CLT — II

In the previous session we illustrated the famous Central Limit theorem. It says that you start with any collection of iid random variables the probability distribution of corresponding sum of variables or sample mean ultimately becomes normal.

In fact CLT also allows us to approximate various probabilities of many events such $P(a < S_n \leq b)$ or $P(a < \bar{X}_n \leq b)$ for various values of a and b such that $a < b$.

Note: Suppose we want to approximate the probability $P(a < S_n \leq b)$ using the CLT. In this case we We proceed as follows. In this case we use the prob. distⁿ of S_n which is $N(n\mu, \sigma^2 n)$ under the CLT. So let us approximate the required probability.

$$P(a < S_n \le b) = P\left(\frac{a-n\mu}{\sqrt{\sigma^2 n}} < \frac{S_n - n\mu}{\sqrt{\sigma^2 n}} \le \frac{b-n\mu}{\sqrt{\sigma^2 n}}\right)$$

$$= P\left(\frac{a-n\mu}{\sqrt{\sigma^2 n}} < Z \le \frac{b-n\mu}{\sqrt{\sigma^2 n}}\right) \quad \{Z \sim N(0,1)\}$$

$$= \Phi\left(\frac{b-n\mu}{\sqrt{\sigma^2 n}}\right) - \Phi\left(\frac{a-n\mu}{\sqrt{\sigma^2 n}}\right)$$

$$\therefore \boxed{P(a < S_n \le b) = \Phi\left[\frac{b-n\mu}{\sigma\sqrt{n}}\right] - \Phi\left[\frac{a-n\mu}{\sigma\sqrt{n}}\right]} \quad \text{---} (*)$$

Prob in (*) can be compute easily using normal table for given values a, b, μ and σ, n.

Similarly you can approximate $P(a < \bar{X}_n \le b)$ using pdf of $\bar{X}_n$ as follows:

$$P(a < \bar{X}_n \le b) = \left(\frac{a-\mu}{\sigma/\sqrt{n}} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \le \frac{b-\mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{a-\mu}{\sigma/\sqrt{n}} < Z \le \frac{b-\mu}{\sigma/\sqrt{n}}\right)$$

$$= \Phi\left(\frac{\sqrt{n}(b-\mu)}{\sigma}\right) - \Phi\left(\frac{\sqrt{n}(a-\mu)}{\sigma}\right).$$

Ex: A random variable of size 100 is taken from a population whose mean $\mu = 60$ and variance $\sigma^2 = 400$. Use the central limit theorem to find with what probability can we assert that the mean of the sample will not differ from the mean μ by more than 4.

Sol^n: Here our variable of interest is $\bar{X}_n$, the sample mean. Under the CLT $\bar{X}_n \sim N(\mu, \sigma^2/n)$.
Here $\mu = 60$, $\sigma^2 = 400$, $n = 100$. We want to compute $P(|\bar{X} - \mu| \leq 4) = P(|\bar{X} - 60| \leq 4)$.

Now $P(|\bar{x} - 60| \leq 4) = P(-4 \leq \bar{x} - 60 \leq 4)$

$$= P\left(\frac{-4}{20/\sqrt{100}} \leq \frac{\bar{x} - 60}{20/\sqrt{100}} \leq \frac{4}{20/\sqrt{100}}\right)$$

$$= P(-2 \leq Z \leq 2) = 2\Phi(2) - 1$$

$$= 2(0.5 + 0.4772) - 1$$

$$= 0.9544.$$

<u>Ex</u>: A distribution with uknown mean μ has variance σ^2 equal to 1.5. Use the central limit theorem to find <u>how large a sample</u> should be taken to form the distribution in order that the probability will be at least 0.95 that the <u>sample mean</u> will be within 0.5 of the population mean μ.

<u>Solution</u>: Here situation like that $X_1, X_2, \cdots, X_n$ are iid with $E(X_i) = \mu$ and $V(X_i) = \sigma^2 = 1.5$.

from the CLT $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$.

Our problem is to determine n such that

$$P(|\bar{X} - \mu| \le 0.5) \geq 0.95$$

$$\Rightarrow P\left(\frac{|X - \mu|}{\sigma/\sqrt{n}} \le \frac{0.5}{\sigma/\sqrt{n}}\right) \geq 0.95$$

$$= P\left(|Z| \le \frac{0.5}{\sigma/\sqrt{n}}\right) \geq 0.95$$

[By $N(0,1)$ property $P(|Z| \le 1.96) = 0.95$]

Hence we must have

$$\frac{0.5}{\sigma/\sqrt{n}} = 1.96 \Rightarrow \sqrt{n} = \frac{1.96 \times \sigma}{0.5}$$

$$\Rightarrow n = \frac{(1.96)^2 \times 1.5}{0.25} \simeq 24$$

$\therefore$ required sample size is atleast 24.

Ex: Let X_i $(i=1,2,\cdots,75)$ be iid Poisson random variables with parameter $\lambda = 2$ $(X_i \sim P(2))$

Use the CLT to estimate the probability $P(120 \leq S_n \leq 160)$ where $S_n = X_1 + X_2 + \cdots + X_{75}$.

Sol^n $\Rightarrow$ Since $X_i \overset{iid}{\sim} P(2)$, $i = 1,2,\cdots,75$, and so we have $E(X_i) = 2$, $V(X_i) = 2$.

$$\therefore E(S_n) = E(X_1 + \cdots + X_n) = n\lambda \; ; \; \begin{matrix} n = 75 \\ \lambda = 2 \end{matrix}$$

$$= 150$$

$$V(S_n) = 150$$

from the CLT, for large n, $S_n \sim N(n\lambda, n\lambda)$

or equivalently $S_n \sim N(150, 150)$

Now $P(120 \leq S_n \leq 160) = P\left(\frac{120-150}{\sqrt{150}} < \frac{S_n - 150}{\sqrt{150}} \leq \frac{160-150}{\sqrt{150}}\right)$

$$P(120 \le S_n \le 160) = P(-2.45 < Z \le 0.82)$$

$$= \Phi(0.82) - \Phi(-2.45)$$

$$= \Phi(0.82) - (1 - \Phi(2.45))$$

$$= \Phi(0.82) + \Phi(2.45) - 1$$

$$= (0.5 + 0.2939) + (0.5 + 0.4929) - 1$$

$$= 0.7868$$

Note: Can you observe computing $P(120 \le S_n \le 160)$ using actual prob-distⁿ of S_n which is $P(n\lambda)$, is really difficult as can be seen below:

$$P(120 \le S_n \le 160) = \sum_{s=120}^{160} p_{S_n}(s) \quad \text{---} \ (1)$$

where $p_{S_n}(s) = \dfrac{e^{-150}(150)^s}{s!}$, $s = 0, 1, 2, \ldots$

If you are able to simplify Equation (1) then hopefully its value will be close to 0.7868.

Ex: A Study involving stress is done on a college Campus among the students. The stress scores follow a uniform distribution with the lowest stress Score equal to 1 and the highest equal to 5. Using a sample of 75 students, find

(i) The probability that average score for the 75 students is less than 2.

(ii) The 90th percentile for the average stress score for the 75 students.

(iii) The prob. that the total of the 75 stress score is less than 200.

(iv) The 90th percentile for the total stress score for the 75 students.

Ex: From a large collection of bolts which is known to contain 3% defective bolts, 1000 are chosen at random. Let X be the no of defective bolts among those chosen. Use the CLT to find an approximate value of the prob. that X does not exceed 5% of 1000.