{E(X) = \mu = np}$$

The variance of the binomial distribution X~Bi (n, p) is calculated using the equation

$$\boxed{Var(X) = \sigma^2 = np(1 - p)}$$

The standard deviation of the binomial distribution X~Bi (n, p) is calculated using the equation

$$\boxed{SD(X) = \sigma = \sqrt{[np(1 - p)]}}$$

Example 1

A test consists of 20 multiple choice questions, each with 5 alternatives for the answer. A student has not studied for the test, so she chooses the answers at random. Let X be the discrete random variable that describes the number of correct answers.

a. Find the expected number of correct questions answered.

$E(X) = \mu \quad = np$

$\quad = 20*(1/5) \quad = 4$

The expected number of questions correct is 4.

b. Find the variance of the correct number of questions answered.

$Var(X) = \sigma^2 \quad = np(1 - p)$.

$\quad = 20(1/5)(4/5) \quad = 3.2$

Example 2

A binomial random variable, Z, has a mean of 8.4 and a variance of 3.696.

a. Find the probability of success, p.

$E(Z) = \mu \quad = np \quad = 8.4$ ------ ①

$Var(X) = \sigma^2 \quad = np(1 - p). \quad = 3.696$ ------ ②

②/① $\quad = np(1 - p)/ np \quad = 3.696/8.4$

$(1 - p) \quad = 0.44$ and $p \quad = 0.56$

The probability of success is 0.56.

b. Find the number of trials, n.

$\mu = np \rightarrow 8.4 = n \times 0.56 \rightarrow n \quad = 15$

There are 15 trials.

Example 3

An examination consists of 30 multiple-choice questions, each question having three possible answers. If a student guesses the answer to every question, how many will she expect to get right? Discuss the result.

Since the number of correct answers is a binomial random variable, with parameters p = 1/3 and n = 30, the student who guesses has an expected result of μ = np = 10 correct answers.

(not enough to pass if the pass mark is 50%).

Example 4

The probability of contracting influenza this winter is known to be 0.2. Of the 100 employees at a certain business how many would the owner expect to get influenza? Find the standard deviation of the number who will get influenza and calculate $\mu \pm 2\sigma$.

Interpret the interval $[\mu - 2\sigma, \ \mu + 2\sigma]$ for this example.

The number of people who get influenza is a binomial random variable, with parameters p = 0.2 and n = 100.

The owner will expect μ = np = 20 of her employees to contract influenza, with a

variance $\sigma^2 = np(1 - p) = 16$ and hence a standard deviation

$\sigma = 4$.

Then $\mu \pm 2\sigma \quad = 20 \pm (2 \times 4)$

$\qquad\qquad\quad = 20 \pm 8$ or from 12 to 28

Thus, the owner of the business knows there is a probability of about 0.95 that from 12 to 28 of her employees will contract influenza this winter.

3 Application of the binomial distribution

The binomial distribution has important applications in medical research, quality control, simulation, and genetics. In this section we will explore some of these areas.

3.1 Finding the sample size

While we can never be absolutely certain about the outcome of a random experiment, sometimes we are interested in knowing what size sample would be required to observe a certain outcome. For example, how many times do you need to roll a die to be reasonably sure of observing a six, or how many lotto tickets must you buy to be reasonably sure that you will win a prize?

Example 1

The probability of winning a prize in a game of chance is 0.48.

(a) What is the least number of games that must be played to ensure that the probability of winning at least once is more than 0.95?

Since the probability of winning each game is the same each time the game is played, this is an example of a binomial distribution, with the probability of success $p = 0.48$.

The required answer is the smallest value of n such that $\Pr(X \geq 1) > 0.95$.

$\Pr(X \geq 1) \quad > 0.95$

$\Leftrightarrow 1 - \Pr(X = 0) \quad > 0.95$

$\Leftrightarrow \Pr(X = 0) \quad < 0.05$

$\Leftrightarrow 0.52^n \quad < 0.05 \quad [\text{since } \Pr(X = 0) = 0.52^n]$

This can be solved by taking logarithms of both sides:

$\ln(0.52^n) \quad < \ln(0.05)$

$n \ln(0.52) \quad < \ln(0.05)$

$n \quad < \ln(0.05)/\ln(0.52) = 4.58$

Thus, the game must be played at least five times to ensure that the probability of winning at least once is more than 0.95.

(b) What is the least number of games that must be played to ensure that the probability of winning at least twice is more than 0.95?

The required answer is the smallest value of n such that $\Pr(X \geq 2) > 0.95$, or equivalently, such that $\Pr(X < 2) \quad < 0.05$. We have

$\Pr(X < 2) \quad = \Pr(X = 0) + \Pr(X = 1)$

$= {}^{n}C_0\,(0.48)^0\,(0.52)^n + {}^{n}C_1\,(0.48)^1\,(0.52)^{n-1}$

$= (0.52)^n + 0.48n(0.52)^{n-1}$

So, the answer is the smallest value of n such that

$(0.52)^n + 0.48n(0.52)^{n-1} < 0.05$

This equation cannot be solved algebraically; but a graphics calculator can be used to find the solution $n > 7.7985 \ldots$.

Thus, the game must be played at least eight times to ensure that the probability of winning at least twice is more than 0.95.

Example 2

It has been found that 9% of the population have diabetes. A sample of 15 people were tested for diabetes. Let X be the random variable that gives the number of people who have diabetes.

(a). Determine $P(X \leq 5)$.

$n = 15$ and $p = 0.09$ and $(1 - p) = 0.91$

$X \sim Bi(15, 0.09)$ Let X = number of people who have diabetes.

Therefore, $0 \leq k \leq 5$.

$$\begin{aligned}
Pr(X \leq 5) &= Pr(X = 0) + Pr(X = 1) + Pr(X = 2) + Pr(X = 3) + Pr(X = 4) + Pr(X = 5) \\
&= {}^{15}C_0\,(0.09)^0\,(0.91)^{15} + {}^{15}C_1\,(0.09)^1\,(0.91)^{14} + {}^{15}C_2\,(0.09)^2\,(0.91)^{13} \\
&\quad + {}^{15}C_3\,(0.09)^3\,(0.91)^{12} + {}^{15}C_4\,(0.09)^4\,(0.91)^{11} + {}^{15}C_5\,(0.09)^5\,(0.91)^{10} \\
&= 0.243\,008 + 0.360\,507 + 0.249\,582 + 0.106\,964 + 0.031\,736 + 0.006\,905 \\
&= 0.9987
\end{aligned}$$

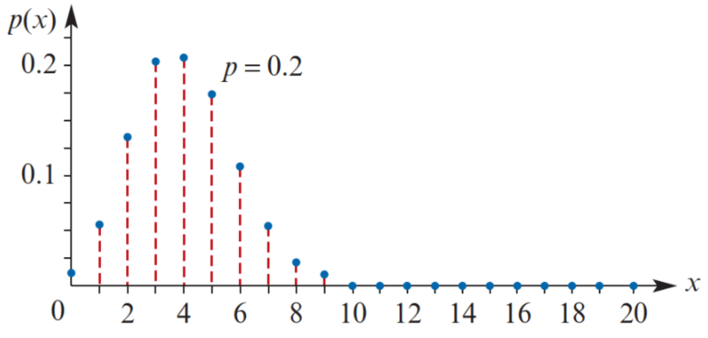
The probability that 5 or fewer people of the 15 selected have diabetes is 0.9987.

(b). Determine E(X) and SD(X).

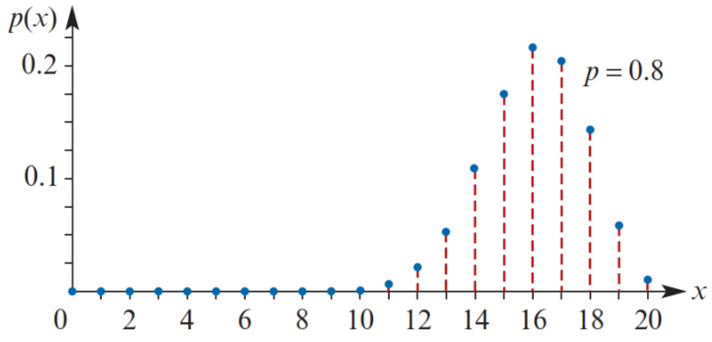
$$\begin{aligned}
E(X) &= np \\
E(X) &= 15 \times 0.09 && = 1.35 \\
Var(X) &= np(1 - 9) \\
&= 15 \times 0.09 \times 0.91 && = 1.2285 \\
SD(X) &= \sqrt{Var(X)} && = 1.1084
\end{aligned}$$

4 The graph of the binomial probability distribution

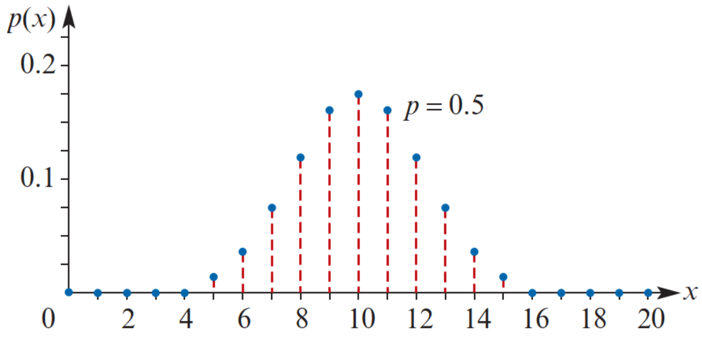
The shape of the graph of the binomial probability function depends on the values of n and p.



The shape of this graph is positively skewed. It indicates that the probability of success is low (as p = 0.2), as the larger x-values (number of successful outcomes) have corresponding low probabilities.



The shape of this graph is negatively skewed. It indicates that the probability of success is high (as p = 0.8), as the larger x-values (number of successful outcomes) have corresponding high probabilities.



The shape of this graph is symmetrical. It indicates that the probability of success (as p = 0.5) is equal to the probability of failure.

If the number of trials increased, the graph would approach the shape of a bell-shaped curve.

5 Practice

Question 1

Executives in the New Zealand Forestry Industry claim that only 5% of all old sawmills sites contain soil residuals of dioxin (an additive previously used for anti-sap-stain treatment in wood) higher than the recommended level. If Environment Canterbury randomly selects 20 old sawmill sites for inspection, assuming that the executive claim is correct:

a) Calculate the probability that less than 1 site exceeds the recommended level of dioxin.

b) Calculate the probability that less than or equal to 1 site exceed the recommended level of dioxin.

c) Calculate the probability that at most (i.e., maximum of) 2 sites exceed the recommended level of dioxin.

Question 2

Inland Revenue audits 5% of all companies every year. The companies selected for auditing in any one year are independent of the previous year's selection.

a) What is the probability that the company 'Ross Waste Disposal' will be selected for auditing exactly twice in the next 5 years?

b) What is the probability that the company will be audited exactly twice in the next 2 years?

c) What is the exact probability that this company will be audited at least once in the next 4 years?

Question 3

The probability that a driver must stop at any one traffic light coming to Lincoln University is 0.2. There are 15 sets of traffic lights on the journey.

a) What is the probability that a student must stop at exactly 2 of the 15 sets of traffic lights?

b) What is the probability that a student will be stopped at 1 or more of the 15 sets of traffic lights?

Answers

Question 1

(a) 0.3585 (b) 0.7359 (c) 0.9246

Question 2

(a) 0.0214 = 2% (b) 0.0025 = 0.25% (c) 0.1854

Question 3

(a) 0.2309 = 23% (b) 0.9648 = 97%

Reference

Book	Chapter	Exercise
Yr 12 MM VCE Units 3 & 4 (2016e MQ Jacaranda) Solution	11 Binomial Distributions	Ex 11.3 & Ex 11.4
Yr 12 MM QLD Unit 3 & 4 (2018e MQ Jacaranda)	10 Discrete random variables	Ex 10.4 to Ex 10.7
Yr 12 MM (Haese)	07 Discrete random variables	Ex 7G
Yr 12 MM VCE Unit 3 & 4 (2018e MQ Jacaranda)	11 Discrete random variables	Ex 11.3 to Ex 10.5