Central Limit Theorem Cheat Sheet

Edexcel FS1

In Chapter 3 of Statistics and Mechanics Year 2, you learnt that for a random sample of size n taken from a random variable X which is normally distributed with mean μ and variance σ^2, the sample mean $\bar{X}$ is approximately normally distributed, with mean μ and variance $\frac{\sigma^2}{n}$. It turns out that this is in fact true regardless of the distribution that X follows. This result is known as the Central Limit Theorem (C.L.T.).

The central limit theorem states that:

This notation tells you that the sample mean is approximately normally distributed with mean μ and variance σ^2.

- If $X_1, X_2, \dots, X_n$ is a random sample of size n from a population with mean μ and variance σ^2, then $\bar{X} \approx \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

This is a very powerful result because regardless of the underlying distribution, the sample mean is always normally distributed. This makes any further questions much easier to answer because we only need to analyse a normal distribution, rather than the underlying distribution which could potentially be much more complex than a normal one.

- It is important to remember that the sample mean is **approximately** normally distributed. As the sample size n increases, the approximation becomes better.
- You do not need to apply the continuity correction when using the central limit theorem because while the underlying distribution might be discrete, the sample mean itself is not discrete and can theoretically take non-integer values (e.g. 2.5).

You need to be able to apply the central limit theorem to problems where a random variable is modelled by any of the distributions you have met so far. We will go through an example for each distribution.

C.L.T. with a discrete distribution (that does not fit any specific distribution you have learnt about)

Example 1: The random variable X has the probability distribution shown in the table.
(a) Find the value of k.

x	0	2	3	5
$P(X=x)$	0.1	$3k$	k	0.3

A random sample of 100 observations of X is taken.
(b) Use the central limit theorem to estimate the probability that the mean of these observations is greater than 3.
(c) Comment on the accuracy of your estimate.

To find k, we use the fact that the total probability must equal 1.	$\sum P(X=x) = 0.1 + 3k + k + 0.3 = 1$ $\Rightarrow 4k = 0.6$ $\Rightarrow k = 0.15$
Before we can write the distribution of X, we need to find the mean and variance of the distribution at hand.	$E(X) = \sum xP(X=x)$ $= 0(0.1) + 2(3(0.15)) + 3(0.15) + 5(0.3)$ $= 2.85$
To find the variance we need to use $Var(X) = E(X^2) - E(X)^2$	$E(X^2) = 0^2(0.1) + 2^2(3(0.15)) + 3^2(0.15) + 5^2(0.3) = 10.65$ $\therefore Var(X) = 10.65 - 2.85^2 = 2.5275$
The sample mean will be normally distributed with mean $E(X)$ and variance $\frac{Var(X)}{n}$.	$\mu = E(X) = 2.85, \quad \frac{\sigma^2}{n} = \frac{2.5275}{100} = 0.0253$ $\therefore \bar{X} \approx \sim N(2.85, 0.0253)$
Finding the required probability using the sample mean distribution:	$P(required) = P(\bar{X} > 3) = 0.173$ (by calculator)
Using the fact that the approximation becomes better as the sample size increases.	Our estimate is fairly accurate since the sample size (100) is large.

C.L.T. with a Poisson distribution

Example 2: A supermarket manager is trying to model the number of customers that visit her store each day. She observes that, on average, 20 new customers enter the store every minute.
(a) Calculate the probability that fewer than 15 customers arrive in a given minute.
(b) Find the probability that in one hour no more than 1150 arrive.
(c) Use the central limit theorem to estimate the probability that in one hour no more than 1150 customers arrive. Compare your answer to part **b**.

(a) Defining the distribution for the number of customers visiting in a given minute:	Let X denote the number of customers arriving every minute, then $X \sim Po(20)$
Finding the required probability using a calculator:	$P(required) = P(X \leq 14) = 0.1049\ (4\ d.p, by\ calculator)$.
(b) Defining the distribution for the number of customers visiting in a given hour:	Let Y denote the number of customers arriving every minute, then $Y \sim Po(20 \times 60) \Rightarrow Y \sim Po(1200)$
Finding the required probability using a calculator:	$P(required) = P(Y \leq 1150) = 0.0756$ (4 d.p.)
(c) We need to use the C.L.T; this means using the distribution of the sample mean. We can take a sample of 60 observations from X, so instead of finding $P(Y \leq 1150)$, we can find $P\left(\bar{X} \leq \frac{1150}{60}\right)$. This is equivalent but makes use of the central limit theorem.	This is the same question as part b, but we need to use the central limit theorem to estimate this probability. To do this, we take a sample of 60 observations of X and calculate the probability that the sample mean is less than or equal to $\frac{1150}{60} = 19.1666\dots$
Using $\bar{X} \approx \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.	For a sample of 60 observations, $\bar{X} \approx \sim N\left(20, \frac{20}{60}\right) \rightarrow \bar{X} \approx \sim N\left(20, \frac{1}{3}\right)$
Finding the required probability using a calculator:	$P(required) = P(\bar{X} \leq 19.1666\dots) = 0.0745\ (4\ d.p.)$
Since the answers to parts b and c are very close we can conclude the approximation is good.	Our answer is quite close to part b, so it is fair to say that our approximation using the central limit theorem is very good.

C.L.T. with a geometric distribution

Example 3: A married couple plan to have children and are desperate to have a daughter. They decide they will keep having children until they have a daughter and then stop. You can assume that giving birth to a boy or girl is equally likely, and independent of the gnnder of any other children the couple have had.
(a) Find the probability that they will have more than 2 children.

Suppose a group of 10 couples all decide on the same plan.
(b) Estimate the probability that between them, the 10 couples have more than 24 children.

(a) The distribution we are dealing with here is geometric, since we are carrying out a number of trials (each with a constant probability of success) until one success.	Let X denote the number of children birthed until they have a daughter, then $X \sim Geo\left(\frac{1}{2}\right)$,
If $X \sim Geo(p)\ then\ P(X \geq x) = (1-p)^{x-1}$	$P(required) = P(X \geq 3) = (1-0.5)^{3-1} = 0.25$
(b) We are asked to estimate the probability, so we need to use the central limit theorem. Finding the mean and variance for X using the mean/variance results for a geometric distribution:	$Sample\ size\ n = 10$ $E(X) = \frac{1}{p} = \frac{1}{0.5} = 2, \quad Var(X) = \frac{1-p}{p^2} = \frac{1-0.5}{0.5^2} = 2$
Now we can define the distribution of the sample mean:	$\therefore \bar{X} \approx \sim N\left(2, \frac{2}{10}\right) \rightarrow \bar{X} \approx \sim N\left(2, \frac{1}{5}\right)$
The probability that the 10 couples have more than 24 children is equivalent to the probability that the sample mean is greater than $\frac{24}{10} = 2.4$.	$P(required) = P\left(\bar{X} > \frac{24}{10}\right) = P(\bar{X} > 2.4) = 1 - P(\bar{X} < 2.4)$ $= 1 - 0.8145 = 0.1855\ (4\ d.p.)$ by calculator.

C.L.T. with a binomial distribution

Example 5: In a group of 20 students, each rolls a fair six-sided dice 10 times and records the number of sixes. Estimate the probability that the average number of sixes rolled by each student is greater than 2.

We first consider each student separately. They each roll a fair six-sided dice 10 times. The number of sixes can therefore be modelled by a binomial distribution since the probability of a six is constant $\left(\frac{1}{6}\right)$ and we have a fixed number of trials.	Let X represent the number of sixes achieved by one student, then $X \sim B\left(10, \frac{1}{6}\right)$
Now we need to use the central limit theorem. We are told that there is a group of 20 students, so the sample size is 20. Finding the mean and variance using the mean/variance results for a binomial distribution:	$Sample\ size\ n = 10$ $E(X) = np = 10\left(\frac{1}{6}\right) = \frac{10}{6} = \frac{5}{3}$ $Var(X) = np(1-p) = \frac{10}{6}\left(1 - \frac{1}{6}\right) = \frac{50}{36} = \frac{25}{18}$
Now we can define the distribution of the sample mean:	$\therefore \bar{X} \approx \sim N\left(\frac{5}{3}, \frac{25}{18(20)}\right) \rightarrow \bar{X} \approx \sim N\left(\frac{5}{3}, \frac{5}{72}\right)$
We need to find the probability that the sample mean is greater than 2:	$P(required) = P(\bar{X} > 2) = 1 - P(\bar{X} < 2) = 1 - 0.8970 = 0.1030$

C.L.T. with a negative binomial distribution

Example 4: David is selling raffle tickets from door to door to raise money for charity. To reach his daily fundraising goal, he needs to sell 10 tickets. He observes that, on average, an occupant in one in every three houses he visits will buy a ticket.
(a) Find the probability that on a given day he reaches his daily goal after visiting exactly 35 houses.
(b) In one month, David worked on 20 days, and met his daily goal on each day. Estimate the probability that the average number of houses he visited per day was 35 or fewer.

(a) The distribution we are dealing with here is a negative binomial one, since we are finding the probability that he achieves 10 successes in 35 trials. (i.e. it takes him 35 houses to sell 10 tickets)	Let X denote the number of houses visited until 10 tickets are sold, then $X \sim NB\left(10, \frac{1}{3}\right)$.
If $X \sim NB,\ then\ P(X=x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$.	$P(required) = P(X = 35) = \binom{34}{9}\left(\frac{1}{3}\right)^{10}\left(\frac{2}{3}\right)^{25} = 0.0352\ (4\ d.p.)$
(b) This is where we need to use the central limit theorem. David worked on 20 days so the sample size is 20. Finding the mean and variance using the mean/variance results for a negative binomial distribution:	$Sample\ size\ n = 10$ $E(X) = \frac{r}{p} = \frac{10}{\frac{1}{3}} = 30, \quad Var(X) = \frac{r(1-p)}{p^2} = \frac{10\left(1-\frac{1}{3}\right)}{\frac{1}{3}^2} = 60$
Now we can define the distribution of the sample mean:	$\therefore \bar{X} \approx \sim N\left(30, \frac{60}{20}\right) \rightarrow \bar{X} \approx \sim N(30, 3)$
We need to find the probability that the sample mean is less than or equal to 35:	$P(required) = P(\bar{X} \leq 35) = 0.9981\ (4\ d.p.)$ by calculator.

C.L.T. with a normal distribution

Example 6: An automatic coffee machine uses milk powder. The mass, S grams, of milk powder used in one cup of coffee is modelled by $S \sim N(4.9, 0.8^2)$. 'Semi skimmed' milk powder is sold in $500g$ packs. Find the probability that one pack will be sufficient for 100 cups of coffee.

We need to take a sample of 100 observations from S and find the probability that the mean of this sample is less than or equal to 5. We are given the mean and variance of our distribution so we can use the central limit theorem right away:	Let $S = S_1 + S_2 + S_3 + \dots + S_{100}$. Then $\bar{S} \approx \sim N\left(4.9, \frac{0.8^2}{100}\right) \rightarrow \bar{S} \approx \sim N(4.9, 0.0064)$
Now we just need to find the probability the sample mean is less than or equal to 5:	$P(required) = P(\bar{S} \leq 5) = 0.8944\ (4\ d.p.)$