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, $\mathrm{B}(n, p)$, 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

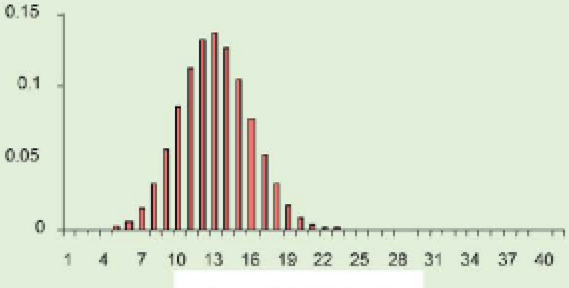
Consider an experiment with a fixed number, n, of independent repeated trials, with two possible outcomes, of which one is defined as '**success**' and the other '**failure**'. Suppose the probability of success is a constant, p, and the probability of failure is $1-p$ (or q).

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 $X \sim \mathrm{B}(n, p)$ 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

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

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 $X \sim \mathrm{B}(40, 0.3)$. We can also use a diagram to illustrate the respective probabilities of the distribution.



$X \sim \mathrm{B}(40, 0.3)$

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 **fixed number**, n, of repeated trials, where $n \in \mathbb{Z}^+$.
2. There are **2 mutually exclusive outcomes**, namely success and failure.
3. The **outcome** of each trial **is independent** of the outcomes of other trials.
4. The **probability of success**, p, **is constant** 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.

Solution	**Think Zone**
• The colour of the n cars must be independent of the colour of any other cars. • The probability that a car is red is the same throughout the sample.	Recall the conditions for binomial distribution and infer which is relevant in the context of the question.

When solving questions involving distributions, it is useful to remember the **DICE** framework:

D: Define the random variable
I: Identify the distribution of the random variable
C: Compute the probability
E: Evaluate if the answer makes sense ($0 \le p \le 1$)

Example 4 (2007/II/9):

A random variable X has a binomial distribution with $n=6$ and the probability of success p.

Write down an expression, in terms of p, for $\mathrm{P}(X=4)$.

It is given that $p=\frac{1}{4}$. Find $\mathrm{P}(X=4)$, giving your answer as a fraction.

Solution	**Think Zone**
$X \sim \mathrm{B}(6,p)$ $\mathrm{P}(X=4)=\binom{6}{4}p^4(1-p)^{6-4}=15p^4(1-p)^2$ When $p=\frac{1}{4}$, we have $X \sim \mathrm{B}\left(6,\frac{1}{4}\right)$ $\mathrm{P}(X=4)=15\left(\frac{1}{4}\right)^4\left(1-\frac{1}{4}\right)^2=\frac{135}{4096}$	Apply the **DICE** framework: **D**: Define the random variable **I**: Identify the distribution of the random variable **C**: Compute the probability **E**: Evaluate if the answer makes sense $(0 \le p \le 1)$

Example 5:

Given that the probability of a randomly chosen person having blood type A is 0.4, find the probability that in a group of four randomly chosen people,

(i) no one has blood type A,

(ii) more than two have blood type A.

Solution	**Think Zone**
Step 1: Define the random variable Let X be the random variable denoting the number of people who have blood type A, out of four people. **Step 2: Identify the distribution of X** $X \sim \mathrm{B}(4,0.4)$ **Step 3: Compute & Step 4: Evaluate** **(i)** $\mathrm{P}(X=0)=\binom{4}{0}(0.4)^0(0.6)^4=0.130\,(3\text{ s.f.})$ **(ii)** $\mathrm{P}(X>2)=\mathrm{P}(X=3\text{ or }4)$ $=\mathrm{P}(X=3)+\mathrm{P}(X=4)$ $=\binom{4}{3}(0.4)^3(0.6)+\binom{4}{4}(0.4)^4(0.6)^0$ $=0.1536+0.0256$ $=0.180\ (3\text{ s.f.})$	Apply the **DICE** framework: What are we interested to find in this experiment? To do this, check: 1. Is there a maximum number that X can take? 2. Is the probability of a person having blood type A constant throughout? Once we have established that X follows a binomial distribution, we can apply the formula to calculate the probabilities. In order to apply the formula, we have to convert the '>' to '=' by interpreting the meaning of the question.

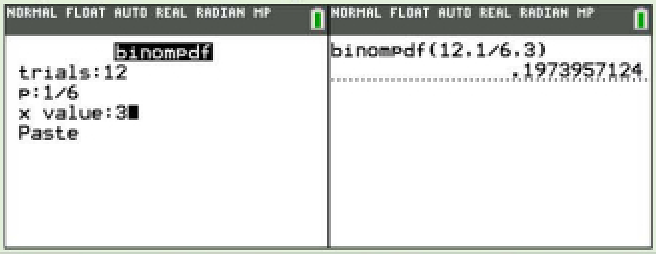
4.2 Use of G.C. for questions with Binomial Random Variable

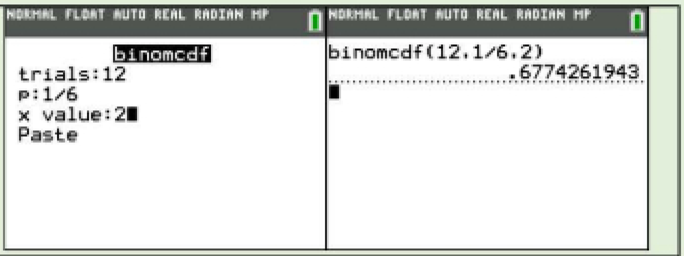
The G.C. allows us to compute the value of $\mathrm{P}(X = x)$ and $\mathrm{P}(X \le x)$directly via the command [**binompdf**] and [**binomcdf**] respectively. We would need to rewrite the expressions if we require other probabilities.

Example 6

An ordinary die is thrown 12 times. Find the probability of obtaining

(i) exactly three 6's,
(ii) less than three 6's,
(iii) more than five 6's,
(iv) between two and seven 6's,
(v) between two and seven 6's (inclusive of both).

Solution	**Think Zone**
Let X be the random variable denoting the number of 6's obtained out of 12 throws. Then, $X \sim \mathrm{B}\left(12, \frac{1}{6}\right)$	Define the random variable first and establish its distribution by asking yourself: - What are we interested to find in this experiment? - Does it fulfil the criteria of the binomial random variable?
(i) Using GC, $\mathrm{P}(X = 3) = 0.197$ (3 s.f.)	Press 2nd VARS and scroll down to '**A:binompdf**' and press ENTER. Key in the values: '**trials:** $\boxed{12}$, **p:** $\boxed{1/6}$ and **x value:** $\boxed{3}$. Highlight '**Paste**' and press ENTER.  NORMAL FLOAT AUTO REAL RADIAN MP binompdf trials:12 p:1/6 x value:3 Paste NORMAL FLOAT AUTO REAL RADIAN MP binompdf(12,1/6,3) .1973957124
(ii) Using GC, $\mathrm{P}(X < 3) = \mathrm{P}(X \le 2)$ $= 0.677$ (3 s.f.)	In this case, the GC does not allow us to compute the probability directly. Thus, we need to rewrite into $P(X = x)$ or $P(X \le x)$. Which is the more appropriate one? Press 2nd VARS and scroll down to '**B:binomcdf**' and press ENTER. Key in the values: '**trials:** $\boxed{12}$, **p:** $\boxed{1/6}$ and **x value:** $\boxed{2}$. Highlight '**Paste**' and press ENTER.

	 binomcdf trials:12 p:1/6 x value:2 Paste binomcdf(12,1/6,2) .6774261943
(iii) Using GC, $\mathrm{P}(X>5)=1-\mathrm{P}(X\le 5)$ $=0.00792$ (3 s.f.)	In this case, the GC does not allow us to compute the probability directly. Thus, we need to rewrite into $P(X=x)$ or $P(X\le x)$. Which is the appropriate one?
(iv) Using GC, $\mathrm{P}(2<X<7)=\mathrm{P}(X\le 6)-\mathrm{P}(X\le 2)$ $=0.321$ (3 s.f.)	Again, the GC does not allow us to compute the probability directly. Be careful when answering this question so as not to under or over count probabilities **Note:** $\mathrm{P}(2 < X < 7) = \mathrm{P}(X = 3) + \ldots + \mathrm{P}(X = 6)$
(v) Using GC, $\mathrm{P}(2\le X\le 7)=P(X\le 7)-P(X\le 1)$ $=0.619$ (3 s.f.)	**Note:** $\mathrm{P}(2\le X\le 7) = \mathrm{P}(X = 2) + \ldots + \mathrm{P}(X = 7)$

Example 7 (Independent Reading):

A multiple-choice test has 10 questions, each with five different possible responses, of which only one is correct. If a student is totally unskilled and selects at random a response to each question, so that, for each question, the probability of choosing the correct answer is $\frac{1}{5}$, find the probability that he gets

(i) all the questions wrong,

(ii) less than 50% of the questions correct,

(iii) **at most** three questions correct,

(iv) **not more than** 2 correct and

(v) **at least** three questions correct.

Solution	**Think Zone**
Let X be the random variable denoting the number of questions that are correct, out of 10 questions. $X\sim \mathrm{B}\left(10,\frac{1}{5}\right)$	Read the entire question before deciding the "successful" outcome, ie "number of questions that are correct". Have you used the **DICE** framework?
(i) $\mathrm{P}(X=0)=0.107$ (3 s.f.)	Interpret – Having all the questions wrong is equivalent to having 0 correct questions.
(ii) $\mathrm{P}(X<5)=\mathrm{P}(X\le 4)=0.967$ (3 s.f.)	Interpret – 50% of the questions is equivalent to 5 questions out of a total of 10.
(iii) $\mathrm{P}(X\le 3)=0.879$ (3 s.f.)	Interpret – **at most** 3 is equivalent to having a **maximum** of 3.

(iv) $\mathrm{P}(X \le 2) = 0.678$ (3 s.f.)	Interpret – **not more than** 2 is equivalent to having a **maximum** of 2.
(v) $\mathrm{P}(X \ge 3) = 1 - \mathrm{P}(X \le 2) = 0.322$ (3 s.f.)	Interpret – **at least** 3 is equivalent to having a **minimum** of 3.

Self Review 2:

The probability that the people in town X support Party A is 0.6. Find the probability that in a randomly selected sample of 8 voters, there are

(i) exactly 3 who support Party A,

(ii) more than 5 who support Party A. [0.124; 0.315]

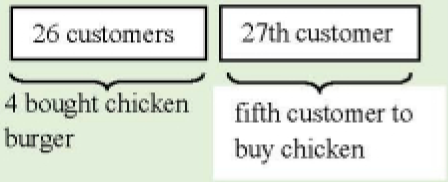
Solution	**Think Zone**
Let X be the random variable denoting the number of voters who support Party A, out of 8 voters. $X \sim \mathrm{B}(8, 0.6)$	
(i) $\mathrm{P}(X = 3) = 0.124$ (3 s.f.)	
(ii) $\mathrm{P}(X > 5) = 1 - \mathrm{P}(X \le 5) = 0.315$ (3 s.f.)	

Example 8 (Question involving conditional probability)

From past records, the purchases at a fast food outlet show that 9% of its customers prefer to buy chicken burgers. Find the probability that, in a random sample of 27 customers,

(a) exactly 4 customers buy chicken burgers given that fewer than 6 customers buy chicken burgers.

(b) the last customer is the fifth customer to buy chicken burger.

Solution	**Think Zone**
(a) Let X be the random variable denoting the number of customers who buy chicken burgers, out of 27 customers. $X \sim \mathrm{B}(27, 0.09)$ $\mathrm{P}(X = 4 \mid X < 6) = \dfrac{\mathrm{P}(X = 4 \text{ and } X < 6)}{\mathrm{P}(X < 6)}$ $= \dfrac{\mathrm{P}(X = 4)}{\mathrm{P}(X \le 5)}$ $= 0.136$ (3 s.f.)	Have you used the **DICE** framework?
(b) Let Y be the random variable denoting the number of customers who buy chicken burgers, out of 26 customers. $Y \sim \mathrm{B}(26, 0.09)$	

$\text{P}(\text{last customer is 5th customer to buy chicken burger})$ $= \text{P}($ 4 customers out of 26 customers buy chicken burger and last customer buys chicken burger) $= \text{P}(Y=4)(0.09)$ $= 0.0111$ (3 s.f.)	

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. $X \sim \text{B}(5, 0.35)$ Using G.C., $\text{P}(X \le 2) \approx 0.76483$ $= 0.765$ (3 s.f.)	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. 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 Y be the random variable denoting the number of samples with more than 2 people with blood group A, out of 3 samples $\text{P}(X>2) = 1 - \text{P}(X \le 2) \approx 0.23517$ Thus, $Y \sim \text{B}(3, 0.23517)$ Using GC, $\text{P}(Y=3) \approx 0.013006$ $= 0.0130$ (3 s.f.)	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 define a new random variable Y and identify its distribution.

Alternatively, Required probability $= \left[\mathrm{P}(X > 2)\right]^3$ $= \left[1 - \mathrm{P}(X \le 2)\right]^3$ $\approx 0.23517^3$ ≈ 0.013006 $= 0.0130$	
(iii) Let W be the random variable denoting the number of samples with two or fewer people with blood group A, out of 20 samples. Thus $W \sim \mathrm{B}(20, \mathrm{P}(X \le 2))$ Using G.C., $\mathrm{P}(W \ge 15) = 1 - \mathrm{P}(W \le 14) = 0.676$ (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, $X \sim \mathrm{B}(n, p)$, its expectation, variance and standard deviation are as follows:

$$\mathrm{E}(X) = \mu = np \qquad \textbf{(In MF 26)}$$

$$\mathrm{Var}(X) = \sigma^2 = np(1-p) \qquad \textbf{(In MF 26)}$$

$$\text{standard deviation, } \sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{npq}\text{, where } q = 1-p$$

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. $X \sim \mathrm{B}(400, 0.4)$	Establish the distribution of X and use the formulae accordingly.
$\mathrm{E}(X) = np = (400)(0.4) = 160$ $\mathrm{Var}(X) = npq = (400)(0.4)(0.6) = 96$	Since the amount fined is a constant at $120, the expected fine collected is proportional to the expected

Thus, the expected fine collected is $160\times\$120=\$19200$.	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): $8(1-p)=4.8$ $1-p=0.6$ $p=0.4$ $\qquad\therefore n=\dfrac{8}{0.4}=20$	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?

<table>
<tr><th>Solution</th><th>ThinkZone</th></tr>
<tr><td>(i) Assume that whether a passenger show up or do not show up is independent of other passengers.

Let X be the random variable denoting the number of passengers who shows up for the flight, out of 12 passengers.
$X \sim \mathrm{B}(12, 0.8)$.
$\mathrm{P}(\text{not enough seats}) = \mathrm{P}(X=11)+\mathrm{P}(X=12)$
$= 0.27488$
$= 0.275 \text{ (3 s.f.)}$</td><td>Nowadays, people often travel in groups of two or more. Does this affect the independence assumption about passenger behaviour?

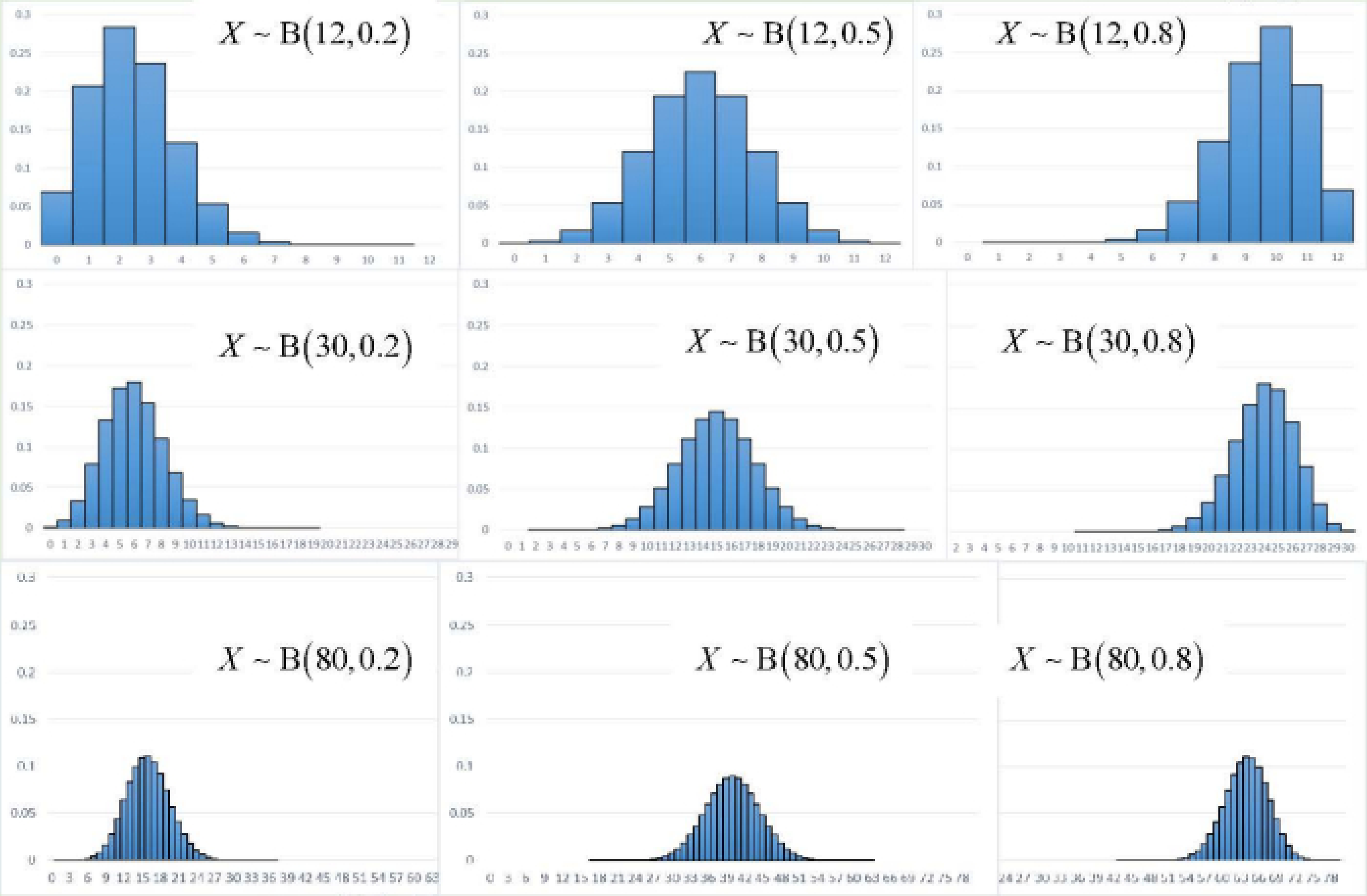
We are looking at the number of tickets sold.

The small jet has only 10 seats. If 11 or 12 passengers turn up, there will not be enough seats.</td></tr>
<tr><td>(ii) Average number of passengers on each flight
$= \mathrm{E}(X) = (12)(0.8) = 9.6$</td><td></td></tr>
<tr><td>(iii) We can find the expected compensation by letting Y be the compensation made and form the following table:
<table>
<tr><td>Y</td><td>0</td><td>2500</td><td>5000</td></tr>
<tr><td>$\mathrm{P}(Y=y)$</td><td>$\mathrm{P}(X \le 10)$</td><td>$\mathrm{P}(X=11)$</td><td>$\mathrm{P}(X=12)$</td></tr>
</table>
Expected revenue
= selling price of 12 tickets − $\mathrm{E}(Y)$
$= 1000 \cdot 12 - \left[2500 \cdot \mathrm{P}(X=11) + 5000 \cdot \mathrm{P}(X=12)\right]$
$= 12000 - \left[2500(0.20616) + 5000(0.06872)\right]$
$= \$11141$</td><td></td></tr>
<tr><td>(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.</td><td></td></tr>
</table>

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.

Effects of varying *n* and *p* on the shape of the binomial distribution

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

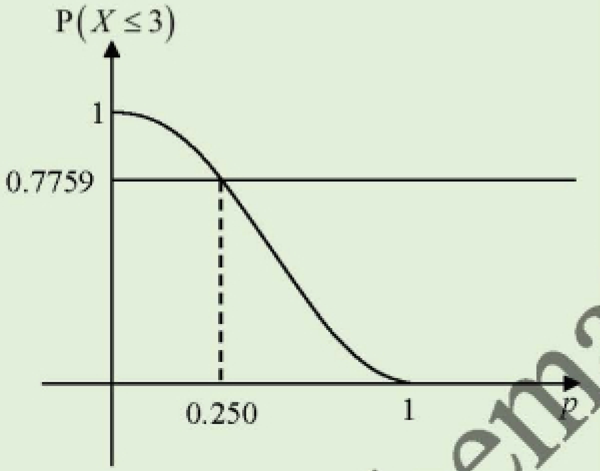
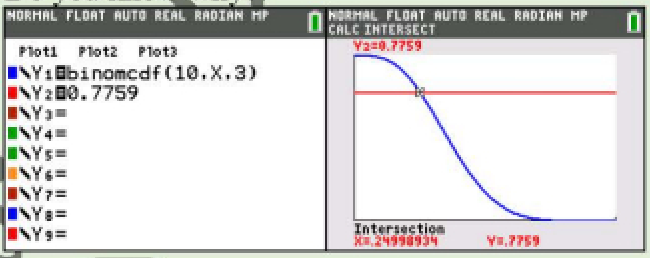


What observations can you make and what conclusion can you come to?

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 12, it corresponds to the case where the no-show rate is 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.

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. $X \sim \mathrm{B}(10, p)$ $\mathrm{P}(X < 4) = 0.7759$ $\mathrm{P}(X \leq 3) = 0.7759$  Thus, by G.C., $p = 0.250 (3 \text{ s.f.})$	**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 $\mathrm{X_{min}} = 0$, $\mathrm{X_{max}} = 1$, $\mathrm{Y_{min}} = 0$ and $\mathrm{Y_{max}} = 1$. Do you know why?  NORMAL FLOAT AUTO REAL RADIAN MP Plot1 Plot2 Plot3 \Y1=binomcdf(10,X,3) \Y2=0.7759 \Y3= \Y4= \Y5= \Y6= \Y7= \Y8= \Y9= NORMAL FLOAT AUTO REAL RADIAN MP CALC INTERSECT Y2=0.7759 Intersection X=.24998934 Y=.7759 Press GRAPH to get the graph of Y_1 and Y_2. 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]

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, $W \sim \mathrm{B}\left(8, \frac{1}{12}\right)$

$$\mathrm{P}(W \geq 1) = 1 - \mathrm{P}(W = 0) = 0.50147 \text{ (5 s.f.)}$$
$$= 0.501 \text{ (3 s.f.)}$$

$$\mathrm{P}(W \geq 2 | W \geq 1) = \frac{\mathrm{P}(W \geq 2 \text{ and } W \geq 1)}{\mathrm{P}(W \geq 1)}$$
$$= \frac{\mathrm{P}(W \geq 2)}{\mathrm{P}(W \geq 1)} = \frac{1 - \mathrm{P}(W \leq 1)}{\mathrm{P}(W \geq 1)}$$
$$= \frac{0.13890}{0.50147}$$
$$= 0.277$$

(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 $\mathrm{P}(R \geq 2) > 0.8$

$1 - \mathrm{P}(R \leq 1) > 0.8$

Using G.C.,

M	$1 - \mathrm{P}(R \leq 1)$
34	0.78767
35	0.80104

Thus, the least value of M is 35.