Supporting Australian Mathematics Project

A guide for teachers – Years 11 and 12

Probability and statistics: Module 20

Binomial distribution

Education Services Australia

AMSI
AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Binomial distribution

Assumed knowledge

The content of the modules:

- *Probability*
- *Discrete probability distributions.*

Motivation

This module focusses on the **binomial distribution**. The module *Discrete probability distributions* includes many examples of discrete random variables. But the binomial distribution is such an important example of a discrete distribution that it gets a module in its own right.

The importance of the binomial distribution is that it has very wide application. This is because at its heart is a binary situation: one with two possible outcomes. Many random phenomena worth studying have two outcomes. Most notably, this occurs when we examine a sample from a large population of 'units' for the presence of a characteristic; each unit either has the characteristic or it doesn't. The generic term 'unit' is used precisely because the situation is so general. The population is often people, in which case a unit is a person; but a unit might be a school, an insect, a bank loan, a company, a DNA sequence, or any of a number of other possibilities.

This module starts by introducing a Bernoulli random variable as a model for a binary situation. Then we introduce a binomial random variable as the number of 'successes' in n independent Bernoulli trials, each with the same probability of success p. We show how to calculate probabilities associated with a binomial distribution, and illustrate the use of binomial distributions to solve practical problems. The last section covers the mean and variance of a binomial distribution.

Content

Bernoulli trials

The module *Discrete probability distributions* discusses the idea of a sequence of independent trials, where each trial has the same probability of success p. This structure leads to a number of random variables with different distributions. In that module, the trials structure is used to introduce the *geometric distribution*. Because of the general importance of the trials structure, we examine it systematically in this module.

The central idea is a Bernoulli trial — named after Jacob Bernoulli (1655–1705), who was one of a family of prominent mathematicians. A **Bernoulli trial** is a random procedure that can have one of two outcomes, which are arbitrarily labelled 'success' and 'failure'.

Example: Tetris

The following random procedure is considered in the module *Probability*.

During a game of Tetris, we observe a sequence of three consecutive pieces. Each Tetris piece has one of seven possible shapes: `I,J,L,O,S,T,Z`. So in this random procedure, we can observe a sequence such as `JLL`, `ZOS`, `ZSZ`, `III` and so on.

For each sequence of three pieces, we may ask: Does it contain at least one `Z`? The sequence of three pieces is the test or 'trial'. A sequence of three pieces containing a `Z` is regarded as a success, and one without a `Z` as a failure.

From the point of view of this question, it does not matter what the other shapes in a sequence are. All we are concerned about is whether or not the sequence has a `Z`.

A Bernoulli trial has a corresponding **Bernoulli random variable**, which counts the number of successes in a single trial. The only possible values of this random variable are zero and one; the random variable takes the value one if a success occurs, and the value zero if a failure occurs.

Let X be a Bernoulli random variable with parameter p, where $0 < p < 1$. The probability function $p_X(x)$ of X is given by

$$p_X(x) = \Pr(X = x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases}$$

The mean of this random variable is

$$\mu_X = \mathrm{E}(X) = \sum x\,p_X(x)$$

$$= \big(0 \times (1-p)\big) + \big(1 \times p\big)$$

$$= p.$$

The variance of X is equal to $p(1-p)$, a result obtained as follows:

$$\mathrm{var}(X) = \sum (x - \mu_X)^2 p_X(x)$$

$$= (0-p)^2(1-p) + (1-p)^2 p$$

$$= p(1-p)\big(p + (1-p)\big)$$

$$= p(1-p).$$

Bernoulli random variables arise in a way that occurs very widely indeed. Suppose we are interested in a population of 'units', in which the proportion of units with a particular characteristic is p. Here a 'unit' may be a person, an animal, a plant, a school, a business, or many other entities, according to the population under study. We take a random sample of units from the population, and observe whether or not each unit has this characteristic of interest.

If the population is infinite, or so large that we can regard it as effectively infinite, then sampling without replacement is the same as sampling with replacement, and if each unit is sampled independently from all the others, the probability of any single sampled unit having the characteristic is equal to p.

If we define 'success' as 'having the characteristic of interest', then each observation can be regarded as a Bernoulli trial, independent of the other observations, with probability of success equal to p.

The importance of this insight is that it is so widely applicable. Here are some examples:

- A political poll of voters is carried out. Each polled voter is asked whether or not they currently approve of the Prime Minister.

- A random sample of schools is obtained. The schools are assessed on their compliance with a suitable policy on sun exposure for their students.

- A random sample of police personnel are interviewed. Each person is assessed as to whether or not they show appropriate awareness of different cultures.

- A random sample of drivers are drug tested, and it is recorded whether or not they are positive for recent methamphetamine use.

- A random sample of footballers is chosen and their record of injuries assessed, according to whether or not they have had more than three episodes of concussion.

Consideration of these examples suggests that we are interested in the number of successes, in general, and not just the examination of individual responses. If we have n trials, we want to know how many of them are successes.

Binomial random variables

Define X to be the number of successes in n independent Bernoulli trials, each with probability of success p. We now derive the probability function of X.

First we note that X can take the values $0, 1, 2, \ldots, n$. Hence, there are $n + 1$ possible discrete values for X. We are interested in the probability that X takes a particular value x. If there are x successes observed, then there must be $n - x$ failures observed. One way for this to happen is for the first x Bernoulli trials to be successes, and the remaining $n - x$ to be failures. The probability of this is given by

$$\underbrace{p \times p \times \cdots \times p}_{x \text{ of these}} \times \underbrace{(1 - p) \times (1 - p) \times \cdots \times (1 - p)}_{n-x \text{ of these}} = p^x (1 - p)^{n-x}.$$

However, this is only one of the many ways that we can obtain x successes and $n - x$ failures. They could occur in exactly the opposite order: all $n - x$ failures first, then x successes. This outcome, which has the same number of successes as the first outcome but with a different arrangement, also has probability $p^x(1-p)^{n-x}$. There are many other ways to arrange x successes and $n - x$ failures among the n possible positions. Once we position the x successes, the positions of the $n - x$ failures are inevitable.

The number of ways of arranging x successes in n trials is equal to the number of ways of choosing x objects from n objects, which is equal to

$$\binom{n}{x} = \frac{n!}{x!(n - x)!} = \frac{n \times (n - 1) \times (n - 2) \times \cdots \times (n - x + 2) \times (n - x + 1)}{x \times (x - 1) \times (x - 2) \times \cdots \times 2 \times 1}.$$

(This is discussed in the module *The binomial theorem*.) Recall that 0! is defined to be 1, and that $\binom{n}{n} = \binom{n}{0} = 1$.

We can now determine the total probability of obtaining x successes. Each outcome with x successes and $n - x$ failures has individual probability $p^x(1 - p)^{n-x}$, and there are $\binom{n}{x}$ possible arrangements. Hence, the total probability of x successes in n independent Bernoulli trials is

$$\binom{n}{x} p^x (1 - p)^{n-x}, \quad \text{for } x = 0, 1, 2, \ldots, n.$$

This is the **binomial distribution**, and it is arguably the most important discrete distribution of all, because of the range of its application.

If X is the number of successes in n independent Bernoulli trials, each with probability of success p, then the probability function $p_X(x)$ of X is given by

$$p_X(x) = \Pr(X = x) = \binom{n}{x} p^x(1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, \ldots, n,$$

and X is a **binomial random variable** with parameters n and p. We write $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(n, p)$.

Notes.

- The probability of n successes out of n trials equals $\binom{n}{n} p^n(1-p)^0 = p^n$, which is the answer we expect from first principles. Similarly, the probability of zero successes out of n trials is the probability of obtaining n failures, which is $(1-p)^n$.

- Consider the two essential properties of any probability function $p_X(x)$:

  1 $p_X(x) \geq 0$, for every real number x.
  2 $\sum p_X(x) = 1$, where the sum is taken over all values x for which $p_X(x) > 0$.

  In the case that X is a binomial random variable, we clearly have $p_X(x) \geq 0$, since $0 < p < 1$. It looks like a more challenging task to verify that $\sum p_X(x) = 1$. However, we can use the binomial theorem:

  $$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r.$$

  (See the module *The binomial theorem*.) Note the form of each summand, and the similarity to the probability function of the binomial distribution. It follows by the binomial theorem that

  $$\sum p_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x(1-p)^{n-x} = \big((1-p) + p\big)^n = 1.$$

- If we view the formula

  $$p_X(x) = \binom{n}{x} p^x(1-p)^{n-x}$$

  in isolation, we may wonder whether $p_X(x) \leq 1$. After all, the number $\binom{n}{x}$ can be very large: for example, $\binom{300}{150} \approx 10^{89}$. But in these circumstances, the number $p^x(1-p)^{n-x}$ is very small: for example, $0.4^{150} \times 0.6^{150} \approx 10^{-93}$. So it seems that, in the formula for $p_X(x)$, we may be multiplying a very large number by a very small number. Can we be sure that the product $p_X(x)$ does not exceed one? The simplest answer is that we know that $p_X(x) \leq 1$ because we derived $p_X(x)$ from first principles as a probability. Alternatively, since $\sum p_X(x) = 1$ and $p_X(x) \geq 0$, for all possible values of X, it follows that $p_X(x) \leq 1$.

Example: Multiple-choice test

A teacher sets a short multiple-choice test for her students. The test consists of five questions, each with four choices. For each question, she uses a random number generator to assign the letters A, B, C, D to the choices. So for each question, the chance that any letter corresponds to the correct answer is equal to $\frac{1}{4}$, and the assignments of the letters are independent for different questions. One of her students, Barry, has not studied at all for the test. He decides that he might as well guess 'A' for each question, since there is no penalty for incorrect answers.

Let X be the number of questions that Barry gets right. His chance of getting a question right is just the chance that the letter 'A' was assigned to the correct answer, which is $\frac{1}{4}$. There are five questions. The outcome for each question can be regarded as 'correct' or 'not correct', and hence as one of two possible outcomes. Each question is independent. We therefore have the structure of a sequence of independent Bernoulli trials, each with probability of success (a correct answer) equal to $\frac{1}{4}$. So X has a binomial distribution with parameters 5 and $\frac{1}{4}$, that is, $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(5, \frac{1}{4})$.

The probability function of X is shown in figure 1.
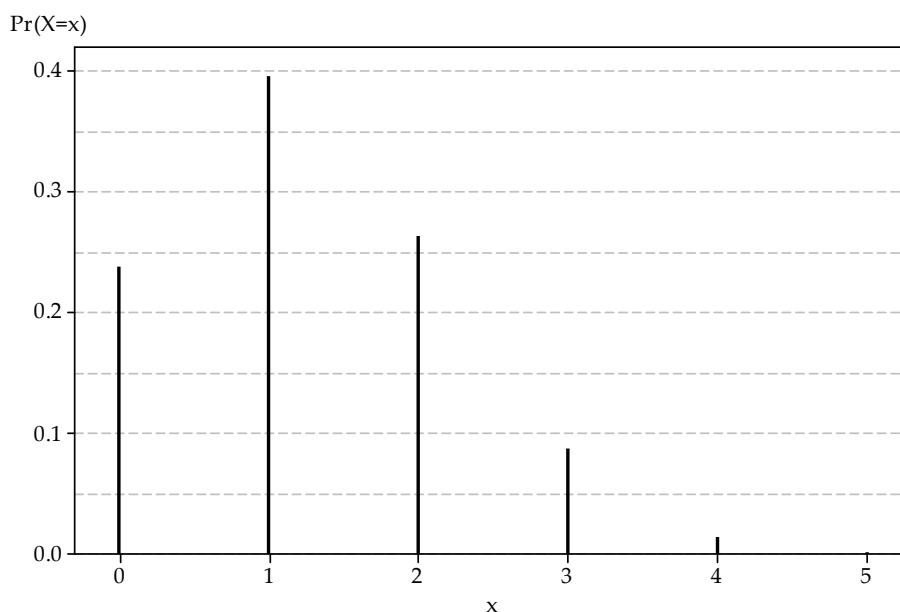
Figure 1: The probability function $p_X(x)$ for $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(5, \frac{1}{4})$.

We can see from the graph that Barry's chance of getting all five questions correct is very small; it is just visible on the graph as a very small value. On the other hand, the chance that he gets none correct (all wrong) is about 0.24, and the chance that he gets one correct is almost 0.4. What are these probabilities, more precisely?

The probability function of X is given by

$$p_X(x) = \binom{5}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{n-x}, \quad \text{for } x = 0, 1, 2, 3, 4, 5.$$

So $p_X(5) = \left(\frac{1}{4}\right)^5 \approx 0.0010$; Barry's chance of 'full marks' is one in a thousand. On the other hand, $p_X(0) = \left(\frac{3}{4}\right)^5 \approx 0.2373$ and $p_X(1) = 5 \times \frac{1}{4} \times \left(\frac{3}{4}\right)^4 \approx 0.3955$. Completing the distribution, we find that $p_X(2) \approx 0.2637$, $p_X(3) \approx 0.0879$ and $p_X(4) \approx 0.0146$.

In general, it is easier to calculate binomial probabilities using technology. Microsoft Excel provides a function for this, called BINOM.DIST. To use this function to calculate $p_X(x)$ for $X \overset{d}{=} \text{Bi}(n, p)$, four arguments are required: x, n, p, FALSE. For example, to find the value of $p_X(2)$ for $X \overset{d}{=} \text{Bi}(5, \frac{1}{4})$, enter the formula

=BINOM.DIST(2, 5, 0.25, FALSE)

in a cell; this should produce the result 0.2637 (to four decimal places). The somewhat off-putting argument FALSE is required to get the actual probability function $p_X(x)$. If TRUE is used instead, then the result is the cumulative probability $\sum_{k=0}^{x} p_X(k)$, which we do not consider here.

Exercise 1

Casey buys a Venus chocolate bar every day during a promotion that promises 'one in six wrappers is a winner'. Assume that the conditions of the binomial distribution apply: the outcomes for Casey's purchases are independent, and the population of Venus chocolate bars is effectively infinite.

a What is the distribution of the number of winning wrappers in seven days?

b Find the probability that Casey gets no winning wrappers in seven days.

c Casey gets no winning wrappers for the first six days of the week. What is the chance that he will get a winning wrapper on the seventh day?

d Casey buys a bar every day for six weeks. Find the probability that he gets at least three winning wrappers.

e How many days of purchases are required so that Casey's chance of getting at least one winning wrapper is 0.95 or greater?

Exercise 2

Which of the following situations are suitable for modelling with the binomial distribution? For those which are not, explain what the problem is.

a A normal six-sided die is rolled 15 times, and the number of fours obtained is observed.

b For a period of two weeks in a particular city, the number of days on which rain occurs is recorded.

c There are five prizes in a raffle of 100 tickets. Each ticket is either blue, green, yellow or pink, and there are 25 tickets of each colour. The number of blue tickets among the five winning tickets drawn in the raffle is recorded.

d Twenty classes at primary schools are chosen at random, and all the students in these classes are surveyed. The total number of students who like their teacher is recorded.

e Genetic testing is performed on a random sample of 50 Australian women. The number of women in the sample with a gene associated with breast cancer is recorded.

Exercise 3

Ecologists are studying the distributions of plant species in a large forest. They do this by choosing n quadrats at random (a quadrat is a small rectangular area of land), and examining in detail the species found in each of these quadrats. Suppose that a particular plant species is present in a proportion k of all possible quadrats, and distributed randomly throughout the forest. How large should the sample size n be, in order to detect the presence of the species in the forest with probability at least 0.9?

It is useful to get a general picture of the binomial distribution. Figure 2 shows nine binomial distributions, each with $n = 20$; the values of p range from 0.01 to 0.99.
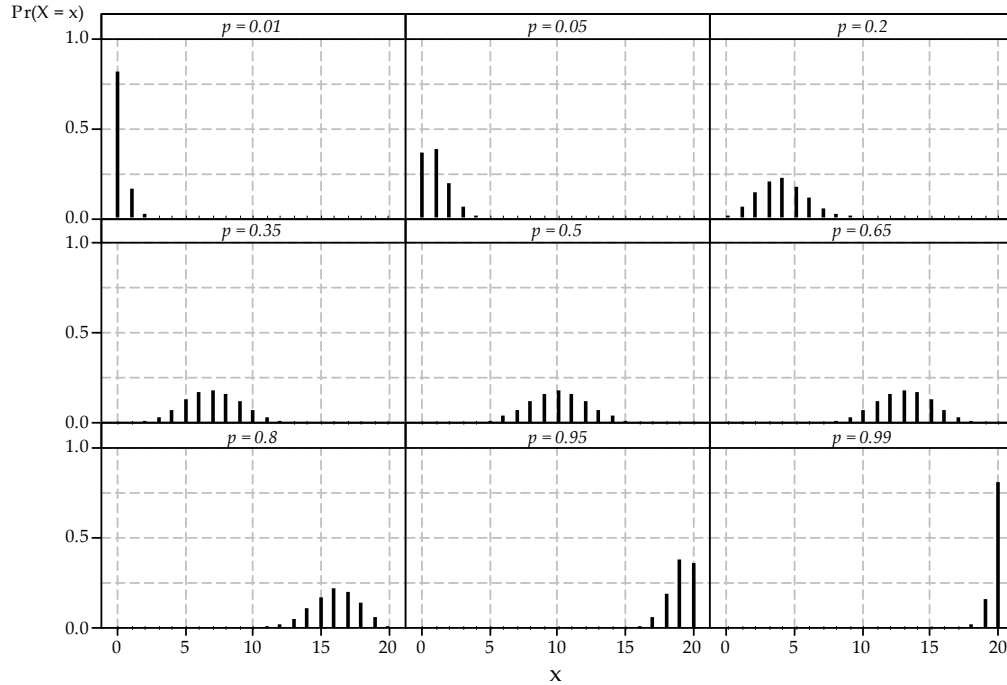

Figure 2: Nine binomial distributions with parameters $n = 20$ and p as labelled.

A number of features of this figure are worth noting:

- First, we may wonder: Where are all the other probabilities? We know that a binomial random variable can take any value from 0 to n. Here $n = 20$, so there are 21 possible values. But in all of the graphs, there are far fewer than 21 spikes showing. Why? The reason is simply that many of the probabilities are too small to be visible on the scale shown. The smallest visible probabilities are about 0.01; many of the probabilities here are 0.001 or smaller, and therefore invisible. For example (taking an extreme case that is easy to work out), for the top-left graph where $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(20, 0.01)$, we have $\Pr(X = 20) = 0.01^{20} = 10^{-40}$.

- When p is close to 0 or 1, the distribution is not symmetric. When p is close to 0.5, it is closer to being symmetric. When $p = 0.5$, it is symmetric.

- Consider the pairs of distributions for $p = \theta$ and $p = 1 - \theta$: $p = 0.01$ and $p = 0.99$; $p = 0.05$ and $p = 0.95$; $p = 0.2$ and $p = 0.8$; $p = 0.35$ and $p = 0.65$. Look carefully at the two distributions for any one of these pairs. The two distributions are mirror images, reflected about $x = 10$. This follows from the nature of the binomial distribution.

  According to the definition of the binomial distribution, we count the number of successes. Since each of the Bernoulli trials has only two possible outcomes, it must be equivalent, in some sense, to count the number of failures. If we know the number of

successes is equal to x, then clearly the number of failures is just $n - x$. It follows that, if $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(n, p)$ and Y is the number of failures, then $Y = n - X$ and $Y \stackrel{\mathrm{d}}{=} \mathrm{Bi}(n, 1 - p)$. Note that X and Y are not independent.

- Finally, the *spread* of the binomial distribution is smaller when p is close to 0 or 1, and the spread is greatest when $p = 0.5$.

Mean and variance

Let $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(n, p)$. We now consider the mean and variance of X.

The mean μ_X is np, a result we will shortly derive. Note that this value for the mean is intuitively compelling. If we are told that 8% of the population is left-handed, then on average how many left-handers will there be in a sample of 100? We expect, on average, 8% of 100, which is 8. In a sample of 200, we expect, on average, 8% of 200, which is 16. The formula we are applying here is $n \times p$, where n is the sample size and p the proportion of the population with the binary characteristic (in this example, being left-handed).

We will use the following two general results without proving them: one for the mean of a sum of random variables, and the other for the variance of a sum of independent random variables.

1 For n random variables $X_1, X_2, \ldots, X_n$, we have
$$E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).$$

2 If $Y_1, Y_2, \ldots, Y_n$ are *independent* random variables, then
$$\mathrm{var}(Y_1 + Y_2 + \cdots + Y_n) = \mathrm{var}(Y_1) + \mathrm{var}(Y_2) + \cdots + \mathrm{var}(Y_n).$$

Note the important difference in the conditions of these two results. The mean of the sum equals the sum of the means, without any conditions on the random variables. The corresponding result for the variance, however, requires that the random variables are independent.

Recall that a binomial random variable is defined to be the number of successes in a sequence of n independent Bernoulli trials, each with probability of success p. Let B_i be the number of successes on the ith trial. Then B_i is a Bernoulli random variable with parameter p. Counting the total number of successes over n trials is equivalent to summing the B_i's. In the section *Bernoulli trials*, we showed that $E(B_i) = p$. Since $X = B_1 + B_2 + \cdots + B_n$, it follows that

$$E(X) = E(B_1) + E(B_2) + \cdots + E(B_n)$$

$$= p + p + \cdots + p$$

$$= np.$$

We may also establish this result in a far less elegant way, using the probability function of X. This requires a result involving combinations: for $a \geq b \geq 1$, we have

$$
\begin{aligned}
\binom{a}{b} &= \frac{a!}{(a-b)!b!} \\
&= \frac{a(a-1)\cdots(a-b+1)}{b!} \\
&= \frac{a}{b}\frac{(a-1)(a-2)\cdots(a-b+1)}{(b-1)!} \\
&= \frac{a}{b}\binom{a-1}{b-1}.
\end{aligned}
$$

We can use this result to calculate $\mathrm{E}(X)$ directly:

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{n} x\, p_X(x) && \text{(definition of expected value)} \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1}(1-p)^{(n-1)-(x-1)} && \text{(using the result above)} \\
&= np \sum_{y=0}^{m} \binom{m}{y} p^y (1-p)^{m-y} && \text{(where } m = n-1 \text{ and } y = x-1) \\
&= np \times 1 && \text{(by the binomial theorem)} \\
&= np.
\end{aligned}
$$

Now we consider the variance of X. It is possible, but cumbersome, to derive the variance directly, using the definition $\mathrm{var}(X) = \mathrm{E}[(X - \mu_X)^2]$ and a further result involving combinations. Instead, we will apply the general result for the variance of a sum of independent random variables. As before, we note that $X = B_1 + B_2 + \cdots + B_n$, where $B_1, B_2, \ldots, B_n$ are *independent* Bernoulli random variables with parameter p. In the section *Bernoulli trials*, we showed that $\mathrm{var}(B_i) = p(1-p)$. Hence,

$$
\begin{aligned}
\mathrm{var}(X) &= \mathrm{var}(B_1) + \mathrm{var}(B_2) + \cdots + \mathrm{var}(B_n) \\
&= p(1-p) + p(1-p) + \cdots + p(1-p) \\
&= np(1-p).
\end{aligned}
$$

It follows that, for $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(n, p)$, the standard deviation of X is given by

$$
\mathrm{sd}(X) = \sqrt{np(1-p)}.
$$

Note that the spread of the distribution of X, reflected in the formulas for $\text{var}(X)$ and $\text{sd}(X)$, is the same for $X \overset{\text{d}}{=} \text{Bi}(n, p)$ and $X \overset{\text{d}}{=} \text{Bi}(n, 1 - p)$. This agrees with the pattern observed in figure 2: the distribution for $p = \theta$ is a mirror image of the distribution for $p = 1 - \theta$, and therefore has the same spread.

Exercise 4

Consider the nine binomial distributions represented in figure 2.

a   Determine the mean and standard deviation of X in each case.

b   Among the nine distributions, when is the standard deviation smallest? When is it largest?

Exercise 5

Suppose that $X \overset{\text{d}}{=} \text{Bi}(n, p)$.

a   Sketch the graph of the variance of X as a function of p.

b   Using calculus, or otherwise, show that the variance is largest when $p = 0.5$.

c   Find the variance and standard deviation of X when $p = 0.5$.

Exercise 6

Let $0 < p < 1$ and suppose that $X \overset{\text{d}}{=} \text{Bi}(n, p)$. Consider the following claim:

As n tends to infinity, the largest value of $p_X(x)$ tends to zero.

Is this true? Explain.

Answers to exercises

Exercise 1

a   Let X be the number of winning wrappers in one week (7 days). Then $X \overset{\text{d}}{=} \text{Bi}(7, \frac{1}{6})$.

b $\Pr(X = 0) = \left(\frac{5}{6}\right)^7 \approx 0.279$.

c   The purchases are assumed to be independent, and hence the outcome of the first six days is irrelevant. The probability of a winning wrapper on any given day (including on the seventh day) is $\frac{1}{6}$.

d Let Y be the number of winning wrappers in six weeks (42 days). Then $Y \overset{\mathrm{d}}{=} \mathrm{Bi}(42, \frac{1}{6})$. Hence,

$$\Pr(Y \geq 3) = 1 - \Pr(Y \leq 2)$$

$$= 1 - \left[\Pr(Y = 0) + \Pr(Y = 1) + \Pr(Y = 2) \right]$$

$$= 1 - \left[\left(\tfrac{5}{6}\right)^{42} + 42 \times \tfrac{1}{6} \times \left(\tfrac{5}{6}\right)^{41} + 861 \times \left(\tfrac{1}{6}\right)^{2} \times \left(\tfrac{5}{6}\right)^{40} \right]$$

$$\approx 1 - \left[0.0005 + 0.0040 + 0.0163 \right] \approx 0.979.$$

This could also be calculated with the help of the BINOM.DIST function in Excel.

e Let U be the number of winning wrappers in n purchases. Then $U \overset{\mathrm{d}}{=} \mathrm{Bi}(n, \frac{1}{6})$ and so

$$\Pr(U \geq 1) \geq 0.95 \iff 1 - \Pr(U = 0) \geq 0.95$$

$$\iff \Pr(U = 0) \leq 0.05$$

$$\iff \left(\tfrac{5}{6}\right)^{n} \leq 0.05$$

$$\iff n \log_e\left(\tfrac{5}{6}\right) \leq \log_e(0.05)$$

$$\iff n \geq \frac{\log_e(0.05)}{\log_e\left(\tfrac{5}{6}\right)} \approx 16.431.$$

Hence, he needs to purchase a bar every day for 17 days.

Exercise 2

a Binomial.

b Not binomial. The weather from one day to the next is not independent. Even though, for example, the chance of rain on a day in March may be 0.1 based on long-term records for the city, the days in a particular fortnight are not independent.

c Not binomial. When sampling without replacement from a small population of tickets, the chance of a 'success' (in this case, a blue ticket) changes with each ticket drawn.

d Not binomial. There is likely to be clustering within classes; the students within a class are not independent from each other.

e Binomial.

Exercise 3

This problem is like that of Casey and the winning wrappers. Let X be the number of sampled quadrats containing the species of interest. We assume that $X \overset{\mathrm{d}}{=} \mathrm{Bi}(n, k)$. We

want $\Pr(X \geq 1) \geq 0.9$, since we only need to observe the species in one quadrat to know that it is present in the forest. We have

$$\Pr(X \geq 1) \geq 0.9 \iff \Pr(X = 0) \leq 0.1$$

$$\iff (1 - k)^n \leq 0.1$$

$$\iff n \geq \frac{\log_e(0.1)}{\log_e(1 - k)}.$$

For example: if $k = 0.1$, we need $n \geq 22$; if $k = 0.05$, we need $n \geq 45$; if $k = 0.01$, we need $n \geq 230$.

Exercise 4

a

p	0.01	0.05	0.2	0.35	0.5	0.65	0.8	0.95	0.99
μ_X	0.2	1	4	7	10	13	16	19	19.8
$\mathrm{sd}(X)$	0.445	0.975	1.789	2.133	2.236	2.133	1.789	0.975	0.445

b   The standard deviation is smallest when $|p - 0.5|$ is largest; in this set of distributions, when $p = 0.01$ and $p = 0.99$. The standard deviation is largest when $p = 0.5$.

Exercise 5

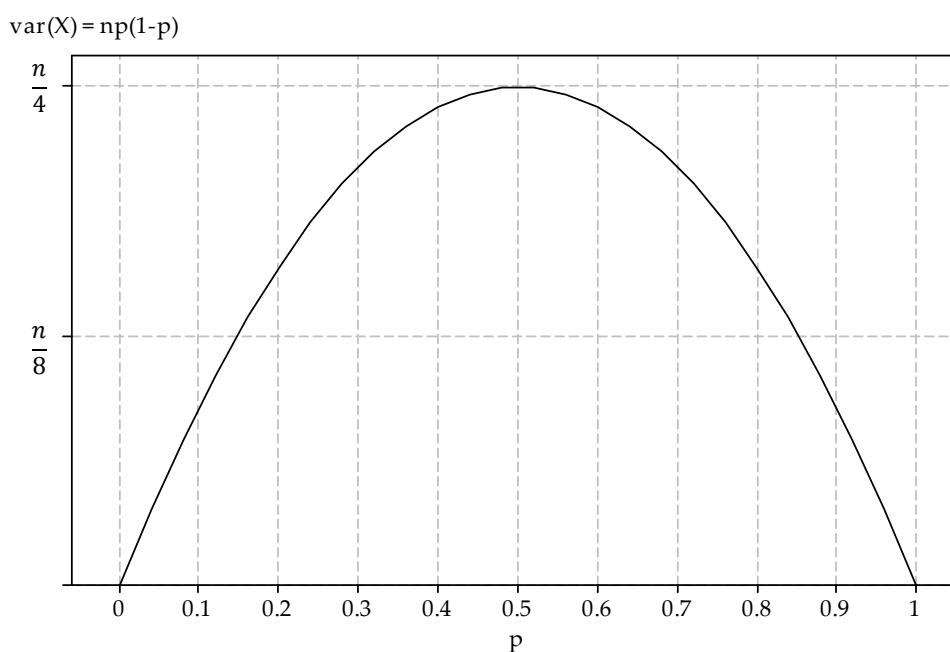a   The graph of the variance of X as a function of p is as follows.

Figure 3: Variance of X as a function of p, where $X \stackrel{\mathrm{d}}{=} \mathrm{Bi}(n, p)$.

b The variance function is $f(p) = np(1-p)$. So $f'(p) = n(1-2p)$. Hence $f'(p) = 0$ when $p = 0.5$, and this is clearly a maximum.

c When $p = 0.5$, we have $\text{var}(X) = \frac{n}{4}$ and $\text{sd}(X) = \frac{\sqrt{n}}{2}$.

Exercise 6

Yes, it is true that, as n tends to infinity, the largest value of $p_X(x)$ tends to zero. We will not prove this explicitly, but make the following two observations.

● First consider the case where $p = 0.5$ and n is even; let $n = 2m$. The largest value of $p_X(x)$ occurs when $x = \frac{n}{2} = m$ (see figure 2, for example). This probability is given by

$$p_X(m) = \binom{2m}{m} \frac{1}{4^m}.$$

This tends to zero as m tends to infinity. Proving this fact is beyond the scope of the curriculum (it can be done using Stirling's formula, for example), but evaluating the probabilities for large values of m will suggest that it is true:

- for $m = 1000$, $p_X(m) \approx 0.0178$
- for $m = 10\,000$, $p_X(m) \approx 0.0056$
- for $m = 100\,000$, $p_X(m) \approx 0.0018$
- for $m = 1\,000\,000$, $p_X(m) \approx 0.0006$.

● In general, for $X \stackrel{\text{d}}{=} \text{Bi}(n, p)$, we have $\text{sd}(X) = \sqrt{np(1-p)}$. As n tends to infinity, the standard deviation grows larger and larger, which means that the distribution is more and more spread out, leading to the largest probability gradually diminishing in size.

0   1   2   3   4   5   6   7   8   9   10   11   12