Exam-style practice A Level

1 H_0: Geo (0.4) is a good model to represent the data.
H_1: Geo (0.4) is not a good model to represent the data.

The total observed frequencies = 52 + 31+ 12 + 7 +1 +1 = 104

Calculate the expected frequencies as follows:

$E_1 = 104 \times P(X=1) = 104 \times 0.4 \times 0.6^0 = 41.6$

$E_2 = 104 \times P(X=2) = 104 \times 0.4 \times 0.6^1 = 24.96$

$E_3 = 104 \times P(X=3) = 104 \times 0.4 \times 0.6^2 = 14.976$

$E_4 = 104 \times P(X=4) = 104 \times 0.4 \times 0.6^3 = 8.9856$

$E_5 = 104 \times P(X=5) = 104 \times 0.4 \times 0.6^4 = 5.3914$

$E_{i\geq 6} = 200 - (E_1 + E_2 + E_3 + E_4 + E_5) = 104 - 95.913 = 8.087$

The observed and expected results are shown in the table. Note that all expected frequency values are greater than 5.

x	1	2	3	4	5	⩾6	Total
Observed (O_i)	52	31	12	7	1	1	104
Expected (E_i)	41.6	24.96	14.976	8.9856	5.3914	8.087	104
$\dfrac{(O_i - E_i)^2}{E_i}$	2.6	1.462	0.519	0.439	3.577	6.211	14.879

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 14.879$$

There are 6 cells and 1 restriction (the total expected frequency must be 104) and so the number of degrees of freedom = 5

From the tables, the critical value $\chi_5^2(5\%) = 11.070$

As 14.8793 > 11.070, reject H_0. At the 5 % significance level it appears that Geo(0.4) is not a good model for the data.

2 **a** Using the fact that $G_X(1) = 1$

$k(1+2+3)^3 = 1$

So $6^3 k = 1 \Rightarrow k = \dfrac{1}{216}$

b Find the coefficient of the t^2 term

$G_X(t) = \dfrac{1}{216}\left(1+2t+3t^2\right)^3 = \dfrac{1}{216}\left(1+2t+3t^2\right) \times \left(1+2t+3t^2\right) \times \left(1+2t+3t^2\right)$

Collecting up the t^2 terms: $\dfrac{1}{216}\left(3t^2 + 4t^2 + 3t^2 + 4t^2 + 4t^2 + 3t^2\right) = \dfrac{1}{216} \times 21t^2 = \dfrac{7}{72}t^2$

So $P(X=2) = \dfrac{7}{72}$

2 c $G_X(t)=\frac{1}{216}\left(1+2t+3t^2\right)^3$

$$G'_X(t)=\frac{3}{216}\left(1+2t+3t^2\right)^2\times(2+6t)$$
$$=\frac{(2+6t)}{72}\left(1+2t+3t^2\right)^2$$
$$=\frac{(1+3t)}{36}\left(1+2t+3t^2\right)^2$$
$$G'_X(1)=\frac{(1+3)}{36}(1+2+3)^2=\frac{4\times6^2}{36}=4$$
$$G''_X(t)=\frac{(1+3t)}{36}\times2\times\left(1+2t+3t^2\right)(2+6t)+\frac{3}{36}\left(1+2t+3t^2\right)^2$$
$$G''_X(1)=\frac{(1+3)}{36}\times2\times(1+2+3)(2+6)+\frac{3}{36}(1+2+3)^2$$
$$=\frac{(4\times2\times6\times8)+(3\times36)}{36}$$
$$=\frac{384+108}{36}=\frac{492}{36}=\frac{41}{3}$$
$$\text{Var}(X)=G''_X(1)+G'_X(1)-(G'_X(1))^2$$
$$=\frac{41}{3}+4-4^2=\frac{41+12-48}{3}=\frac{5}{3}$$

d Using $G_Y(t)=t^bG_X(t^a)$ and $Y=2X+3$ $\quad a=2 \quad b=3$

$$G_Y(t)=t^3G_X\left(t^2\right)$$
$$=\frac{t^3}{216}(1+2\left(t^2\right)+3\left(t^2\right)^2)^3$$
$$=\frac{t^3}{216}\left(1+2t^2+3t^4\right)^3$$

3 a A suitable model for this situation would be a negative binomial distribution because Jagdeep wants to hit the bullseye r times. If the random variable Y is the number of darts thrown then $Y\sim$ Negative $B(r,p)$

Two assumptions that must be made for this distribution to be valid are that the probability of hitting the bullseye stays constant and that the throws are independent.

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

$$\text{Var}(Y)=\sigma^2=\frac{r(1-p)}{p^2}=37\frac{1}{7}=\frac{260}{7}$$

Substituting for r gives: $\frac{20p(1-p)}{p^2}=\frac{260}{7}$

Which simplifies to: $140(1-p)=260p\Rightarrow140-140p=260p\Rightarrow140=400p$

Hence $p=0.35$

3 c $r = 20p = 20 \times 0.35 = 7$

4 a Mean and variance of a Poisson distribution $= \lambda = 2.1$ in this case

Using the central limit theorem, $\bar{X} \sim N\left(2.1, \frac{2.1}{200}\right)$

$$\mathrm{P}(\bar{X} > 2.3) = 1 - (\bar{X} < 2.3)$$
$$= 1 - 0.9745 = 0.0255 \text{ (4 d.p.)} \qquad \text{(by using a calculator)}$$

b On average there are 2.1 flaws in a ten-metre length of cloth. So in a twenty-metre length of cloth there will be 4.2 flaws. Let the random variable Y be the number of flaws in a twenty-metre length of cloth, then $Y \sim \mathrm{Po}(4.2)$

H_0: $\lambda = 4.2$ $\quad$ H_1: $\lambda < 4.2$

Assume H_0, so that $X \sim \mathrm{Po}(4.2)$

Significance level 5%

By calculation

$$\mathrm{P}(X \leqslant 1) = \mathrm{P}(Y = 1) + \mathrm{P}(Y = 1) = \frac{e^{-4.2}4.2^0}{1!} + \frac{e^{-4.2}4.2^1}{1!}$$
$$= 0.0150 + 0.0630 = 0.0780$$

As $0.0780 > 0.05$, there is insufficient evidence at the 5% significance level to reject H_0. There is not enough evidence to say that the number of flaws has been reduced.

5 a Sum of probabilities = 1, i.e. $\sum \mathrm{P}(X = x) = 1$

$\Rightarrow p + 3q + 2r = 1 \qquad$ **(1)**

$$X = \frac{Y-5}{2} \Rightarrow \mathrm{E}(X) = E\left(\frac{Y-5}{2}\right) = 0.5\mathrm{E}(Y) - 2.5$$

So $\mathrm{E}(X) = 0.5 \times 4.9 - 2.5 = -0.05$

$\mathrm{E}(X) = \sum x\mathrm{P}(X = x) = -3p - 2q + 0 \times r + 2q + 3q + 5r$

$\Rightarrow -3p + 3q + 5r = -0.05 \qquad$ **(2)**

$\mathrm{P}(Y < 9) = 0.55 \Rightarrow \mathrm{P}(2X + 5 < 9) = 0.55$

So $\mathrm{P}(2X < 4) = 0.55$, $\mathrm{P}(X < 2) = 0.55$

$\Rightarrow p + q + r = 0.55 \qquad$ **(3)**

Equation **(1)** – Equation **(2)** gives $\qquad 4p - 3r = 1.05 \qquad$ **(4)**

3 × Equation **(3)** – Equation **(2)** gives $\qquad 6p - 2r = 1.7 \qquad$ **(5)**

2 × Equation **(5)** – 3 × Equation **(4)** gives $\qquad 5r = 3.4 - 3.15 = 0.25$, so $r = 0.05$

Substituting for r in Equation **(5)** gives $\qquad 6p = 1.8$, so $p = 0.3$

Substituting for p and r in Equation **(3)** gives $\qquad q = 0.55 - 0.35 = 0.2$

5 b $P(X > 2Y - 3) = P(X > 2(2X + 5) - 3)$

$= P(X > 4X + 10 - 3)$

$= P(X > 4X + 7)$

$= P(3X < -7)$

$= P\left(X < -\frac{7}{3}\right)$

$= P(X \leqslant -3)$

So $P(X > 2Y - 3) = P(X \leqslant 3) = p = 0.3$

6 a Let the random variable X represent the number of penalties the footballer misses each day. So $X \sim B(500, p)$

$P(\text{scores}) = (1 - p)$

$P(\text{scores four consecutive penalties}) = (1 - p)^4 = 0.8853$

$\Rightarrow (1 - p) = \sqrt[4]{0.8853}$

$\Rightarrow p = 1 - \sqrt[4]{0.8853} = 1 - 0.970\,002 = 0.0300$ (4 d.p.)

$\text{Mean} = np = 500 \times 0.03 = 15$

$\text{Variance} = np(1 - p) = 500 \times 0.03 \times 0.97 = 14.55$

b n (= 500) is large, p (= 0.03) is small, and the mean (15) is approximately equal to the variance (14.55). Therefore, a Poisson distribution is an appropriate approximation for the binomial distribution.

c Using the Poisson approximation $X \sim \text{Po}(500 \times 0.03)$, i.e. $X \sim \text{Po}(15)$

$P(X > 17) = P(X \geqslant 18) = 1 - P(X \leqslant 17)$

$= 1 - 0.74886$ (using a calculator)

$= 0.2511$ (4 d.p.)

d Using the binomial distribution $X \sim B(500, 0.03)$

$P(X > 17) = P(X \geqslant 18) = 1 - P(X \leqslant 17)$

$= 1 - 0.75146$ (using a calculator)

$= 0.2485$ (4 d.p.)

The two results are the same to 2 significant figures
0.2511 = 0.25 (2 s.f.)
0.2485 = 0.25 (2 s.f.)

So the Poisson distribution is a good approximation for the binomial distribution in this case.

7 a $X \sim B\left(25, \frac{1}{6}\right)$

Number of sixes	Binomial probability	Cumulative binomial distribution
0	$P(X=0)=\binom{25}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{25}=0.0105$	$P(X=0)=0.0105$
1	$P(X=1)=\binom{25}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{24}=0.0524$	$P(X\leqslant 1)=0.0629$
2	$P(X=2)=\binom{25}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{23}=0.1258$	$P(X\leqslant 2)=0.1887$

$H_0: p=\frac{1}{6}$, $H_1: p<\frac{1}{6}$

Significance level 10%, so require c such that $P(X\leqslant c)<0.1$

From the table above $P(X\leqslant 1)<0.1$ and $P(X\leqslant 2)>0.15$

So the critical value is 1, hence the critical region is $X\leqslant 1$

$\text{Size} = \text{P(Type I error)} = P\left(X\leqslant 3 \mid p=\frac{1}{6}\right)=0.0629$

b $\text{Power function} = \text{P(reject } H_0 \text{ when it is false)}$

$= P(X\leqslant 1 \mid X\sim B(25,p))$

$= P(X=0 \mid X\sim B(25,p)) + P(X=1 \mid X\sim B(25,p))$

$=(1-p)^{25}+25p(1-p)^{24}=(1-p)^{24}((1-p)+25p)$

$=(1-p)^{24}(1+24p)$

c $Y\sim \text{Geo}\left(\frac{1}{6}\right)$

$\text{Size} = \text{P(Type I error)} = P\left(Y\geqslant 13 \mid p=\frac{1}{6}\right)==\left(1-\frac{1}{6}\right)^{13-1}=\left(\frac{5}{6}\right)^{12}=0.1122 \text{ (4 d.p.)}$

$\text{Power function} = P\left(Y\geqslant 13 \mid Y\sim \text{Geo}(p)\right)=(1-p)^{13-1}=(1-p)^{12}$

d If $p=0.09$, $p<\frac{1}{6}$ (0.1667) so the dice is biased against rolling a six: the null hypothesis should therefore be rejected.

Using the power functions from parts **b** and **c**, when $p = 0.09$:

the power of Philip's test = $(1-0.09)^{24}(1+24\times 0.9)=0.3286\ 0.3289$

the power of Gemma's test = $(1-0.09)^{12}=0.3225$

The probability of rejecting H_0 when it is false (the power) is greater in Philip's test (0.3286) than in Gemma's test (0.3225) so Philip's test is more likely to arrive at the correct conclusion. The probability of rejecting H_0 when it is true (the size) is smaller in Philip's test (0.0629) than in Gemma's test (0.1122), so Philip's test is less likely to arrive at the incorrect conclusion.