Central Limit Theorem (The distribution of means of large samples)

Starter

1. (Review of previous material) The heights of a new variety of sunflower are normally distributed with mean 2 cm and standard deviation 40 cm. 100 samples of 50 flowers each are measured. In how many samples would you expect the sample mean to be:
   (a) greater than 210 cm
   (b) between 195 cm and 205 cm

Working: (a) Let $X \sim \text{N}(2, 0.4^2)$

The size of ***each sample*** is 50 so $n = 50$ $\Rightarrow$ $\bar{X}_{50} \sim \text{N}\left(2, \dfrac{0.2^2}{50}\right)$

$P(\bar{X} > 210) \approx 0.0385$
Expected number of samples is $0.0385 \times 100 = 3.85$.

(b) $P(\bar{X} > 210) \approx 0.623$
Expected number of samples is $0.623 \times 100 = 62.3$.

E.g. 1 A random sample of 30 observation of the random variable X were taken. Given that X has mean 40 and standard deviation 7, calculate the probability that the sample mean is less than 39.
N.B. No information is given about the original distribution.

Working: Using the CLT, $\bar{X}_{30} \sim \text{N}\left(40, \dfrac{7^2}{30}\right)$.
$P(\bar{X} < 39) \approx 0.217$
The probability that the mean of the sample is less than 39 is 0.217 (3 s.f.)

E.g. 2 A random sample of size 35 is taken from each of these distributions. Calculate the probability that the sample mean is greater than 9.
(a) $X \sim \text{Po}(8.5)$ (b) $X \sim \text{B}(12, 0.8)$ (c) $X \sim \text{R}(8, 10.2)$

Working: (a) $X \sim \text{Po}(8.5)$ $\Rightarrow$
$E(X) = 8.5$ and $\text{Var}(X) = 8.5$
Using the CLT, $\bar{X}_{35} \sim \text{N}\left(8.5, \dfrac{8.5}{35}\right)$
$P(\bar{X} > 9) = 0.155$ (3 s.f.)

(b) $X \sim \text{B}(12, 0.8)$ $\Rightarrow$
$E(X) = np = 12 \times 0.8 = 9.6$
$\text{Var}(X) = np(1-p) = 12 \times 0.8 \times 0.2 = 1.92$
Using the CLT, $\bar{X}_{35} \sim \text{N}\left(9.6, \dfrac{1.92}{35}\right)$
$P(\bar{X} > 9) = 0.995$ (3 s.f.)

(c) $X \sim \text{R}(8,\ 12)$

$$E(X) = \frac{1}{2}(a+b) = \frac{1}{2}(8+10.2) = 9.1$$

$$\text{Var}(X) = \frac{1}{12}(b-a)^2 = \frac{1}{12}(10.2-8)^2 = \frac{121}{300}$$

Using the CLT, $\bar{X}_{35} \sim \text{N}\left(9.1,\ \dfrac{121}{300 \times 35}\right)$

$P(\bar{X} > 9) = 0.824$ (3 s.f.)

E.g. 3 The random variable X is such that $\mu = E(X)$ and $\sigma^2 = \text{Var}(X)$. Decide whether these statements about the distribution of $\bar{X}$ are true or false. If false, give a reason for your answer.

(a) The Central Limit Theorem states that $\bar{X}$ is normally distributed for any distribution of X.

(b) $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$ are true for any distribution of X and any value of n.

(c) The Central Limit Theorem states that the mean of the sample is normally distributed for large values of n.

(d) If $X \sim \text{N}(\mu,\ \sigma^2)$, then $\bar{X}_n \sim \text{N}\left(\mu,\ \dfrac{\sigma^2}{n}\right)$ only for large value of n.

Working:

(a) False; $\bar{X}$ is ***approximately*** a normal distribution and only when n is large $(n > 25)$.

(b) True.

(c) False; $\bar{X}$ is ***approximately*** a normal distribution

(d) False, the statement is true for any size of n.

E.g. 4 Decide whether the Central Limit Theorem would be used in these situations. If not, give a reason for your answer.

(a) Given that $X \sim \text{Po}(32)$, find $P(X \le 34)$.

(b) Given that $X \sim \text{B}(30,\ 0.5)$, $P(\bar{X} \ge 16)$

(c) Given that $X \sim \text{R}(1,\ 9)$, find $P(\bar{X}_{25} > 5.2)$.

(d) Given that $X \sim \text{N}(45,\ 2^2)$, find $P(\bar{X}_{16} > 45.8)$

Working:

(a) No, the Central Limit Theorem is only used with the mean of samples.

(b) Possibly but no information is given for the size of the sample — if $n > 25$, the Central Limit Theorem can be used.

(c) No, the size of the sample, 25, is too small.

(d) No, the Central Limit Theorem is not needed when the original distribution is normally distributed

E.g. 5 If a large number of samples of size n are taken from $X \sim$ R$(2, 30)$ and approximately $80\,\%$ of the sample means are less than 17.15, estimate n.

Working: $X \sim$ R$(2, 30)$

$$E(X) = \frac{1}{2}(a+b) = \frac{1}{2}(2+30) = 16$$

$$\text{Var}(X) = \frac{1}{12}(b-a)^2 = \frac{1}{12}(30-2)^2 = \frac{196}{3}$$

Using the CLT, $\bar{X}_n \sim \text{N}\left(16, \frac{196}{3n}\right)$

$P(\bar{X}_n < 17.15) = 0.8$: $\quad P(Z < z) = 0.8 \quad \Rightarrow \quad z = 0.84162$

$z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$: $\quad 0.84162 = \dfrac{17.15-16}{\frac{\sqrt{196}}{\sqrt{3n}}}$

$$\sqrt{n} = \frac{\sqrt{196} \times 0.84162}{\sqrt{3}(17.15-16)} \approx 5.915$$

$$n = 35$$

E.g. 6 A large number of samples of size n is taken from $X \sim$ Po(5) and approximately $11.5\,\%$ of the sample means are more than 5.3. Estimate the value of n.

Working: $X \sim$ Po(5) $\quad \Rightarrow \quad E(X) = 5$ and Var$(X) = 5$ so $\sigma = \sqrt{5}$

Using the CLT, $\bar{X}_n \sim \text{N}\left(5, \frac{5}{n}\right)$

$P(\bar{X}_n > 5.3) = 0.115$: $\quad P(Z > z) = 0.115$

So $P(Z < z) = 0.885 \quad \Rightarrow \quad z = 1.2004$

$z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$: $\quad 1.2004 = \dfrac{5.3-5}{\frac{\sqrt{5}}{\sqrt{n}}}$

$$\sqrt{n} = \frac{\sqrt{5} \times 1.2004}{5.3-5} \approx 8.947$$

$$n = 80$$

Video: Introducing the Central Limit Theorem
Video: Central Limit Theorem
Video: Central Limit Theorem example

Solutions to Starter and E.g.s

Exercise
p163 8E Qu 1i, 2i, 3-6, (7-8 red)