The Central Limit Theorem

Suppose that a sample of size n is selected from a population that has mean μ and standard deviation σ. Let $X_1, X_2, \cdots, X_n$ be the n observations that are independent and identically distributed (i.i.d.). Define now the sample mean and the total of these n observations as follows:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$T = \sum_{i=1}^{n} X_i$$

The *central limit theorem* states that the sample mean $\bar{X}$ follows approximately the normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$, where μ and σ are the mean and standard deviation of the population from where the sample was selected. The sample size n has to be large (usually $n \geq 30$) if the population from where the sample is taken is nonnormal. If the population follows the normal distribution then the sample size n can be either small or large.

To summarize: $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$.

To transform $\bar{X}$ into z we use: $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Example: Let X be a random variable with $\mu = 10$ and $\sigma = 4$. A sample of size 100 is taken from this population. Find the probability that the sample mean of these 100 observations is less than 9. We write $P(\bar{X} < 9) = P(z < \frac{9-10}{\frac{4}{\sqrt{100}}}) = P(z < -2.5) = 0.0062$ (from the standard normal probabilities table).

Similarly the central limit theorem states that sum T follows approximately the normal distribution, $T \sim N(n\mu, \sqrt{n}\sigma)$, where μ and σ are the mean and standard deviation of the population from where the sample was selected.

To transform T into z we use: $z = \frac{T-n\mu}{\sqrt{n}\sigma}$

Example: Let X be a random variable with $\mu = 10$ and $\sigma = 4$. A sample of size 100 is taken from this population. Find the probability that the sum of these 100 observations is less than 900. We write $P(T < 900) = P(z < \frac{900-100(10)}{\sqrt{100}(4)}) = P(z < -2.5) = 0.0062$ (from the standard normal probabilities table).

Below you can find some applications of the central limit theorem.

EXAMPLE 1
A large freight elevator can transport a maximum of 9800 pounds. Suppose a load of cargo containing 49 boxes must be transported via the elevator. Experience has shown that the weight of boxes of this type of cargo follows a distribution with mean $\mu = 205$ pounds and standard deviation $\sigma = 15$ pounds. Based on this information, what is the probability that all 49 boxes can be safely loaded onto the freight elevator and transported?

EXAMPLE 2
From past experience, it is known that the number of tickets purchased by a student standing in line at the ticket window for the football match of $UCLA$ against USC follows a distribution that has mean $\mu = 2.4$ and standard deviation $\sigma = 2.0$. Suppose that few hours before the start of one of these matches there are 100 eager students standing in line to purchase tickets. If only 250 tickets remain, what is the probability that all 100 students will be able to purchase the tickets they desire?

EXAMPLE 3
Suppose that you have a sample of 100 values from a population with mean $\mu = 500$ and with standard deviation $\sigma = 80$.

a. What is the probability that the sample mean will be in the interval $(490, 510)$?

b. Give an interval that covers the middle 95% of the distribution of the sample mean.

EXAMPLE 4
The amount of regular unleaded gasoline purchased every week at a gas station near $UCLA$ follows the normal distribution with mean 50000 gallons and standard deviation 10000 gallons. The starting supply of gasoline is 74000 gallons, and there is a scheduled weekly delivery of 47000 gallons.

a. Find the probability that, after 11 weeks, the supply of gasoline will be below 20000 gallons.

b. How much should the weekly delivery be so that after 11 weeks the probability that the supply is below 20000 gallons is only 0.5%?

Solutions:
EXAMPLE 1
We are given $n = 49, \mu = 205, \sigma = 15$. The elevator can transport up to 9800 pounds. Therefore these 49 boxes will be safely transported if they weigh in total less than 9800 pounds. The probability that the total weight of these 49 boxes is less than 9800 pounds is $P(T < 9800) = P(z < \frac{9800-49(205)}{\sqrt{49}15}) = P(z < -2.33) = 1 - 0.9901 = 0.0099$.

EXAMPLE 2
We are given that $\mu = 2.4, \sigma = 2, n = 100$. There are 250 tickets available, so the 100 students will be able to purchase the tickets they want if all together ask for less than 250 tickets. The probability for that is $P(T < 250) = P(z < \frac{250-100(2.4)}{\sqrt{100}2}) = P(z < 0.5) = 0.6915$.

EXAMPLE 3
We are given $\mu = 500, \sigma = 80, n = 100$.

a. $P(490 < \bar{x} < 510) = P(\frac{490-500}{\frac{80}{\sqrt{100}}} < z < \frac{490-500}{\frac{80}{\sqrt{100}}}) = P(-1.25 < z < 1.25) = 0.8944 - (1 - 0.8944) = 0.7888$.

b. $\pm 1.96 = \frac{\bar{x}-500}{\frac{80}{\sqrt{100}}} \Rightarrow \bar{x} = 484.32, \bar{x} = 515.68$. Therefore $P(484.32 < \bar{x} < 515.68) = 0.95$.

EXAMPLE 4
We are given that $\mu = 50000, \sigma = 10000, n = 11$. The starting supply is 74000 gallons and the weekly delivery is 47000 gallons. Therefore the total supply for the 11-week period is $74000 + 11 \times 47000 = 591000$ gallons.

a. The supply will be below 20000 gallons if the total gasoline purchased in these 11 weeks is more than $591000 - 20000 = 571000$ gallons. Therefore we need to find $P(T > 571000) = P(z > \frac{571000-11(50000)}{\sqrt{11}10000}) = P(z > 0.63) = 1 - 0.7357 = 0.2643$.

b. Let A be the unknown schedule delivery. Now the total gasoline purchased must be more than $74000 + 11 \times A - 20000$. We want this with probability 0.5%, or $P(T > 74000 + 11A - 20000) = 0.005$. The z value that corresponds to this probability is 2.575. So, $2.575 = \frac{74000+11A-20000-11(50000)}{\sqrt{11}10000} \Rightarrow A = 52854.8$. The weekly delivery must be 52854.8 gallons.