Central limit theorem 5A

1 a The population is normal so $X \sim \text{N}(10, 2^2)$

Variance of the sample $= \frac{\sigma^2}{n} = \frac{2^2}{6} = \frac{2}{3}$, so $\bar{X} \sim \text{N}\left(10, \frac{2}{3}\right)$

Using a calculator:

$\text{P}(\bar{X} > 12) = 1 - \text{P}(\bar{X} < 12) = 1 - 0.9928 = 0.0072$ (4 d.p.)

b As the population is normally distributed, the sample mean is also normally distributed and the answer to part a is not an approximation.

2 a Let X be the amount of drink in a carton.

Variance of the sample $= \frac{\sigma^2}{n} = \frac{1.5^2}{4} = 0.5625$, so $\bar{X} \sim \text{N}(40, 0.5625)$

Using a calculator:

$\text{P}(\bar{X} > 40.5) = 1 - \text{P}(\bar{X} < 40.5) = 1 - 0.7475 = 0.2525$ (4 d.p.)

b Variance of the sample $= \frac{\sigma^2}{n} = \frac{1.5^2}{49} = 0.04918\ldots$, so $\bar{X} \sim \text{N}(40, 0.045918)$

$\text{P}(\bar{X} > 40.5) = 1 - \text{P}(\bar{X} < 40.5) = 1 - 0.9902 = 0.0098$ (4 d.p.)

3 a Let L be the length of a bolt. The distribution of L is unknown.

Variance of the sample $= \frac{\sigma^2}{n} = \frac{0.2^2}{100} = 0.0004$

So using the central limit theorem $\bar{L} \approx\sim \text{N}(3.03, 0.0004)$

Using a calculator:

$\text{P}(\bar{L} > 3) = 0.0668$ (4 d.p.)

b Let $\bar{M}$ be the mean length of a bolt from a sample of n bolts. Then

$\bar{M} \sim \text{N}\left(3.03, \frac{0.2^2}{n}\right)$ and require $\text{P}(\bar{M} < 3) < 0.01$

Standardise the sample mean

$\text{P}(\bar{M} < 3) = \text{P}\left(Z < \frac{3 - 3.03}{\frac{0.2}{\sqrt{n}}}\right)$ and require $\text{P}\left(Z < \frac{3 - 3.03}{\frac{0.2}{\sqrt{n}}}\right) < 0.01$

Using the table for the percentage points of the normal distribution (see Appendix, page 190):

$\text{P}(Z < -2.3263) = 0.01$

So $\frac{3 - 3.03}{\frac{0.2}{\sqrt{n}}} < -2.3263$

$\Rightarrow \frac{0.03\sqrt{n}}{0.2} > 2.3263$ (dividing by -1 so reversing the inequality)

$\Rightarrow \sqrt{n} > 15.5086\ldots$

$\Rightarrow n > 240.51\ldots$

So $n = 241$ is the minimum sample size required for $\text{P}(\bar{M} < 3) < 0.01$

4 $\mu = \text{E}(X) = \frac{1}{5}(1+2+3+4+5) = 3$

$\sigma^2 = \text{Var}(X) = \frac{1}{5}(1+2^2+3^2+4^2+5^2) - \mu^2 = \frac{55}{5} - 9 = 2$

Using the central limit theorem $X \approx\sim \text{N}\left(3, \frac{2}{40}\right)$

$\text{P}(\bar{X} > 3.2) = 1 - \text{P}(\bar{X} < 3.2) \approx 1 - 0.8145 = 0.1855$ (4 d.p.)

5 a Let the random variable S = score on the dice

$\text{E}(S) = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6} = 3.5$

$\text{Var}(S) = \frac{1}{6}(1+2^2+3^2+4^2+5^2+6^2) - 3.5^2 = \frac{91}{6} - \frac{49}{4} = \frac{182}{12} - \frac{147}{12} = \frac{35}{12}$

So by the central limit theorem, $\bar{S} \approx\sim \text{N}\left(3.5, \frac{\frac{35}{12}}{35}\right)$, i.e. $\bar{S} \approx\sim \text{N}\left(3.5, \frac{1}{12}\right)$

$\text{P}(\bar{S} > 4) = 1 - \text{P}(\bar{S} > 4) = 1 - 0.9584 = 0.0416$ (4 d.p.)

b Let the random variable T = total score of 35 rolls of the dice, so $T = 35\bar{S}$

$\text{P}(T < 100) = \text{P}\left(S < \frac{100}{35}\right)$

$\text{P}\left(S < \frac{100}{35}\right) = 0.0130$ (4 d.p.)

6 Let the random variable S = the number of sixes recorded in 30 rolls of the dice, so $S \sim \text{B}\left(30, \frac{1}{6}\right)$

Using the formula for the mean and variance of a binomial distribution

$\text{E}(S) = np = 30 \times \frac{1}{6} = 5$

$\text{Var}(S) = np(1-p) = 30 \times \frac{1}{6} \times \frac{5}{6} = \frac{25}{6}$

So by the central limit theorem $\bar{S} \approx\sim \text{N}\left(5, \frac{\frac{25}{6}}{25}\right)$, i.e. $\bar{S} \approx\sim \text{N}\left(5, \frac{1}{6}\right)$

And by calculator $\text{P}(\bar{S} < 4.5) = 0.1103$ (4 d.p.)

7 a Probabilities sum to 1 so

$0.1 + 3k + k + 0.3 = 1 \Rightarrow 4k = 0.6 \Rightarrow k = 0.15$

b $\text{E}(X) = 2 \times (3 \times 0.15) + 3 \times 0.15 + 5 \times 0.3 = 2.85$

$\text{Var}(X) = (4 \times 0.45 + 9 \times 0.15 + 25 \times 0.3) - 2.85^2$

$= 10.65 - 8.1225 = 2.5275$

So by the central limit theorem $\bar{X} \approx\sim \text{N}(2.85, 0.025275)$

And by calculator $\text{P}(\bar{X} > 3) = 1 - \text{P}(\bar{X} < 3) \approx 1 - 0.8273 = 0.1727$ (4 d.p.)

c Answer is an approximation, but as n (= 100) is large it will be fairly accurate.

8 Let the random variable S = score on the dice. Then $E(S) = \mu = 3.5$ and $Var(S) = \sigma^2 = \frac{35}{12}$

So by the central limit theorem $\bar{S} \approx\sim N\left(3.5, \frac{35}{12n}\right)$

Require $P(\bar{S} < 3.4) + P(\bar{S} > 3.6) < 0.01$ and as the sample mean is normally distributed and symmetrical about the mean, this is equivalent to $P(\bar{S} > 3.6) < 0.005$

Standardise the sample mean

$P(\bar{S} > 3.6) = P\left(Z > \frac{0.1}{\sqrt{\frac{35}{12n}}}\right)$ and require $P\left(Z > \frac{0.1}{\sqrt{\frac{35}{12n}}}\right) < 0.005$

Using the table for the percentage points of the normal distribution:

$P(Z > 2.5758) = 0.005$

So $\frac{0.1}{\sqrt{\frac{35}{12n}}} > 2.5758$

$\Rightarrow \frac{\sqrt{12n}}{10\sqrt{35}} > 2.5758$

$\Rightarrow \frac{12n}{3500} > 6.63474...$

$\Rightarrow n > 1935.13\ldots$

So n = 1936 is the minimum sample size required for $P(\bar{S} < 3.4) + P(\bar{S} > 3.6) < 0.01$, i.e. for there being a less than 1% chance that the mean of all scores differs from 3.5 by more than 0.1

9 **a** The salaries in a company are unlikely to be symmetrically distributed so a normal distribution would not be a good model.

b Let the random variable X = the salary of an employee. Then using the central limit theorem

$\bar{X} \approx\sim N\left(28500, \frac{6800^2}{15}\right)$

i $P(\bar{X} < 25\,000) = 0.0231$ (4 d.p.)

ii $P(25\,000 < \bar{X} < 30\,000) = P(\bar{X} < 30\,000) - P(\bar{X} < 25\,000)$
$= 0.80354 - 0.02311 = 0.7804$ (4 d.p.)

c The estimates are likely to be inaccurate given the distribution of employee salaries are unlikely to be normal and given the relatively small sample size.

10 Let the random variable T be the length of time taken to repair this particular fault. Assume the time taken to repair the fault is normally distributed. Then

$T \sim \mathrm{N}(\mu, 2.5^2)$ and $\bar{T} \sim \mathrm{N}\left(\mu, \frac{2.5^2}{n}\right)$ and require $\mathrm{P}(\bar{T} > \mu + 0.5) = 0.025$

Standardise the sample mean

$\mathrm{P}(\bar{T} > \mu + 0.5) = \mathrm{P}\left(Z > \frac{\mu + 0.5 - \mu}{\frac{2.5}{\sqrt{n}}}\right)$ and so require $\mathrm{P}\left(Z > \frac{\sqrt{n}}{5}\right) < 0.025$

Using the table for the percentage points of the normal distribution: $\mathrm{P}(Z > 1.9600) = 0.025$

So $\frac{\sqrt{n}}{5} > 1.9600$

$\Rightarrow \sqrt{n} > 9.8$

$\Rightarrow n > 96.04$

So $n = 97$ is the minimum sample size required.