Review exercise 1

1 **a** The probability distribution for X is:

x	1	2	3	4	5	6
P(*X* = *x*)	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

b $P(2 < X \leqslant 5) = P(X = 3) + P(X = 4) + P(X = 5)$

$$= \frac{5+7+9}{36} = \frac{21}{36} = \frac{7}{12} = 0.5833 \text{ (4 d.p.)}$$

c $E(X) = \frac{1}{36}(1 + 2\times 3 + 3\times 5 + 4\times 7 + 5\times 9 + 6\times 11) = \frac{161}{36} = 4.4722 \text{ (4 d.p.)}$

d Show all the steps when asked to show that Var(X) = 1.97

$E(X^2) = \sum x^2 P(X = x)$

$$= \frac{1}{36}(1 + 2^2\times 3 + 3^2\times 5 + 4^2\times 7 + 5^2\times 9 + 6^2\times 11)$$

$$= \frac{791}{36}$$

$\mathrm{Var}(X) = E(X^2) - (E(X))^2$

$$= \frac{791}{36} - \frac{25921}{1296} = \frac{28476 - 25921}{1296} = \frac{2555}{1296} = 1.97 \text{ (3 s.f.)}$$

e Using $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$, so:

$$\mathrm{Var}(2 - 3X) = (-3)^2\,\mathrm{Var}(X) = 9 \times \frac{2555}{1296} = 17.7 \text{ (3 s.f.)}$$

2 The probability distribution for X is:

x	1	2	3	4	5
P(*X* = *x*)	k	$2k$	$3k$	$5k$	$6k$

a Probabilities sum to 1, so:

$$k + 2k + 3k + 5k + 6k = 1$$

$$17k = 1$$

$$k = \frac{1}{17} = 0.0588 \text{ (4 d.p.)}$$

b As the question requires an exact answer work in fractions

$$E(X) = \frac{1}{17}(1 + 2\times 2 + 3\times 3 + 4\times 5 + 5\times 6) = \frac{64}{17}$$

2 **c** Show all the steps when asked to show that Var(X) = 1.47

$$\mathrm{E}(X^2) = \sum x^2\,\mathrm{P}(X = x)$$
$$= \frac{1}{17}(1 + 2^2 \times 2 + 3^2 \times 3 + 4^2 \times 5 + 5^2 \times 6)$$
$$= \frac{266}{17}$$
$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$
$$= \frac{266}{17} - \frac{4096}{289} = \frac{4522 - 4096}{289} = \frac{426}{289} = 1.47 \text{ (3 s.f.)}$$

d Using Var($aX + b$) = a^2Var(X), so:

$$\mathrm{Var}(4 - 3X) = (-3)^2\,\mathrm{Var}(X) = 9 \times \frac{426}{289} = 13.3 \text{ (1 d.p.)}$$

3 **a** Probabilities sum to 1, so:

$0.1 + p + 0.2 + q + 0.3 = 1$

$\Rightarrow p + q = 0.4$ **(1)**

$\mathrm{E}(X) = \sum x\,\mathrm{P}(X = x) = 3.5$

So $1 \times 0.1 + 2 \times p + 3 \times 0.2 + 4 \times q + 5 \times 0.3 = 3.5$

$\Rightarrow 2p + 4q = 1.3$ **(2)**

b Multiply equation (1) by 2

$2p + 2q = 0.8$ **(3)**

Subtract equation (3) from equation (2) gives

$2q = 0.5 \Rightarrow q = 0.25$

Substitute value for q into equation (1)

$p + 0.25 = 0.4 \Rightarrow p = 0.15$

Solution: $p = 0.15,\ q = 0.25$

c $\mathrm{E}(X^2) = \sum x^2\,\mathrm{P}(X = x) = 1^2 \times 0.10 + 2^2 \times 0.15 + 3^2 \times 0.2 + 4^2 \times 0.25 + 5^2 \times 0.30 = 14$

$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 14 - 3.5^2 = 1.75$

d Using Var($aX + b$) = a^2Var(X), so:

$\mathrm{Var}(3 - 2X) = (-3)^2\,\mathrm{Var}(X) = 4 \times 1.75 = 7$

4 **a** Probabilities sum to 1, so:

$0.2 + p + 0.2 + q + 0.15 = 1$

$\Rightarrow p + q = 0.45$ **(1)**

$\mathrm{E}(X) = \sum x\,\mathrm{P}(X = x) = 4.5$

So $1 \times 0.2 + 3 \times p + 5 \times 0.2 + 7 \times q + 9 \times 0.15 = 4.5$

$\Rightarrow 3p + 7q = 1.95$ **(2)**

4 b Multiply equation (1) by 3

$3p + 3q = 1.35 \qquad \textbf{(3)}$

Subtract equation (3) from equation (2) gives

$4q = 0.6 \Rightarrow q = 0.15$

Substitute value for q into equation (1)

$p + 0.15 = 0.45 \Rightarrow p = 0.3$

Solution: $p = 0.3,\ q = 0.15$

c $\text{P}(4 < X \leqslant 7) = \text{P}(X = 5) + \text{P}(X = 7)$

$= 0.2 + q = 0.35$

d $\text{Var}(X) = \text{E}(X^2) - (\text{E}(X))^2 = 27.4 - 4.5^2 = 7.15$

e Using $\text{E}(aX + b) = a\text{E}(X) + b$ so:

$\text{E}(19 - 4X) = -4\text{E}(X) + 19 = 19 - (4 \times 4.5) = 1$

f Using $\text{Var}(aX + b) = a^2\text{Var}(X)$, so:

$\text{Var}(19 - 4X) = (-4)^2\,\text{Var}(X) = 16 \times 7.15 = 114.4$

5 a Probabilities sum to 1, so:

$0.2 + 0.3 + a + b = 1$

$\Rightarrow a + b = 0.5$

$\text{E}(Y) = 2 - 3\text{E}(X) = 2.9$

$\Rightarrow \text{E}(X) = -0.3$

So $-2(0.2) - 1(0.3) + b = -0.3$

$\Rightarrow b = 0.4$

$\Rightarrow a = 0.1$

b $\text{E}\left(X^2\right) = 0.2(-2)^2 + 0.3(-1)^2 + 0.4(1)^2 = 1.5$

$\text{Var}(X) = \text{E}\left(X^2\right) - (\text{E}(X))^2 = 1.5 - (-0.3)^2 = 1.41$

c Using $\text{Var}(aX + b) = a^2\text{Var}(X)$:

$\text{Var}(Y) = \text{Var}(2 - 3X) = (-3)^2\,\text{Var}(X) = 9 \times 1.41 = 12.69$

d $Y + 1 = 3 - 3X$

So $Y + 1 < X \Rightarrow 3 - 3X < X \Rightarrow X > \frac{3}{4}$

$\text{P}(Y + 1 < X) = \text{P}\left(X > \frac{3}{4}\right) = \text{P}(X = 1) = 0.4$

6 a Probabilities sum to 1, so:

$2a + 2b + c = 1$ **(1)**

$\mathrm{E}(Y) = \frac{1}{4} - \frac{1}{2}\mathrm{E}(X) = 0.25 - 0.5\mathrm{E}(X)$

So $0.25 - 0.5(-3a - 2b + a + 3c) = -0.05$

$\Rightarrow -2a - 2b + 3c = 0.6$ **(2)**

$Y > 0 \Rightarrow \frac{1-2X}{4} > 0 \Rightarrow X < 0.5$

So $\mathrm{P}(Y > 0) = \mathrm{P}(X < 0.5) = 0.5$

$\Rightarrow a + 2b = 0.5$ **(3)**

Add equations (1) and (2) to get :

$4c = 1.6 \Rightarrow c = 0.4$

Substitute for c in equation (1), and subtract (3) from (1):

$a + 0.4 = 0.5 \Rightarrow a = 0.1$

Substitute for a in equation (3):

$0.1 + 2b = 0.5 \Rightarrow b = 0.2$

$\Rightarrow b = 0.2$

So the probability distribution for X is:

x	–3	–2	0	1	3
P(*X* = *x*)	0.1	0.2	0.2	0.1	0.4

b $-3X < 5Y \Rightarrow -3X < 5\left(\frac{1-2X}{4}\right) \Rightarrow -12X < 5 - 10X \Rightarrow X > -\frac{5}{2}$

So $\mathrm{P}(-3X < 5Y) = \mathrm{P}(X > -2.5) = 1 - \mathrm{P}(X = -3) = 0.9$

7 a Let the random variable X denote the number of accidents per week on the stretch of motorway. The term ‘rate’ used in the question indicates this is a Poisson model, so $X \sim \mathrm{Po}(1.5)$

b $\mathrm{P}(X = 2) = \frac{\mathrm{e}^{-1.5}1.5^2}{2!} = 0.2510$ (4 d.p.)

Note that as X is a discrete variable, $\mathrm{P}(X = 2) = \mathrm{P}(X \leqslant 2) - \mathrm{P}(X \leqslant 1)$ and can therefore be calculated using the tables:

$\mathrm{P}(X = 2) = \mathrm{P}(X \leqslant 2) - \mathrm{P}(X \leqslant 1) = 0.8088 - 0.5578 = 0.2510$

c $\mathrm{P}(X \geqslant 1) = 1 - \mathrm{P}(X = 0) = 1 - 0.2231 = 0.7769$ (from the tables)

P(at least one accident per week for 3 weeks) $= \mathrm{P}(X \geqslant 1) \times \mathrm{P}(X \geqslant 1) \times \mathrm{P}(X \geqslant 1)$

$= 0.7760^3 = 0.4689$

7 d Let the random variable Y denote the number of accidents in a two-week period on the stretch of motorway, so $X \sim \text{Po}(3)$

From the tables

$\text{P}(X > 4) = 1 - \text{P}(X \leqslant 4) = 1 - 0.8153 = 0.1847$

8 a There is no context stated, but a Poisson distribution requires an event to occur. For a Poisson distribution to be a suitable model, events should occur at a constant rate; they should occur independently or randomly; and they should occur singly.

b Let the random variable X denote the number of cars passing the point in a 60-minute period, so $X \sim \text{Po}(6)$

i $\text{P}(X = 4) = \dfrac{e^{-6}6^4}{4!} = 0.1339$ (4 d.p.)

Note that as X is a discrete variable, $\text{P}(X = 4) = \text{P}(X \leqslant 4) - \text{P}(X \leqslant 3)$ and can therefore be calculated using the tables:

$\text{P}(X = 4) = \text{P}(X \leqslant 4) - \text{P}(X \leqslant 3) = 0.2851 - 0.1512 = 0.1339$

ii $$\begin{aligned}\text{P}(X \geqslant 5) &= 1 - \text{P}(X \leqslant 4)\\ &= 1 - 0.2851\\ &= 0.7149\\ &= 0.715 \text{ (3 s.f.)}\end{aligned}$$

c Let the random variable Y denote the number of cars and other vehicles passing the observation point in a 10-minute period. On average 2 other vehicles $(12 \div 6)$ pass the point so $Y \sim \text{Po}(1+2)$, i.e. $Y \sim \text{Po}(3)$

$\text{P}(Y = 1) = 3e^{-3} = 0.1494$ (4 d.p.)

Alternatively, let the random variable Z denote the number of other vehicles passing the observation point in a 10-minute period, so $Z \sim \text{Po}(2)$

$$\begin{aligned}\text{P(1 car and 0 other)} + \text{P(0 car and 1 other)} &= \text{P}(X = 1)\text{P}(Z = 0) + \text{P}(X = 0)\text{P}(Z = 1)\\ &= e^{-1} \times e^{-2} + e^{-1} \times 2e^{-2}\\ &= 0.3679 \times 0.1353 + 0.3679 \times 0.2707\\ &= 0.1494 \text{ (4 d.p.)}\end{aligned}$$

9 a Let the random variable X denote the number of lawn-mowers hired out by *Quikmow* in a one-hour period, so $X \sim \text{Po}(1.5)$; and let the random variable Y denote the number of lawn-mowers hired out by *Easitrim* in a one-hour period, so $Y \sim \text{Po}(2.2)$

As the variables are independent $\text{P}((X = 1) \cap (Y = 1)) = \text{P}(X = 1) \times \text{P}(Y = 1)$

$\text{P}(X = 1) \times \text{P}(Y = 1) = \dfrac{e^{-1.5}1.5^1}{1!} \times \dfrac{e^{-2.2}2.2^1}{1!} = 0.3347 \times 0.2438 = 0.0816$ (4 d.p.)

b Let the random variable Z denote the number of lawn-mowers hired out by *Quikmow* and *Easitrim* in a one-hour period, so $Z \sim \text{Po}(1.5 + 2.2)$, i.e. $Z \sim \text{Po}(3.7)$

$\text{P}(Z = 4) = \dfrac{e^{-3.7}3.7^4}{4!} = 0.1931$ (4 d.p.)

9 c Let the random variable *M* denote the number of lawn-mowers hired out by *Quikmow* and *Easitrim* in a one-hour period, so $M \sim \text{Po}(3\times 3.7)$, i.e. $M \sim \text{Po}(11.1)$

By calculator $\text{P}(M < 12) = \text{P}(M \leqslant 11) = 0.5673$ (4 d.p.)

10 a Mean $= \bar{x} = \frac{\sum x}{n} = \frac{290}{200} = 1.45$

Variance $= \frac{\sum x^2}{n} - (\bar{x})^2 = \frac{702}{200} - 2.1025 = 1.4075$

b The fact that the mean is close to the variance supports the use of a Poisson distribution.

c Let the random variable X denote the number of toys in a cereal box and use the model, $X \sim \text{Po}(1.45)$

By calculator $\text{P}(X \geqslant 2) = 1 - \text{P}(X \leqslant 1) = 1 - 0.5747 = 0.4253$ (4 d.p.)

11 a If $X \sim \text{B}(n, p)$ and *n* is large and *p* is small, then *X* can be approximated by Po(*np*).

b $\text{P(2 consecutive calls connected to wrong agent)} = 0.01 \times 0.01 = 0.0001\text{P}$

c Let the random variable *X* denoted the number of calls wrongly connected in 5 consecutive calls, so $X \sim B(5, 0.01)$

$$\text{P}(X > 1) = 1 - \text{P}(X = 1) - \text{P}(X = 0) = 1 - \binom{5}{1}\times(0.01)(0.99)^4 - \binom{5}{0}\times(0.01)^0(0.99)^5$$
$$= 1 - 0.048\,03 - 0.950\,99 = 0.000\,98 \text{ (5 d.p.)}$$

d Let the random variable Y denoted the number of calls wrongly connected in a day, so $Y \sim \text{B}(1000, 0.01)$

Mean $= \bar{Y} = np = 10$ Variance $= np(1-p) = 10 \times 0.99 = 9.9$

e Approximate the binomial distribution using $X \sim \text{Po}(np)$, i.e. $X \sim \text{Po}(10)$, and use tables

$\text{Po}(X > 6) = 1 - \text{Po}(X \leqslant 6) = 1 - 0.1301 = 0.8699 = 0.870$ (3 d.p.)

12 a $\text{P}(X = 3) = \binom{150}{3}\times(0.02)^3(0.98)^{147} = 0.2263$ (4 d.p.)

b $\lambda = np = 150 \times 0.02 = 3$

The Poisson approximation is justified in this case because *n* is large and *p* is small.

13 a $X \sim \text{B}(200, 0.015)$

b $\text{P}(X = 4) = \binom{200}{4}\times(0.015)^4(0.985)^{196} = 0.1693$ (4 d.p.)

c If $X \sim \text{B}(n, p)$ and *n* is large and *p* is small, then *X* can be approximated by Po(*np*), so in this case $X \approx\sim \text{Po}(200\times 0.015)$, i.e. $X \approx\sim \text{Po}(3)$

13 d $P(X=4)=\dfrac{e^{-3}3^4}{4!}=0.1680$ (4 d.p.)

The percentage error $=\dfrac{0.1693-0.1680}{0.1693}\times 100=0.77\%$

14 a $X \sim \text{Geo}(0.05)$

b $E(X)=\dfrac{1}{p}=\dfrac{1}{0.05}=20$

$\text{Var}(X)=\dfrac{1-p}{p^2}=\dfrac{1-0.05}{0.05^2}=380$

c $P(X\geqslant 15)=(1-p)^{x-1}=(1-0.05)^{14}=0.4877$ (4 d.p.)

15 a Geometric

b $E(Y)=\dfrac{1}{p}=10\Rightarrow p=0.1$

So $Y \sim \text{Geo}(0.1)$

$P(Y=7)=(0.1)(1-0.1)^6=0.0531$ (4 d.p.)

c $\text{Var}(Y)=\dfrac{1-p}{p^2}=\dfrac{1-0.1}{0.1^2}=90$

d Each attempt is independent and the probability of scoring a basket is constant for each attempt.

16 a Let X be the number of spins needed to win a prize, then $X \sim \text{Geo}\left(\dfrac{1}{12}\right)$

Since a player is only allowed 10 spins, the probability of winning a prize is:

$P(X\leqslant 10)=1-(1-p)^{10}=1-\left(\dfrac{11}{12}\right)^{10}=0.5811$ (4 d.p.)

b Let n be the minimum number of spins such that the probability of winning is at least 0.75.

So $P(X\leqslant n)=1-(1-p)^n=1-\left(\dfrac{11}{12}\right)^n\geqslant 0.75$

$\Rightarrow \left(\dfrac{11}{12}\right)^n\leqslant 0.25$

$\Rightarrow \log\left(\dfrac{11}{12}\right)n\leqslant \log 0.25$

$\Rightarrow n\geqslant \dfrac{\log 0.25}{\log\left(\frac{11}{12}\right)}$ (changing the inequality because dividing by a negative number)

$\Rightarrow n\geqslant 15.93$

So Keisha should allow a minimum of 16 spins for the probability of winning to be greater than 0.75.

17 a Let the random variable X denote the number of games it takes Matt to win 3 prizes, then $X \sim \text{Negative B}(3, 0.18)$

$$P(X = 13) = \binom{12}{2} \times (0.18)^3 (0.82)^{10} = 0.0529 \text{ (4 d.p.)}$$

b The model in part **a** assumes that each game is independent and that there is an equal probability of winning each game.

c Let the random variable Z denote the number of games it takes Matt to win 4 prizes, then $Z \sim \text{Negative B}(4, 0.18)$

$$E(Z) = \mu = \frac{r}{p} = \frac{4}{0.18} = 22.2 \text{ (3 s.f.)}$$

$$\text{Var}(Z) = \sigma^2 = \frac{r(1-p)}{p^2} = \frac{4(0.82)}{0.18^2} = 101.23\ldots$$

$$\text{Standard deviation } = \sigma = \sqrt{101.23\ldots} = 10.1 \text{ (3 s.f.)}$$

d $Y \sim \text{Negative B}(r, p)$

$$E(Y) = \frac{r}{p} = 20 \Rightarrow r = 20p$$

$$\text{Var}(Z) = \frac{r(1-p)}{p^2} = 113\frac{1}{3} = \frac{340}{3} \Rightarrow \frac{20p(1-p)}{p^2} = \frac{340}{3}$$

$$\text{So } 3p - 3p^2 = 17p^2 \Rightarrow p(20p - 3) = 0$$

$$\text{Therefore } p = \frac{3}{20} = 0.15$$

18 a Let the random variable X denote the number of rolls of the dice it takes to throw 3 threes, so $X \sim \text{Negative B}(3, p)$

$$\text{Var}(X) = \frac{r(1-p)}{p^2} = 23\frac{1}{3} \Rightarrow \frac{3(1-p)}{p^2} = \frac{70}{3}$$

$$\text{So } 9 - 9p = 70p^2 \Rightarrow 70p^2 + 9p - 9 = 0$$

Solve by factoring or by using the quadratic equation.

Factoring gives $(10p - 3)(7p + 3) = 0$

$$\text{As } p > 0 \Rightarrow p = \frac{3}{10} = 0.3$$

b $P(X = 9) = \binom{8}{2} \times (0.3)^3 (0.7)^6 = 0.0889 \text{ (4 d.p.)}$

c As each roll is independent, this requires finding the probability that it takes 10 more rolls of the dice to throw 2 further threes. If Y is the number of times the dice is rolled until 3 sixes have occurred, then $Y \sim \text{Negative B}(2, 0.3)$ and the required probability is $P(Y = 10)$

$$P(Y = 10) = \binom{9}{1} \times (0.3)^2 (0.7)^8 = 0.0467 \text{ (4 d.p.)}$$

19 a i A hypothesis test is where the value of a population parameter (whose assumed value is given by the null hypothesis H_0) is tested against what value it takes if H_0 is rejected (this could be an increase, a decrease or a change).

ii A range of values of a test statistic that would lead to the rejection of the null hypothesis (H_0).

b Let the random variable X be the number of incoming calls received in a 20-minute interval, so $X \sim \text{Po}(20 \times 0.45)$, i.e. $X \sim \text{Po}(9)$

$H_0 : \lambda = 9 \quad H_1 : \lambda \neq 9$

Assume H_0, so that $X \sim \text{Po}(9)$

Let $X = c_1$ be the upper boundary of the lower critical region

Require $P(X \leqslant c_1)$ to be as close as possible to 2.5%

From the tables

$P(X \leqslant 3) = 0.0212$ and $P(X \leqslant 4) = 0.0550$

0.0212 is closer to 0.025, so $c_1 = 3$ and the lower critical region is $X \leqslant 3$

Let $X = c_2$ is the lower boundary of the upper critical region, r

Require $P(X \geqslant c_2)$ to be as close as possible to 2.5%

From the tables

$P(X \geqslant 15) = 1 - P(X \leqslant 14) = 1 - 0.9585 = 0.0415$

and $P(X \geqslant 16) = 1 - P(X \leqslant 15) = 1 - 0.9780 = 0.0220$

0.0220 is closer to 0.025, so $c_2 = 16$ and the upper critical region is $X \geqslant 16$

Critical region is $X \leqslant 3$ or $X \geqslant 16$

c Actual significance level $= P(X \leqslant 3) + P(X \geqslant 18) = 0.0212 + 0.0220 = 0.0432$ or 4.32%

d Let the random variable Y be the number of incoming calls received in a 10-minute interval, so $Y \sim \text{Po}(10 \times 0.45)$, i.e. $Y \sim \text{Po}(4.5)$

$H_0 : \lambda = 4.5 \quad H_1 : \lambda < 4.5$

Assume H_0, so that $Y \sim \text{Po}(4.5)$

Significance level 5%, so require $P(Y \leqslant c) < 0.05$

From the tables

$P(Y \leqslant 1) = 0.0611$

As $0.061 > 0.05$ there is insufficient evidence to reject H_0

So reject the hypothesis that there are fewer incoming calls during the school holidays.

20 a Let the random variable X denote the number of sales that the agent makes in 300 calls, then $X \sim \text{B}(300, 0.025)$. As n is large and p is small, use $X \sim \text{Po}(300 \times 0.025)$, i.e. $X \sim \text{Po}(7.5)$

$H_0 : \lambda = 7.5 \quad H_1 : \lambda > 7.5$

Assume H_0, so that $X \sim \text{Po}(7.5)$

Significance level 5%

From the tables

$P(X \geqslant 11) = 1 - P(X \leqslant 10) = 1 - 0.8622 = 0.1378$

As $0.1378 > 0.05$ there is insufficient evidence to reject H_0

b Would reject H_0 at a 15% significance level as 15 is the smallest multiple of 5 larger than 13.8.

21 a Let the random variable X be the number of orders received in a 60-minute period, so $X \sim \text{Po}(6 \times 4)$, i.e. $X \sim \text{Po}(24)$

$H_0 : \lambda = 24 \quad H_1 : \lambda \neq 24$

Assume H_0, so that $X \sim \text{Po}(24)$

Significance level 10%

Let $X = c_1$ be the upper boundary of the lower critical region

Require $P(X \leqslant c_1)$ to be as close as possible to 5%

Using a calculator

$P(X \leqslant 15) = 0.0344$ and $P(X \leqslant 16) = 0.0563$

0.0563 is closer to 0.05, so $c_1 = 16$ and the lower critical region is $X \leqslant 16$

Let $X = c_2$ is the lower boundary of the upper critical region

Require $P(X \geqslant c_2)$ to be as close as possible to 5%

From the tables

$P(X \geqslant 32) = 1 - P(X \leqslant 31) = 1 - 0.9322 = 0.0678$

and $P(X \geqslant 33) = 1 - P(X \leqslant 32) = 1 - 0.9533 = 0.0467$

0.0467 is closer to 0.05, so $c_2 = 33$ and the upper critical region is $X \geqslant 33$

Critical region is $X \leqslant 16$ or $X \geqslant 33$

b Actual significance level $= P(X \leqslant 16) + P(X \geqslant 33) = 0.0563 + 0.0467 = 0.1030$ or 10.3%

c 18 is not in the critical region of the test since it is larger than 16 and smaller than 33. Therefore, at the 10% significance level there is no evidence that the average order rate is different to that claimed by the staff

22 a Let the random variable X denote the number of attempts needed before finding a winning ticket. Model this using a geometric distribution with parameter $p = 0.2$, $X \sim \text{Geo}(0.2)$

b $H_0 : p = 0.2 \quad H_1 : p < 0.2$

Assume H_0, so that $X \sim \text{Geo}(0.2)$

Significance level 5%

$P(X \geqslant 15) = (1 - 0.2)^{14} = 0.0440$ (4 d.p.)

$0.0440 < 0.05$

There is sufficient evidence at the 5% significance level to reject H_0, and conclude that Mr Taylor's suspicion is correct.

23 Let the random variable X denote the number of attempts Xander takes before making a basket, so $X \sim \text{Geo}(0.5)$

$H_0 : p = 0.5 \quad H_1 : p < 0.5$

Assume H_0, so that $X \sim \text{Geo}(0.5)$

Significance level 10%

$P(X \geqslant 5) = (1 - 0.5)^4 = 0.0625$ (4 d.p.)

$0.0625 < 0.1$

There is sufficient evidence at the 10% significance level to reject H_0, and conclude that Xander is overstating his ability.

24 a Let the random number X be the number of microchips tested before a faulty one is found

$\mathrm{H}_0 : p = 0.0005 \quad \mathrm{H}_1 : p > 0.0005$

Assume H_0, so that $X \sim \mathrm{Geo}(0.0005)$

Significance level 5%

Require $\mathrm{P}(X \leqslant c) < 0.05$

$\mathrm{P}(X \leqslant c) = 1 - (1 - 0.0005)^c$

So $1 - (1 - 0.0005)^c < 0.05$

$(0.9995)^c > 0.95$

$c < \dfrac{\log 0.95}{\log 0.9995}$

$c < 102.56$

So the critical region is $X \leqslant 102$

b 115 is outside the critical region $X \leqslant 102$ hence at the 5% significance level there is no evidence that the failure rate is greater than claimed.

25 a Let the random variable X be the height of a three-year-old child and let $\bar{X}$ be the sample mean found from a sample of 100 three-year-old children.

As the sample is large, using the central limit theorem $\bar{X} \approx\sim \mathrm{N}\left(90, \dfrac{5^2}{100}\right)$, i.e. $\bar{X} \approx\sim \mathrm{N}(90, 0.25)$

b Using the normal distribution function on a calculator

$\mathrm{P}(\bar{X} \geqslant 91) = 1 - \mathrm{P}(\bar{X} \leqslant 91) = 1 - 0.9772 = 0.0228$ (4 d.p.) $= 0.0228$

26 Using the central limit theorem $\bar{X} \sim \mathrm{N}\left(10, \dfrac{3^2}{5}\right)$, i.e. $\bar{X} \sim \mathrm{N}(10, 1.8)$

Using the normal distribution function on a calculator

$\mathrm{P}(\bar{X} < 10) = 0.5$ and $\mathrm{P}(\bar{X} \leqslant 7) = 0.0127$

So $\mathrm{P}(7 < \bar{X} < 10) = 0.5 - 0.0127 = 0.4873$ (4 d.p.)

27 a Probabilities sum to 1, so:

$0.4 + 2k + 0.3 + k = 1 \Rightarrow 3k = 0.3 \Rightarrow k = 0.1$

b To use the central limit theorem, find the mean and variance of X.

$\mathrm{E}(X) = 0.4 + (2 \times 0.2) + (3 \times 0.3) + (4 \times 0.1) = 2.1$

$\mathrm{E}\left(X^2\right) = 0.4 + (4 \times 0.2) + (9 \times 0.3) + (16 \times 0.1) = 5.5$

$\mathrm{Var}(X) = \mathrm{E}\left(X^2\right) - (\mathrm{E}(X))^2 = 5.5 - 4.41 = 1.09$

So by the central limit theorem, $\bar{X} \approx\sim \mathrm{N}\left(2.1, \dfrac{1.09}{200}\right)$, i.e. $\bar{X} \approx\sim \mathrm{N}(2.1, 0.00545)$

Using the normal distribution function on a calculator

$\mathrm{P}(\bar{X} > 2.09) = 1 - \mathrm{P}(\bar{X} \leqslant 2.09) = 1 - 0.4461 = 0.5539$ (4 d.p.)

c This estimate is accurate since the sample size, $n = 200$, is large.

28 a Let the random variable X be the number of calls the centre receives every minute, then $X \sim \text{Po}(15)$

Using a calculator

$\text{P}(X < 10) = \text{P}(X \leqslant 9) = 0.0699$ (4 d.p.)

b Let the random variable Y be the number of calls the centre receives in a 30-minute period, then $Y \sim \text{Po}(450)$

Using a calculator

$\text{P}(Y \leqslant 420) = 0.0810$ (4 d.p.)

c Let the random variable X be the number of calls the centre receives every minute, then $X \sim \text{Po}(15)$. As this is a Poisson distribution the mean and the variance of X is λ, i.e. 15

The number of calls made in 30-minute window is $30\bar{X}$, where $\bar{X}$ is the sample mean of 30 consecutive 1-minute samples.

Model this using the central limit theorem, $\bar{X} \approx\sim \text{N}\left(15, \frac{15}{30}\right)$, i.e. $\bar{X} \approx\sim \text{N}(15, 0.5)$

Require $\text{P}(30\bar{X} \leqslant 420) = \text{P}(\bar{X} \leqslant 14)$

Using a calculator $\text{P}(\bar{X} \leqslant 14) = 0.0786$ (4 d.p.)

The answer is close to the value found in part **b**, so the central limit theorem provides a good approximation in this case.

29 Let the random variable X be the number of attempts a student makes before selecting a green ball then $X \sim \text{Geo}(0.25)$

$\text{E}(X) = \frac{1}{p} = \frac{1}{0.25} = 4 \quad \text{Var}(X) = \frac{1-p}{p^2} = \frac{0.75}{0.0625} = 12$

Let $\bar{X}$ be the sample mean of the number attempts required by all 20 students

Then by the central limit theorem $\bar{X} \approx\sim \text{N}\left(4, \frac{12}{20}\right)$, i.e. $\bar{X} \approx\sim \text{N}(4, 0.6)$

Using the normal distribution function on a calculator

$\text{P}(\bar{X} > 4.5) = 1 - \text{P}(\bar{X} \leqslant 4.5) = 1 - 0.7407 = 0.2593$ (4 d.p.) = 0.0228

30 a Let the random variable X be the number of questions attempted before the student gets a 4th answer correct then, $X \sim$ Negative $\text{B}(4, 0.2)$

$\text{P}(X = 12) = \binom{11}{3} \times (0.2)^4 (0.8)^8 = 0.0443$ (4 d.p.)

b $\text{E}(X) = \frac{r}{p} = \frac{4}{0.2} = 20$

30 c By the central limit theorem, for a sample of 15 students $\bar{X} \approx\sim \mathrm{N}\left(\mu, \frac{\sigma^2}{15}\right)$

$$\mathrm{Var}(X) = \sigma^2 = \frac{r(1-p)}{p^2} = \frac{4\times 0.8}{0.04} = 80$$

So $\bar{X} \approx\sim \mathrm{N}\left(20, \frac{80}{15}\right)$

Using the normal distribution function on a calculator

$\mathrm{P}(\bar{X} < 19) = 0.3325$ (4 d.p.)

Challenge

1 a Let the random variable X represent the difference between the highest and lowest rolls of the three dice, so X ranges between 0 and 3. There are 64 ($= 4^3$) combinations.

If X is 0 then all 3 dice agree which happens with probability $\frac{4}{4^3} = \frac{1}{16}$

If X is 1 then the possible allowed rolls are (1,1,2), (1,2,2), (2,2,3), (2,3,3), (3,3,4), (4,4,3) and for each triple there are three ways to roll these three numbers since there are three ways to choose the dice with the unique number hence there are 18 dice rolls that lead to X being 1, hence

$$\mathrm{P}(X = 1) = \frac{18}{64} = \frac{9}{32}$$

If X is 2 then the possible allowed rolls are (1,1,3), (1,2,3), (1,3,3), (2,2,4), (2,3,4), (2,4,4) and for rolls where each dice rolls a different number there are 6 ways of permuting the dice to give the same result and as before where a number is repeated once there are 3 ways to permute the dice, hence the total number of rolls that give $X = 2$ is 3 + 6 + 3 + 3+ 6+ 3 = 24, hence

$$\mathrm{P}(X = 2) = \frac{24}{64} = \frac{3}{8}$$

Finally since the probabilities sum to 1: $\mathrm{P}(X = 3) = 1 - \frac{4}{64} - \frac{18}{64} - \frac{24}{64} = 1 - \frac{46}{64} = \frac{18}{64} = \frac{9}{32}$

So the probability distribution for X is:

x	0	1	2	3
$\mathrm{P}(X = x)$	$\frac{1}{16}$	$\frac{9}{32}$	$\frac{3}{8}$	$\frac{9}{32}$

b $\mathrm{E}(X) = \frac{1}{32}(9 + 2\times 12 + 3\times 9) = \frac{60}{32} = \frac{15}{8}$

Challenge

2 a If $X \sim B(4, p)$, then $P(X = x) = \binom{4}{x} \times p^x (1-p)^{4-x}$. Substituting $q = 1 - p$, this gives:

$P(X = 0) = q^4$

$P(X = 1) = 4pq^3$

$P(X = 2) = \frac{4!}{2! \times 2!} p^2 q^2 = 6p^2 q^2$

$P(X = 3) = \frac{4!}{3!} pq^3 = 4pq^3$

$P(X = 4) = p^4$

b Using the probability distribution with $q = 1 - p$, gives

$$\begin{aligned} E(X) &= 4pq^3 + 12p^2q^2 + 12p^3q + 4p^4 \\ &= 4p(1-p)^3 + 12p^2(1-p)^2 + 12p^3(1-p) + 4p^4 \\ &= 4p(1-3p+3p^2-p^3) + 12p^2(1-2p+p^2) + 12p^3(1-p) + 4p^4 \\ &= 4p \end{aligned}$$

$$\begin{aligned} E(X^2) &= 4pq^3 + 24p^2q^2 + 36p^3q + 16p^4 \\ &= 4p(1-p)^3 + 24p^2(1-p)^2 + 36p^3(1-p) + 16p^4 \\ &= 4p(1-3p+3p^2-p^3) + 24p^2(1-2p+p^2) + 36p^3(1-p) + 16p^4 \\ &= 4p + 12p^2 \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= 4p + 12p^2 - (4p)^2 \\ &= 4p + 12p^2 - 16p^2 = 4p - 4p^2 \\ &= 4p(1-p) \end{aligned}$$

3 a Let the random variable X be the number of quadrants searched before the metal detectorist finds the fourth item of value so $X \sim$ Negative B$(4, p)$

$H_0 : \lambda = 0.12 \quad H_1 : \lambda > 0.12$

Assume H_0, so that $X \sim$ Negative B$(4, 0.12)$

Significance level 2.5%

Using a calculator

$P(X \leqslant 10) = 0.0239$ and $P(X \leqslant 11) = 0.0341$

So the critical value is 10 and the critical region is $X \leqslant 10$

b Actual significance level = 0.0239 or 2.39%