Chapter **S4**

Binomial Distribution

At the end of this chapter, students should be able to
- relate that binomial distribution is an example of a discrete probability distribution;
- use the binomial distribution, $B(n, p)$, as a probability model to model practical situations;
- recognise conditions under which the binomial distribution is a suitable model and comment on the appropriate use of a model and the assumptions made;
- calculate binomial probabilities using a graphic calculator;
- calculate the mean and variance of a binomial distribution.

4.1 The Binomial Distribution

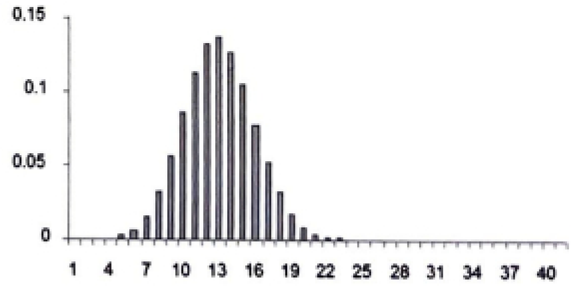
Consider an experiment with a fixed number, n, of independent repeated trials, with two possible outcomes, of which one is defined as '**success**' and the other '**failure**'. Suppose the probability of success is a constant, p, and the probability of failure is $1-p$ (or q).

Let X be the random variable denoting the number of successes occurring in the n independent trials such that the probability of success for each trial is p. Then X can take the values $0, 1, 2, \ldots, n$, so that X is a discrete random variable.

If X is distributed in this way, we write $X \sim B(n,p)$ to denote that the random variable X has a binomial (probability) distribution with **parameters** n and p, where n is the number of independent trials and p is the probability of success. The probability density function (p.d.f.) of X is given by

$$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}\text{, where } x=0, 1, 2, \ldots, n. \quad \textbf{(In MF 26)}$$

For example, if X is the random variable denoting the number of successes occurring in the 40 independent trials such that the probability of success for each trial is 0.3, we write $X \sim B(40, 0.3)$. We can also use a diagram to illustrate the respective probabilities of the distribution.



$X \sim B(40, 0.3)$

As you progress through the chapters, you will learn about other distributions. Thus, it is necessary to correctly identify the random variable and its distribution to ensure that we model a particular situation using the correct probability model. You are also required to justify the appropriateness of the model and the assumptions made **in the context of the question**.

Necessary conditions for a binomial distribution to be an appropriate model: 1. The experiment has a **<u>fixed number</u>**, n, of repeated trials, where $n \in \mathbb{Z}^+$. 2. There are **<u>2 mutually exclusive outcomes</u>**, namely success and failure. 3. The **<u>outcome</u>** of each trial **<u>is independent</u>** of the outcomes of other trials. 4. The **<u>probability of success</u>**, p, **<u>is constant</u>** for each trial, where $0 \le p \le 1$.

Note:

1. **All** conditions must be fulfilled before a random variable can be well-modelled by a binomial distribution.
2. When a question asks for assumptions needed for a random variable to be well-modelled by a binomial distribution, **write down the conditions that you cannot find in the question as assumptions**.
3. All assumptions must be written **in the context of the question**.

Example 1

For each of the random variable described below, state whether the binomial distribution is a suitable model. If it is, justify your answer and give the appropriate values for the parameters. If it is not, give **one** reason.

(i) The number of heads when a fair coin is tossed five times.

(ii) The number of draws, out of 8, (without replacement) a red ball is drawn from a bag containing 10 red and 15 white balls.

(iii) The number of draws, out of 8, (with replacement) a red ball is drawn from a bag containing 10 red and 15 white balls.

(iv) The number of tosses of a coin to get a head.

Solution	**Think Zone**
(i) Yes. - There are $n=5$ repeated tosses - There are 2 mutually exclusive outcomes, namely getting a head or tail. - The event of getting a head in each toss does not affect the event of getting heads in other tosses hence the outcomes are independent. The probability of getting a head, $p=\frac{1}{2}$ is constant	To check: - Does the experiment have a fixed number of repeated trials? - How many possible outcomes are there in each experiment? Are they mutually exclusive? - Are the outcomes of each trial *independent*? - Is the probability of success a constant? In **(ii)**, what if the total number of balls in the bag was large?
(ii) No, p is not constant.	
(iii) Yes. • There are $n = 8$ repeated draws • There are 2 mutually exclusive outcomes, namely drawing a red ball or white ball. • The event of getting a red ball in each draw does not affect the event of getting red balls in other draws hence the outcomes are independent.	

• The probability of drawing a red ball, $p = \frac{2}{5}$, is constant.	
(iv) No, number of trials is not fixed	

Example 2

When a machine is used to dig up potatoes, there is a probability of 0.4 for each individual potato to be damaged in the process. Five potatoes are randomly chosen. The random variable X denotes the number of potatoes in the sample which are damaged. Identify the distribution of X, justifying your answer.

Solution	Think Zone
• Since there are 5 potatoes in the sample, there is a fixed number of repeated trials, i.e, $n=5$, • There are 2 mutually exclusive outcomes, i.e, the potato is damaged or the potato is not damaged. • The event that a potato is damaged is independent of another potato being damaged. • The probability of choosing a damaged potato, $p=0.4$, is constant for each trial. Hence $X \sim \mathrm{B}(5,0.4)$	Is there a maximum value for X? What are the possible outcomes when a potato from the sample is randomly chosen? Are the outcomes independent? Is the probability of choosing a damaged potato constant throughout?

Example 3 (2012/II/9 part)

In an opinion poll before an election, a sample of 30 voters is obtained. The number of voters in the sample who support the Alliance Party is denoted by A. State, in context, what must be *assumed* for A to be well-modelled by a binomial distribution.

Solution	Think Zone
The voters' opinions should be independent Each voter must have the same probability of supporting the Alliance party.	Implied assumptions in the question: 1. Number of voters is fixed. 2. Voters either support or do not support the Alliance Party.

Self Review 1 (2009/II/11 part)

A fixed number, n, of cars is observed and the number of those cars that are red is denoted by R. State, in context, *two assumptions* needed for R to be well modelled by a binomial distribution.

colour of the n cars must be independent of the colour of any other cars.

probability of red is the same throughout the sample.

$=0.136$ (3sf) h) Let Y be the random variable denoting the number of customers who buy chicken burgers, out of 26 customers $Y \sim B(26, 0.09)$ P(4 customers out of 26 customers buy chicken burger and 27th customer buys chicken burger) ~~required prob~~ $= P(Y=4)(0.09)$ $= 0.0111$ (3sf)	26 customers — 27th customer 4 bought chicken burger — fifth customer to buy chicken burger

Example 9 (Question involving binomial distribution within a binomial distribution)

In a large population, the proportion of people having blood group A is 35%. Specimens of blood from the first five people attending a clinic are to be tested. It can be assumed that these five people are a random sample from the population. The random variable X denotes the number of people in the sample who are found to have blood group A.

(i) State the distribution of X, **justifying your answer**. Find the probability that there are two or fewer people with blood group A.

(ii) Three such samples of five people are taken. Find the probability that each of these three samples has more than two people with blood group A.

(iii) Twenty such samples of five people are taken. Find the probability that at least fifteen out of these twenty samples will contain at most two with blood group A.

Solution	Think Zone
(i) X is a binomial distribution since • there is a **fixed number** of number of people to be tested, $n=5$, • each person either has blood group A or does not, • the blood group of each person is **independent** of others, • the probability of a person having blood group A, $p=0.35$, can be considered to be a **constant** since the population is large.	You have to answer in the context of the question. Do not just copy the generic characteristics. The word "random sample" implies that the blood group of each person is independent of one another.
$X \sim B(5, 0.35)$ Using G.C., $P(X \le 2) \approx 0.76483$ $= 0.765$ (3 s.f.)	We usually leave answers to 5 s.f. for accuracy in case we need the value in computing subsequent parts of the questions. However, you still need to leave the **final** answer in **3 s.f.**
(ii) Let be the RV denoting the number of samples with more than 2 people with blood group A, out of 3 samples, $P(X>2) = 1-P(X\le 2) \approx 0.23517$ Thus ~~Y~~ $Y \sim B(3, 0.23517)$	We are no longer interested in the number of people in a sample of 5 with blood group A. Instead we are interested to know how many samples out of the 3 groups of 5 that satisfy the condition of "having more than 2 with blood group A". Hence we need to

	Using GC, $P(Y=3) \approx 0.013006$ $= 0.0130$ (3 s.f.) Alternatively, Required probability $= \left[P(X>2)\right]^3$ $= \left[1-P(X \le 2)\right]^3$ $\approx 0.23517^3$ ≈ 0.013006 $= 0.0130$	define a new random variable Y and identify its distribution.
(iii)	Let W be the RV denoting the number of samples with two or fewer people with blood group A out of 20 samples Thus $W \sim B(20, P(X \le 2))$ Using G.C, $P(W \ge 15) = 1 - P(W \le 14) = 0.676$ (3 s.f.)	Again, in this case, we are interested to know how many samples out of the 20 groups of 5 satisfy the condition of "at most 2 with blood group A". Thus there is a need to define a new random variable.

4.3 Expectation and Variance of a Binomial Distribution

For a binomial random variable, $X \sim B(n,p)$, its expectation, variance and standard deviation are as follows:

$E(X) = \mu = np$	**(In MF 26)**
$\text{Var}(X) = \sigma^2 = np(1-p)$,	**(In MF 26)**
standard deviation, $\sigma = \sqrt{\text{Var}(X)} = \sqrt{npq}$, where $q = 1-p$	

Note: The formulae are derived from calculations based on the definition of expectation and variance for discrete random variables in the previous chapter. *However, you are not required to know the derivation in the current syllabus.*

Example 10

A safety engineer claims that, on average, 60% of all drivers whose cars are equipped with seat belts use them on short trips. In a sample of 400 drivers whose cars are equipped with seat belts, find the mean number of drivers not wearing seat belts. Find also the variance of the number of drivers not wearing seat belts.

If a driver is caught without seat belt, he will be fined \$120. What is the expected fine collected by the traffic police, in a night where 400 drivers are sampled?

Solution	Think Zone
Let X be the random variable denoting the number of drivers who do not wear seat belt, out of 400 drivers. $X \sim \mathrm{B}(400, 0.4)$ $E(X) = np = 400 \times 0.4 = 160$ $var(X) = npq = 400 \times 0.4 \times 0.6 = 96$ The expected fine is 160×120 $= 19200$	Establish the distribution of X and use the formulae accordingly. Since the amount fined is a constant at \$120, the expected fine collected is proportional to the expected number of drivers not wearing seat belts.

Example 11 (Independent Reading)

In a binomial probability distribution, there are n trials and the probability of success for each trial is p. If the mean is 8 and the variance is 4.8, find the values of n and p.

Solution	Think Zone
Let X be the random variable representing the number of success out of n trials. Then, $X \sim \mathrm{B}(n, p)$ $\mathrm{E}(X) = np = 8$ ----------------------(1) $\mathrm{Var}(X) = np(1-p) = 4.8$ ---------------------(2) Subst. (1) into (2): $8(1-p) = 4.8$ $1 - p = 0.6$ $p = 0.4$ $\therefore n = \dfrac{8}{0.4} = 20$	Interpret the significance of having mean 8 and variance 4.8

Example 12 (Airline Overbooking)

Air Canada has apologised to a Canadian family and offered "very generous compensation" after the airline bumped a 10-year-old boy from a flight.

Agencies, S. A. (2017, April 17). Air Canada apologises for bumping boy, 10, from family holiday flight. https://www.theguardian.com/world/2017/apr/18/air-canada-apologises-bumping-boy-family-holiday-flight

United passenger dragged off plane likely to sue airline, attorney says

David Dao, a 69-year-old Vietnamese American doctor, was hospitalized after Chicago aviation police dragged him from the plane as the airline sought to **make space** on a flight from the city's O'Hare international airport to Louisville, Kentucky. Agencies, S. A. (2017, April 13). https://www.theguardian.com/world/2017/apr/13/united-airlines-passenger-lawsuit-david-dao

Past experiences show that about 20% of the passengers who are scheduled to take a particular flight on a private jet fail to show up. For this reason, the airline sometimes overbook flights, selling more tickets than the seats they have, with the expectation that they will have some no shows and reduce revenue loss due to empty seats. Suppose that the airline uses a small jet with 10 seats and consistently sells 12 tickets for every one of these flights.

(i) Find the probability that there are not enough seats for the passengers and state the assumption made in your calculation.

(ii) On average, how many passengers will be on each flight?

(iii) If each ticket is sold at \$1000 and the compensation associated to an overbooked seat is \$2500, which includes the refund of ticket sold, what is the expected revenue for each flight, leaving your answers to the nearest whole number?

(iv) Will the revenue be higher if the airline decides to sell exactly 10 tickets instead of 12?

Solution

(i) $X \sim B(12, 0.8)$

$P(\text{not enough seats})$

$= P(X=11) + P(X=12) = 0.27488 = 0.275$ (to 3 s.f.

Let X be the random variable denoting the number of passengers who shows up for the flight, out of 12 passengers.

Assume that whether a passenger shows up or do not show up is independent of other passengers

ThinkZone

Nowadays, people often travel in groups of two or more. Does this affect the independence assumption about passenger behaviour?

We are looking at the number of tickets sold.

The small jet has only 10 seats. If 11 or 12 passengers turn up, there will not be enough seats.

Solution

(ii) Average number of passengers on each flight

$E(X) = (12)(0.8) = 9.6$

(iii) We can find the expected compensation by letting Y be the compensation made and form the following table:

Y	0	2500	5000
$P(Y=y)$	$P(X \le 10)$	$P(X=11)$	$P(X=12)$

Expected revenue

= selling price of 12 tickets – E(Y)

$12(1000) - 0(P(X \le 10)) - \$2500(P(X=11)) - \$5000(P(X=12))$

$= \cancel{\$0}\ \$11141$

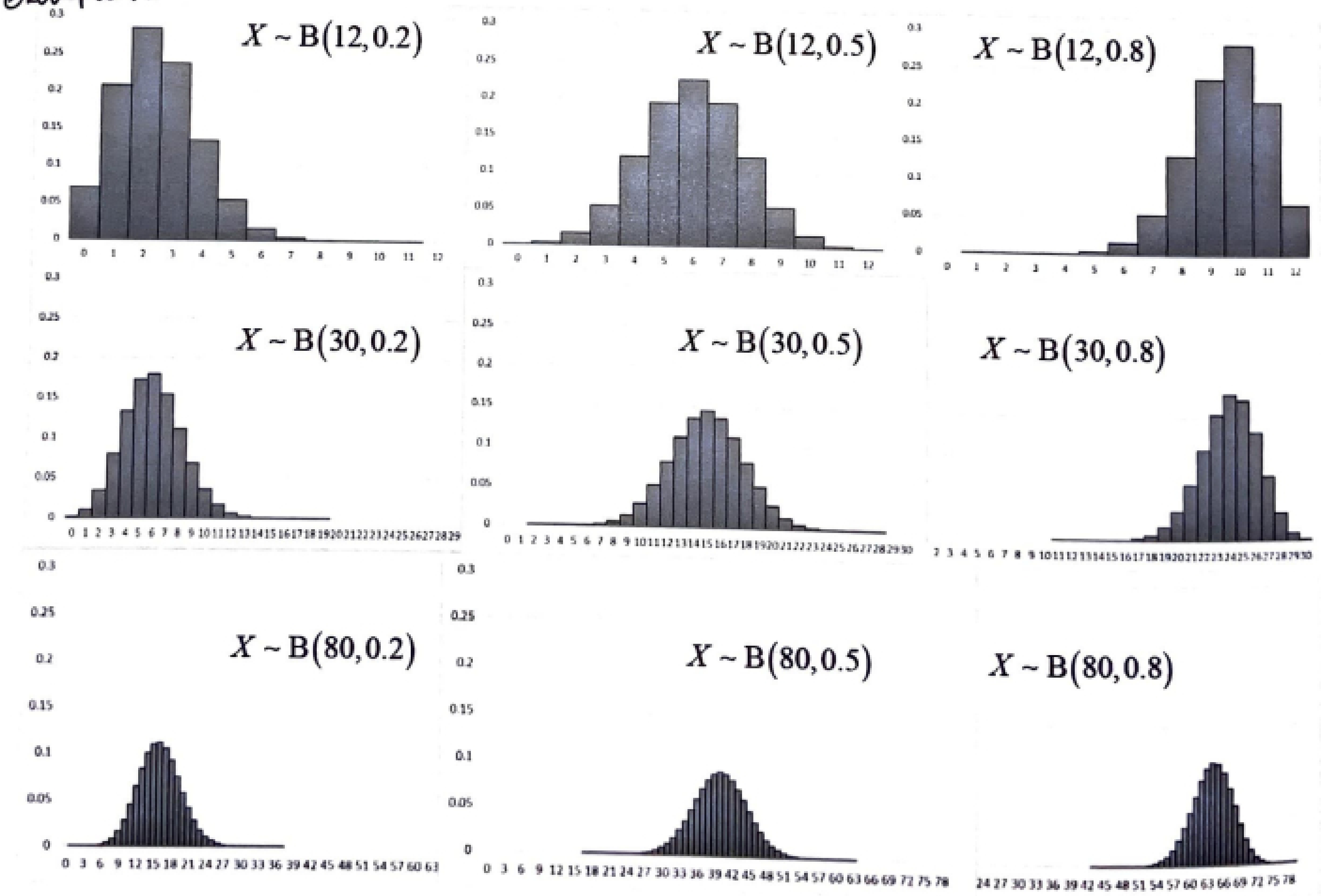
(iv) Revenue $=1000(10)=10000$ Since the expected revenue of selling all 12 tickets (with compensation for overbooking) is more than selling all 10 tickets, the airline should continue with the first method.	

The above example is a simplified version of calculations used by airlines when they overbook flights. As the airline issues more tickets, there is a higher chance of having to bump passengers from the flight, but there is also a higher chance of filling most seats. In reality, the airline has to make a trade-off between these two, by taking, its various costs and revenues and how sensitive the various no-show probabilities are to the number of tickets it issues, into account to determine the optimal number of tickets to issue.

<u>Effects of varying *n* and *p* on the shape of the binomial distribution</u>

Applet to explore how the binomial distribution change as n and p changes:
https://www.geogebra.org/m/GKb9XrA5

example12



What observations can you make and what conclusion can you come to?

1. For small p and small n, e.g. $X \sim \mathrm{B}(12, 0.2)$, the binomial distribution is skewed right (i.e. the bulk of the probability falls in the smaller numbers 0, 1, 2,..., and the distribution tails off to the right). In the context of the airline overbooking in Example ~~13~~ 12, it corresponds to the case where the no-show rate is ~~80%~~ 20% and there is allowance to more than 12 tickets such that the bulk of the probability still falls in the numbers smaller than 10 (number of seats).
2. For small n and large p, the binomial distribution is skewed left.
3. For $p = 0.5$ and large and small n, the binomial distribution is symmetric.
4. For large n, the binomial distribution approaches a bell shape.

4.4 Mode of X or most likely value of X

The **mode** is the value of X that is most likely to occur, i.e., **the most probable value** of X. To find the mode of a binomial distribution, we will calculate all the binomial probabilities and find the value of X with the highest probability.

Important: The **mode** (most probable number) is usually close to the **expected number** but both refer to different concepts and **must not** be used interchangeably.

The expected number refers to the average that is expected to be obtained if a sufficiently large number of the same experiment were repeated under the same set of conditions. The most probable number is the value most likely to occur in any given experiment.

Example 13

Of the inhabitants of a certain African village, 80% are known to have a particular eye disorder. If the eye specialist sees a total of 12 randomly selected inhabitants, what is the most likely number of them with the eye disorder?

Solution

Let X be the random variable denoting the number of inhabitants with eye disorder, out of 12 inhabitants.

$X \sim \mathrm{B}(12, 0.8)$

From GC,

x	$\mathrm{P}(X=x)$
9	0.23622
10	0.28347
11	0.20616

Need to compare 3 values of x to get mode

The most likely number (mode) of X is 10.

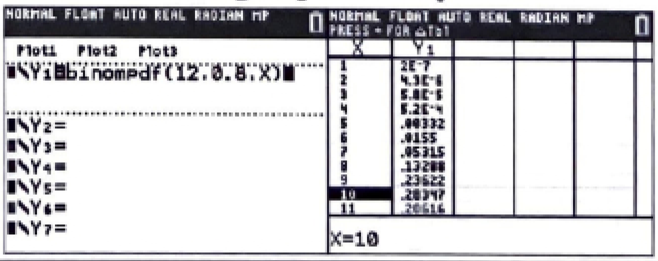
Think Zone

Press [Y=] followed by [2nd] [VARS] and scroll down to '**A:binompdf**' and press [ENTER].

Key in the values: '**trials:** $\boxed{12}$, **p:** $\boxed{0.8}$ and **x value:** $\boxed{X}$.

Highlight '**Paste**' and press [ENTER].

Press [2nd] [GRAPH] to get the table of values and scroll down to find the one with the **largest** Y_1 value, i.e., the **largest probability**.



	Note: The **unknown** that needs to be evaluated is always represented by **X** in the GC. In this example, the mode is 10 while the **expected number** is $np = 12(0.8) = 9.6$.

Example 14 (N82/1/13 (modified))

The random variable X is the number of success in n independent trials of an experiment in which the probability of a success at any one trial is p.

Show that $\dfrac{P(X=k+1)}{P(X=k)} = \dfrac{(n-k)p}{(k+1)(1-p)}$, $k = 0, 1, 2, \ldots, n-1$.

Hence deduce the most probable number of success when $n = 10$ and $p = \dfrac{1}{4}$.

Solution	Think Zone
$X \sim B(n, p)$ $P(X=k+1) = \binom{n}{k+1} p^{k+1}(1-p)^{n-(k+1)}$ $P(X=k) = \binom{n}{k} p^{k}(1-p)^{n-k}$ $\dfrac{P(X=k+1)}{P(X=k)} = \dfrac{\binom{n}{k+1} p^{k+1}(1-p)^{n-(k+1)}}{\binom{n}{k} p^{k}(1-p)^{n-k}}$ $= \dfrac{\dfrac{n!}{(k+1)!\,[n-(k+1)]!}\,p(1-p)^{-1}}{\dfrac{n!}{k!\,(n-k)!}}$ $= \dfrac{(n-k)(p)}{(k+1)(1-p)}$	$k! = k(k-1)(k-2)(k-3)\ldots.3\times2\times1$ $(k+1)! = (k+1)k(k-1)(k-2)(k-3)\ldots.3\times2\times1$ $(n-k)! = (n-k)[n-(k+1)][n-(k+2)]\ldots3\times2\times1$ $[n-(k+1)]! = [n-(k+1)][n-(k+2)]\ldots3\times2\times1$
When $n = 10$ and $p = \dfrac{1}{4}$ $\dfrac{P(X=k+1)}{P(X=k)} = \dfrac{(10-k)\times\frac{1}{4}}{(k+1)\left(1-\frac{1}{4}\right)} = \dfrac{(10-k)}{3(k+1)}$ Let k be the mode. Then $P(X=k) > P(X=k+1)$	

Solution	Think Zone
Hence, $\frac{P(X=k+1)}{P(X=k)}<1$ $\frac{10-k}{3(k+1)}<1$ $10-k<3(k+1)$ $k>\frac{7}{4}$ since $k\in\mathbb{Z}^+, k\geq 2$ When $k=2$, $P(X=2)>P(X=3)$ When $k=3$, $P(X=3)>P(X=4)$ . . . Hence, $P(X=2)>P(X=3)>P(X=4)>....$ Hence, most probable number is 2.	

4.5 Calculating unknown sample size (n), number of success (r) and probability of success (p)

The GC allows us to obtain the values of n, r and p via the command [**binompdf**] or [**binomcdf**]. We will key in the required command into $\boxed{\text{Y=}}$ and use the table of values (for integers) or the graph (for non-integer) to find these values. We will key in "X" in replacement of the unknowns (n, r or p).

Example 15 (Calculating the unknown sample size n)

A door-to-door canvasser tries to persuade people to have a certain type of double-glazing installed. The probability that his canvassing at a house is successful is 0.05. Find the least number of houses he must canvass in order that the probability of him getting at least one success exceeds 0.99.

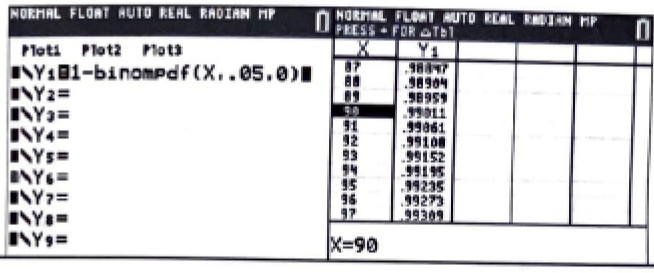
Solution	Think Zone
Let X be the random variable denoting the number of successful canvass, out of n houses. $X\sim B(n,0.05)$ $P(X\geq 1)>0.99$ $1-P(X=0)>0.99$ By G.C., we have n | $P(X\geq 1)$ 89 | 0.9896 90 | 0.9901	In this case, the GC does not allow us to compute the probability directly. Thus, we need to rewrite into $P(X=x)$ or $P(X\leq x)$. Which is the more appropriate one? Press [Y=] and key in '1−' followed by [2nd] [VARS] and scroll down to '**A:binompdf**' and press [ENTER].

Thus, the least number of houses he must canvass is 90

Key in the values: '**trials:** $\boxed{X}$ **(in replacement of n)**, **p:** $\boxed{0.05}$ and **x value:** $\boxed{0}$.

Highlight '**Paste**' and press ENTER.

Press 2nd GRAPH and scroll down to find the value of x corresponding to the first value of Y_1 which is **greater than 0.99**.

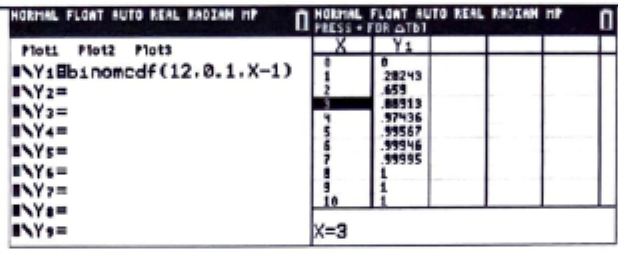


Example 16 (Calculating the unknown number of successes *r*)

The result of an experiment is a discrete random variable X which has a binomial probability distribution with mean 1.2 and variance 1.08. Find the least integer r such that $P(X<r)>0.85$.

Solution	**Think Zone**
$X \sim B(n,p)$ $E(X)=np=1.2$----------(1) $Var(X)=np(1-p)=1.08$----------(2) Solving, we have $n=12$ and $p=0.1$ Thus, $X \sim B(12,0.1)$ $P(X<r)>0.85$ $P(X\le r-1)>0.85$	In this case, the GC does not allow us to compute the probability directly. Thus, we need to rewrite the condition such that $P(X \le r-1)>0.85$. Press Y= followed by 2nd VARS and scroll down to '**B:binomcdf**' and press ENTER. Key in the values: '**trials:** $\boxed{12}$, **p:** $\boxed{0.1}$ and **x value:** $\boxed{X-1}$ **(in replacement of r–1)**
By G.C., we have Thus, the least integer r is 3	Highlight '**Paste**' and press ENTER. Press 2nd GRAPH to get the table of X and Y_1 values. Scroll down until you find the first value in Y_1 to be **greater than** 0.85.

n	$P(X\le r-1)$
2	0.65900
3	0.88913

Self Review 3

The result of an experiment is a discrete random variable X which has a binomial probability distribution with mean 3.2 and standard deviation 1.6. Calculate

$np = 3.2$ $\sqrt{np(1-p)} = 1.6$ $p = 0.2$ $n = 16$

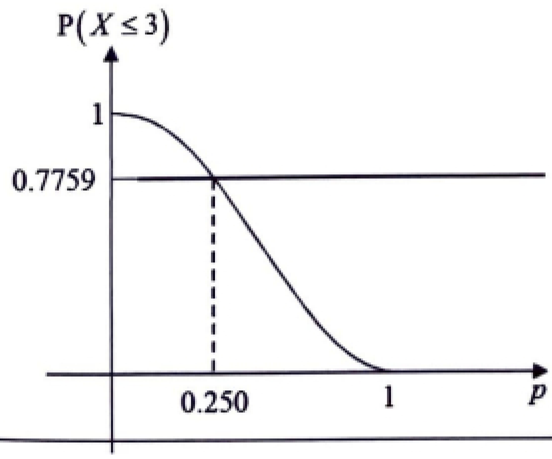
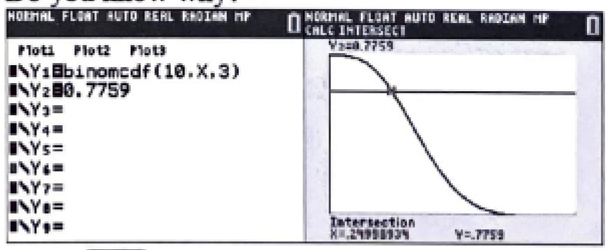
(i) $\mathrm{P}(X=1)$ to 3 significant figures;

(ii) the least integer r such that $\mathrm{P}(X \le r) > 0.85$,

(iii) the value of x such that $\mathrm{P}(X=x)$ is largest. [0.113; 5; 3]

Example 17 (Calculating the unknown probability of success, p)

For a certain strain of a flower, the probability that a seed produces a pink flower is a constant, p. In a sample of 10 such flowers, it is known that the probability of obtaining less than 4 pink flowers is 0.7759. Find the probability that a seed produces a pink flower.

Solution	**Think Zone**
Let X be the random variable denoting the number of seeds that produces pink flowers, out of 10 seeds. $X \sim B(10, p)$ $P(X < 4) = 0.7759$ $P(X \le 3) = 0.7759$ Thus by G.C, $p = 0.250$ (3sf) 	**DICE** Press Y= followed by 2nd VARS and scroll down to '**B:binomcdf**' and press ENTER. Key in the values: '**trials:** $\boxed{10}$, **p:** $\boxed{X}$ **(in replacement of p)** and **x value:** $\boxed{3}$. Highlight '**Paste**' and press ENTER. Key in 0.7759 in Y_2. Press WINDOW to get to 'Window Setup'. Use $X_{min} = 0$, $X_{max} = 1$, $Y_{min} = 0$ and $Y_{max} = 1$. Do you know why?  Press GRAPH to get the graph of Y_1 and Y_2.

Thus, by G.C.,	Press 2nd TRACE and scroll down to '**5:intersect**' to find the intersection between the two curves.

Question: Why can't we press 2nd GRAPH to obtain the value of p?

Self Review 4

In NYJC, $p\%$ of the student population is left-handed. Find p, if in a sample of 30 students, the probability of obtaining less than 9 left-handed students is 81.81%. [21.6]

$P(X<9)=81.81$ $X \sim B(30, p)$

$P(X \leq 8)=81.81$

Example 18 (J84/I/12(modified))

It may be assumed that dates of birth in a large population are evenly distributed throughout the year so that the probability of a randomly chosen person's date of birth in any particular month may be taken as $\frac{1}{12}$.

(i) Find the probability that of 10 people chosen at random, at least 2 will have their birthdays in May, June, July or August.

(ii) Find the probability that of 6 people chosen at random,
(a) exactly two will have their birthdays in January,
(b) more than 2 but less than 6 will have their birthdays in January.

(iii) 8 people are chosen at random. Find the probability that at least 1 people will have his birthday in January. Find also the probability that at least 2 people will have their birthdays in January given that at least 1 person will have a birthday in January.

(iv) M people are chosen at random. Find the least value of M so that the probability that at least two will have their birthdays in January exceeds 0.8.

Solution	
(i)	Let X be the random variable denoting the number of people who have their birthdays in May, June, July or August, out of 10 people. Then $X \sim \mathrm{B}\left(10, \frac{4}{12}\right)$, i.e., $X \sim \mathrm{B}\left(10, \frac{1}{3}\right)$ $\mathrm{P}(X \geq 2) = 1 - \mathrm{P}(X \leq 1) = 0.89595 \text{ (5 s.f.)}$ $= 0.896 \text{ (3 s.f.)}$
(ii)	Let Y be the random variable denoting the number of people who have their birthday in January, out of 6 people. Then $Y \sim \mathrm{B}\left(6, \frac{1}{12}\right)$ **(i)** $\mathrm{P}(Y = 2) = 0.073549 \text{ (5 s.f.)}$ $= 0.0735 \text{ (3 s.f.)}$ **(ii)** $\mathrm{P}(2 < Y < 6) = \mathrm{P}(Y \leq 5) - \mathrm{P}(Y \leq 2) = 0.0095449 \text{ (5 s.f.)}$ $= 0.00954 \text{ (3 s.f.)}]$ [Note: $\mathrm{P}(2 < Y < 6) = \mathrm{P}(Y = 3) + \mathrm{P}(Y = 4) + \mathrm{P}(Y = 5)$]

(iii)	Let W be the random variable denoting the number of people who have their birthday in January, out of 8 people. Then, $W \sim \mathrm{B}\left(8, \frac{1}{12}\right)$ $\mathrm{P}(W \geq 1) = 1 - \mathrm{P}(W = 0) = 0.50147$ (5 s.f.) $= 0.501$ (3.s.f.) $\mathrm{P}(W \geq 2 \mid W \geq 1) = \frac{\mathrm{P}(W \geq 2 \text{ and } W \geq 1)}{\mathrm{P}(W \geq 1)}$ $= \frac{\mathrm{P}(W \geq 2)}{\mathrm{P}(W \geq 1)} = \frac{1 - \mathrm{P}(W \leq 1)}{\mathrm{P}(W \geq 1)}$ $= \frac{0.13890}{0.50147}$ $= 0.277$
(iv)	Let R be the random variable denoting the number of people who have their birthday in January, out of M people. Then, $R \sim \mathrm{B}\left(M, \frac{1}{12}\right)$ Since $\mathrm{P}(R \geq 2) > 0.8$ $1 - P(R \leq 1) > 0.8$ Using G.C., <table><tr><td>M</td><td>$1 - P(R \leq 1)$</td></tr><tr><td>34</td><td>0.78767</td></tr><tr><td>35</td><td>0.8010</td></tr></table> Thus, the least value of M is

NANYANG JUNIOR COLLEGE

DEPARTMENT OF MATHEMATICS

H2 Mathematics **Year 2/2020**

Tutorial S4 – Binomial Distribution

1. In practising the high jump, Darren has 5 attempts at a particular height. The probability that he succeeds in an attempt is p.

(i) State two assumptions needed for the number of successful jumps to be modelled by a binomial distribution.

(ii) Find an expression, in terms of p, for the probability that he succeeds exactly 4 times.

(iii) The probability that he succeeds exactly 4 times is twice the probability that he succeeds exactly 2 times. Without the use of GC, find the value of p. [2/3]

2. A sample of 6 objects is to be drawn from a population in which 20% are defective. Explain why, if there are 10 000 objects in the whole population, the probability of obtaining 3 defective objects in the sample can be very accurately estimated by using the binomial probability formula, whereas, if there are only 20 objects in the population, this method gives quite the wrong answer.

For the large population, use this method to estimate the probability that a sample of 6 contains 3 or fewer defective objects. [0.983]

Another large population contains a proportion p of defective items. Write down an expression in terms of p for P, the probability that a sample of 6 contains exactly 2 defectives. Find $\frac{\mathrm{d}P}{\mathrm{d}p}$ and show that P is greatest when $p=\frac{1}{3}$.

3. Doreen has a 60% chance of scoring from the free-throw line when she plays basketball. If she makes 11 such throws in a typical game, find her most probable number of successful throws and state the probability of this occurrence.
Write down the expected number of successful throws. [7, 0.236, 6.6]

4. J76/1/3 (modified)

X is a binomial variable with mean 4 and variance $\frac{4}{3}$. Show that the largest value X can take is 6 and find $\mathrm{P}(X=5)$. This variable X represents the number of eggs laid each year by a certain species of bird. Find the largest value of r such that the probability of at least r eggs laid in a single year is more than 0.1. [0.263, 5]

5. Nation-wide surveys have shown that 1 in 20 junior college students is a student volunteer.

(i) Twelve junior college students are randomly selected to form a group. Find the probability that there are at least two student volunteers in the group. [0.118]

(ii) Three such groups are randomly chosen. Find the probability that one of these three groups has no student volunteer, another has exactly one student volunteer and the remaining group has at least two student volunteers. [0.131]

(iii) Ten such groups are randomly chosen. Find the probability that there are at least seven groups with less than two student volunteers in the group. [0.977]

6. A crossword puzzle is published in the Times each day of the week, excluding Sunday. A man is able to complete, on average, 8 out of 10 of the crossword puzzles.
   (i) State the expected value and standard deviation of the number of crossword puzzles the man would be able to complete in a given week. [4.8; 0.980]
   (ii) Find the probability that he will complete at least 5 puzzles in a given week. [0.655]
   (iii) Given that he completes the puzzle on Monday, find the probability that he will complete at least 4 puzzles in the rest of the week. [0.737]
   (iv) Given that he completes the puzzle on Monday, find the probability that he will complete at least 4 puzzles for the week. [0.942]
   (v) Find the probability that in a period of 4 weeks, he completes 4 or less puzzles in only one of the 4 weeks. [0.388]

7. In order to be offered a scholarship, a candidate has to pass two rounds of interview. In the first round, there will be a panel of 10 interviewers and the probability of each interviewer passing a candidate is 0.9. The candidate fails to qualify for the second round if more than one interviewer decides not to pass him or her. It is assumed that all interviewers' decisions are independent.
   (i) Find the probability that a candidate passes the first round of interview. [0.736]

   In the second round, there will be a panel of 5 interviewers and the probability of each interviewer passing a candidate is 0.8. The candidate is offered a scholarship only if all interviewers pass him or her in the second round.
   (ii) Show that the probability that the candidate is offered the scholarship is 0.241, correct to three significant figures.
   (iii) There are n candidates going for the interviews. Find the smallest n such that there is at least a 98% chance of 2 or more candidates being offered scholarships. [22]

8. **SAJC/2012/ P2/ Q13 (modified)**
   On the average, 30% of the students in a particular junior college could do the Binomial Distribution question in the JC 2 Mid-year Examination. The Level Head randomly select a class of 30 students to analyse the results.
   (i) State, in the context of this question, two assumptions needed to model the results by a binomial distribution.
   (ii) Find the probability that at least 6 students in that class could do that question. [0.923]
   (iii) Find the probability that only 2 students among the first 8 selected students in that class could do the question given that at least 6 students could do that question. [0.299]
   (iv) The probability of no more than 5 students could do the question in another randomly selected class exceeds 0.9. Find the largest possible number of students in this class. [11]
   (v) Another class of 30 students is randomly chosen. What should be the level of difficulty of the question (in terms of the percentage of students who could do the question on average) so that the probability that more than half of the class could do the question is 0.9? [63%]

9. **N2009/H1/10 (i),(ii) (DIY)**
   Over a long period of time, it is found that 20% of candidates who take a particular piano examination fail the examination.
   (i) Find the probability that, in a group of 10 randomly chosen candidates who take the examination, exactly 2 will fail. [2]
   (ii) It is given that 15% of the candidates who **pass** the piano examination are awarded a distinction. Find the probability that, in a randomly chosen group of 10 candidates who take the examination, fewer than 2 will be awarded a distinction. [3]

[0.302,0.658]

10. N81/2/12

In each batch of manufactured articles, the proportion of defective articles is p. From each batch, a random sample of nine is taken and each of the nine articles is examined. If two or more of the nine articles are found to be defective, that batch is rejected; otherwise it is accepted. Prove that the probability that a batch is accepted is $(1-p)^8(1+8p)$.

It is decided that to modify the sampling scheme so that, when one defective is found in the sample, a second sample of nine is taken and the batch rejected if this contains any defectives. With this exception, the original scheme is continued. Find an expression in terms of p for the probability that a batch is accepted. For this modified scheme, evaluate the average number sampled for manufactured batch over a large number of batches when p has the value 0.1. $[(1-p)^9\left[1+9p(1-p)^8\right]$, 12.5]

11. 2018 MYE/ACJC/8

The random variable X is the number of successes in n independent trials of an experiment in which the probability of a success at any trial is p.

Denoting $\mathrm{P}(X=k)$ by p_k, it is given that $\dfrac{p_k}{p_{k-1}}=\dfrac{(n-k+1)p}{k(1-p)}$, $k=1,\ 2,\ \dots,\ n$.

A distribution is said to be bimodal if it has two modes.

Find the least value of n, and the corresponding modes of X, given that X is bimodal and that $p=\dfrac{18}{25}$.

[17,18]

<u>Assignment</u>

1. **2017/RVHS/II/8 (modified)**

A car park next to a small commercial building has a total of 12 parking lots. Land surveillance officers have been observing the usage of parking lots per day to determine if the land has been efficiently utilised. Each parking lot can be occupied by at most one vehicle per day.

(i) Denoting the number of occupied parking lots per day by X, state in context, two assumptions needed for X to be well modelled by a binomial distribution.

(ii) Explain why one of the assumptions stated in **(i)** may not hold in this context.

It is further observed that for 80% of the days in the survey period, there are at least 4 occupied lots in the car park for each day. Assume that the assumptions in **(i)** hold,

(iii) Find the probability that a parking lot is being occupied in a day.

(iv) Using the value of the probability found in **(iii)**, find the expected number of parking lots occupied.

2. **2015/DHS/II/7 (modified)**

Data is transmitted in bytes, where each byte consists of 8 bits. The probability of a bit being corrupted during its transmission is 0.03. A byte is considered 'corrupted' if it contains at least 2 corrupted bits. Assume that all bits are not corrupted prior to their transmission.

(i) Find the probability that a randomly chosen byte is corrupted during its transmission.

(ii) Given that a randomly chosen byte is not corrupted during its transmission, find the probability that it contains no corrupted bits.

(iii) Find the probability that between 5 and 10 bytes are corrupted during the transmission of 100 bytes.

In another transmission system, the probability of having at least 3 corrupted bytes during the transmission of 20 bytes is 0.85. Find the probability that a bit is corrupted during its transmission.

Extra Practice Questions

(The questions will not be discussed during tutorials. Full solutions will be uploaded to NYJC portal)

1 2017/MJC/MYE/Q9

An airline company sold 105 tickets for a particular flight. Based on past experience, 8% of passengers do not turn up for the flight. Let X be the number of passengers who do not turn up for that flight.

(i) State, in context, two assumptions needed for X to be well modelled by a binomial distribution. [2]

(ii) Find the probability that fewer than 5 do not turn up for the flight. [2]

"Overbooking" occurs when there are more passengers who turn up than the number of seats available for the flight. When this happens, the airline will re-accommodate the excess passengers, that is, arrangements will be made for them to take another flight.

(iii) It is given that the particular flight has only 100 seats. Find the probability that at least 3 passengers have to be re-accommodated given that "overbooking" has occurred. [3]

2 2014/PJC/II/Q9

A female doctor wishes to test a new treatment for a certain genetic disease. On a particular day, she asks n (where n is fixed) patients who come to her clinic suffering from this disease if they are willing to take part in a trial. Based on past experiences, she finds that the probability of such a patient willing to take part is 0.3. The number of patients who are willing to take part is the random variable R.

(i) State, in the context of this question, two assumptions needed to model R by a binomial distribution. [2]

(ii) Explain why one of the assumptions stated in part **(i)** may not hold in this context. [1]

Assume now that these assumptions do in fact hold.

(iii) Given that $n = 8$, find the probability that R is at least 2. [1]

(iv) If the chance of at least 2 patients willing to take part in this trial is more than 90%, find the least possible integer value of n. [3]

3 2016/ACJC/II/11 (modified)

The manager of a car show room wants to study the number of cars sold by a car salesman under his charge. The number of potential car-buyers that he meet in a particular week is 60 and the average probability that he is successful in closing a deal with each customer is 0.2.

(i) It is assumed that the deals closed are independent of one another. State, in context, another assumption needed for the number of deals closed by a car salesman to be well modelled by a binomial distribution. [1]

(ii) Explain why the assumption that the deals closed are independent of one another may not hold in this context. [1]

(iii) Assuming that the assumptions stated in part **(i)** hold, find the probability that the salesman closed a total of more than 20 deals in a particular week. [2]

A new salesman joined the company. During his probation week, he met 60 potential car-buyers. The number of car deals he closed during his probation week is denoted by C with the distribution $\mathrm{B}(60, p)$.

(iv) Given that $\mathrm{P}(C = 30) = 0.03014$. Find an equation for p. Hence find the value of p, correct to 1 decimal place, given that $p < 0.5$. [2]

4 2016/TJC/II/10 (part)

Newmob is a mobile phone service provider which sells several brands of mobile phones. uPhones and Samseng phones are sold at a subsidy to its subscribers. Each subscriber can either buy one uPhone or one Samseng phone or both one uPhone and one Samseng phone. The probability that a randomly chosen subscriber buys a uPhone is 0.3, and independently, the probability that the subscriber buys a Samseng phone is p.

(i) In a random sample of 50 subscribers, the probability that at most 20 subscribers buy a Samseng phone is twice the probability that exactly 15 subscribers buy a uPhone. Find the value of p. [3]

For the remainder of this question, you may take the value of p to be 0.4.

(ii) In a random sample of 50 subscribers, find the probability that the number of subscribers who buy a uPhone is greater than the expected number of subscribers who buy a Samseng phone. [2]

5 2016/MJC/II/6

A box consists of a very large number of balls, of which 20% are red and 80% are white.

A game consists of a player drawing n balls at random from the box and counting the number of red balls drawn. If at most one red ball is drawn, the player wins. If more than two red balls are drawn, the player loses. If exactly two red balls are drawn, the player draws another n balls and if none of these n balls drawn are red, the player wins. Otherwise, the player loses.

Show that the probability that a randomly chosen player wins is P where

$$P=(0.8+0.2n)(0.8)^{n-1}+\binom{n}{2}(0.2)^2(0.8)^{2n-2}.$$ [3]

(i) Given that the probability that a randomly chosen player wins is less than 0.1, write down an inequality in terms of n to represent this information. Hence find the least possible value of n. [3]

(ii) Given instead that $P=0.3$, find the probability that out of 100 games played, at least 40 games are won. [2]

6 2016/HCI/MYE/6 (modified)

A factory produces large numbers of bulbs that are packed in boxes of 10. On average, 5% of the bulbs are spoiled. Assume that the quality of all bulbs are independent of one another.

(i) Show that the probability that one box contains at least two spoiled bulbs is 0.0861. [1]

A box is rejected if there are at least two spoiled bulbs in it.

(ii) A quality test is conducted on 10 randomly chosen boxes. Find the probability that the 10th box is the third rejected box. [2]

(iii) Find the probability that, out of 50 randomly chosen boxes, at most 5 boxes are rejected. [2]

(iv) The probability that, among n randomly chosen boxes, fewer than 2 boxes are rejected is more than 0.7. Find the largest possible value of n. [3]

7 2016/NYJC/MYE/II/7

In a large population the proportion of people with disability is 15%. A random sample of 20 people is taken.

(i) State, in context, an assumption needed for the number of people in the sample with disability to be well modelled by a binomial distribution. [1]

(ii) Four such samples of 20 people are taken. Find the probability that two of these samples have exactly one person with disability, and the other two samples have more than two people with disability. [3]

(iii) Find the least number of such samples of 20 people that must be taken so that the probability of having at least two samples with more than two people with disability is not less than 0.95. [3]

8 2017/CJC/MYE/1/9

The chocolate brand 'Kid Cats' are sold in packets of 20 individual pieces. Each packet is made up of randomly chosen flavours of Kid Cats pieces produced. On average 15% of the Kid Cats pieces are hazelnut flavoured.

(i) Let the random variable X denote the number of hazelnut flavoured Kid Cats pieces in a randomly chosen packet. State, in context, one assumption needed for X to be well modelled by a binomial distribution. [1]

(ii) Find the probability that a randomly chosen packet of Kid Cats contains no more than 4 hazelnut flavoured pieces. [2]

(iii) The Interact Club buys 100 randomly chosen packets of Kid Cats for their event. Find the probability that at least 75 of these packets contain no more than 4 hazelnut flavoured pieces. [3]

On average the proportion of Kid Cats pieces that are almond flavoured is q. It is known that the most common number of almond flavoured pieces in a packet is 6.

(iv) Use this information to find exactly the range of values that q can take. [4]

9 2006/VJC/II/29(Either) (modified)

11 cards marked with the letters from the word MATHEMATICS are placed in a bag. Mary draws 3 cards at random and without replacement. She scores 1 point for every vowel drawn and 2 points for every consonant drawn.

Let X denote Mary's score. Show that

$$\mathrm{P}(X \geq 4) = \frac{161}{165}$$

and find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$. [7]

The cards are now replaced in the bag.

John draws 3 cards at random and without replacement. Like Mary, John scores 1 point for every vowel drawn and 2 points for every consonant drawn. After noting his score, he replaced the 3 cards in the bag. John carries out the same experiment a total of 165 times.

Find the probability that he gets a score of at least 4 for at least 161 times. [2]

10 2005/RJC/II/26

Alex undergoes 4 training sessions each week, on separate days. In each of his training session, Alex has to run 10 laps of 400m. The probability of him meeting his target time for each lap on a sunny day is 0.95. On a rainy day, the probability of achieving his target time is 0.9. Each day is either sunny or rainy, and the probability of a sunny day is 0.8.

(i) Calculate the probability of Alex achieving his target time for at least 8 of the laps on a rainy day. [2]

(ii) Given that there was a week of rainy days, determine the conditional probability of Alex being able to achieve his target time for 8 or more laps, in each of the sessions, on 3 separate days. [4]

There was sunny weather for 2 consecutive weeks. Find the probability of Alex not meeting his target time for more than 4 laps during the 2 weeks. [2]

Answer:

1	(ii) 0.0706 (iii) 0.115
2	(i) (a) The probability of willing to take part is 0.3 remains constant for all patients (b) All patients willing to take parts are independent (ii) The patients may not be independent as the disease is genetic, may affect same people from same family (iii) 0.745 (iv) $n = 12$
3	(iii) 0.00483 (iv) 0.4
4	(i) 0.459 (ii) 0.0478
5	(i) 19 (ii) 0.0210
6	(i) 0.0861 (ii) 0.0122 (iii) 0.741 (iv) 12
7	(ii) 0.0398 (iii) 6
8	(ii) 0.830 (iii)0.985 (iv) $\frac{2}{7} < q < \frac{1}{3}$
9	4.91, 0.555, 0.629
10	(i) 0.930 (ii) 0.226 (iii) 0.371