Central Limit Theorem

If $x_1, x_2 \ldots x_n$ is a sequence of iid rv with finite mean μ and finite variance σ^2

then distribution of $\dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ approaches the standard normal distribution, $N(0,1)$ as $n \to \infty$.

we can also write $\dfrac{\sum x_i - n\mu}{\sqrt{n\sigma^2}}$ approx $N(0,1)$ for large n.

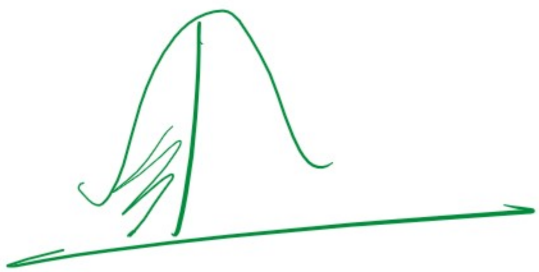
Ex 1: It is assumed that the no. of flights arriving at an airport per working day has a mean of 40 and a s.d. of 12. A survey was conducted over 50 working days. Calculate the probability that the sample mean number of flights arriving per working day was less than 35.

$\mu = 40 \quad \sigma = 12 \quad n = 50$

The central limit theorem states that $\bar{x} \sim N\left(40, \frac{12^2}{50}\right)$

$P(\bar{x} < 35) = P\left(z < \dfrac{35-40}{}\right)$

$$P(\bar{x} < 35) = P\left(z < \frac{35-40}{\sqrt{12^2/50}}\right)$$



$$= P\left[Z < \frac{\sqrt{50}\,(-5)}{12}\right]$$

$$= P\left[Z < -2.95\right]$$

$$= 1 - P\left[Z < 2.95\right]$$

$$= 1 - 0.99841$$

$$= 0.0016 \text{ (ans)}$$

(2) The cost of repairing a vehicle following an accident has mean 6,200 and std deviation 650. A study was carried out into 65 vehicles that had been involved in accidents. Calculate the probability that the total repair bill for the vehicles exceeded 400,000!

We have $\mu = 6200$ $\quad \sigma = 650$ $\quad n = 65$

Let $Z \sim N(0,1)$

$$S \sim N\left[65 \times 6200,\ 65 \times 650^2\right]$$

$$= N\left[403000,\ 5240^2\right]$$

$$P\left[S > 400000\right] = P\left[Z > \frac{400000 - 403000}{5240}\right]$$

$$= P\left[Z > -0.572\right]$$

$$= P\left[Z < 0.572\right]$$

$$= 0.71634$$

(ans).

Normal approximation to the Binomial Distribution

$Bin(n, p)$

Let X_i be iid Bernoulli r.v. $B(1, p)$

$$P(X_i = 1) = p$$
$$P(X_i = 0) = 1 - p$$

in Bernulli's Distribn $\mu = E(X_i) = p$

$$\sigma^2 = V(X_i) = p(1-p)$$

$$\Sigma X_i \sim Bin(n, p)$$

As a result of CLT it can be said that for large n. $\bar{X} \sim N(\mu, \sigma^2/n)$

or $\Sigma X_i \sim N(n\mu, n\sigma^2)$

$\Rightarrow \sum X_i \sim N(n\mu, n\sigma^2)$

$\sum X_i \sim N(np, np(1-p))$

(3) Given that $X \sim Bin(n, p)$ derive the mean and variance of $\bar{X}$ and hence write the distribution of $\bar{X}$.

$$\bar{X} = \frac{1}{n} \sum X_i$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} \sum E(X_i)$$

$$= \frac{1}{\cancel{n}} \times \cancel{n}p$$

$$= p$$

$$V(\bar{X}) = V\left(\frac{\sum X_i}{n}\right) = \frac{1}{n^2} V(\sum X_i)$$

$$= \frac{1}{n^{\cancel{2}}} \times \cancel{n}p(1-p)$$

$$= \frac{p(1-p)}{n}$$

$$\therefore \bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Poisson Distribution

$X_i \quad i = 1, \ldots, n \qquad$ iid $Poi(\lambda)$

so $\mu = E(X_i) = \lambda$

$\sigma^2 = V(X_i) = \lambda$

CLT: $\sum X \sim N(n\lambda, n\lambda)$ for large n.

Ex 4: Show that $\sum X_i \sim Poi(n\lambda)$ where X_i is $Poi(\lambda)$

$$X \sim P(\lambda)$$
$$Y \sim P(\lambda)$$
$$\Rightarrow X+Y \sim P(\lambda+\lambda)$$
$$\therefore \sum X_i \sim P(n\lambda)$$

Gamma Distribution:

Let X_i, $i = 1, 2, \ldots n$ be a sequence of iid exponential (λ) variables and let Y be their sum,

mean, $\mu = \frac{1}{\lambda}$ and variance $\sigma^2 = \frac{1}{\lambda^2}$

for large n, $Y = \sum X_i \sim N(n/\lambda, \frac{n}{\lambda^2})$

$\therefore$ Y which is Gamma(n, λ) will have

$\therefore$ Y which is Gamma (n, λ) will have a normal approximation for large values of n

Recall: $X_i \sim Exp(\lambda)$

then $\Sigma X_i \sim$ Gamma (n, λ)

Since χ^2_k = Gamma $\left(\frac{k}{2}, \frac{1}{2}\right)$

$\therefore$ χ^2_k will have normal approximation $N(k, 2k)$ for large values of its degree of free k

Ex5: Let X be a Poisson Variable with parameter 20. Use the normal approximation to obtain a value for $P(X \leq 15)$.

$X \sim P(20)$

$X \sim N(20, 20) \Rightarrow \frac{X-20}{\sqrt{20}} \sim N(0,1)$

CLT: $\sum X_i \sim N(\mu n, n\sigma^2)$

$\Leftarrow \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$\Rightarrow \frac{\bar{x}-\mu}{\sqrt{\sigma/n}} \sim N(0,1)$ as $n \to \infty$

$P(X \le 15) = P(X < 15.5)$

$= P\left(Z < \frac{15.5-20}{\sqrt{20}}\right)$

$= P(Z < -1.006)$

$= 1 - P(Z > 1.006)$

$= 1 - 0.84279$

$= 0.15721$

$Bin(n,p) \sim N(np, npq)$ $np > 5$, $nq > 5$

$P(\lambda) \sim N(\lambda, \lambda)$ λ large.

Gamma Poi $(\lambda) \sim N\left(\frac{\alpha}{\lambda}, \frac{\alpha}{\lambda^2}\right)$, α large

$\chi^2_{(k)} \sim N(k, 2k)$ k large.

The continuity correction:

Apply for discrete [Convert $\pm$ 0.5]

1. change in equality $\geq$ or $\leq$ between consecutive integers.
2. Upper bound $(+0.5)$
   lower bound (-0.5)

Ex: $S < 100 \rightarrow S \le 99 \rightarrow S \le 99 + 0.5$
$\rightarrow S \le 99.5$

$S \ge 70 \rightarrow S \ge 70 - 0.5 \rightarrow S \ge 69.5$

$50 < S \leq 110 \rightarrow 51 \leq S \leq 110 \rightarrow 50.5 \leq S \leq 110.5$

Ex 6: A company issued questionnaires to clients to obtain feedback on the clarity of their brochure. It is thought that 5% of clients do not find the brochure helpful. Calculate the approximate probability that in a sample of 1000 responses, the number N, of clients who do not find the brochure helpful satisfies $40 < N < 70$.

We have $N \sim Bin(1000, 0.05)$

Normal approximation $N \sim N(np, npq)$

$N \sim N(50, 47.5)$

$$P(40 < N < 70) \approx P(41 \leq N \leq 69)$$

$$P(40.5 < N < 69.5)$$

$$= P\left(Z < \frac{69.5 - 50}{\sqrt{47.5}}\right) - P\left(Z < \frac{40.5 - 50}{\sqrt{47.5}}\right)$$

$$= P(Z < 2.829) - [1 - P(Z > 1.378)]$$

$$\doteq 0.99766 - [1 - 0.9159]$$

$$= 0.39766 - [1 - 0.9159]$$

$$= 0.31356$$

TRY: In a certain large population 45% of people have blood group A. A random sample of 300 individuals is chosen from this population. Calculate an approximate value for the probability that more than 115 of the sample have blood group A.