Sri SAI RAM ENGINEERING COLLEGE
INSTITUTE OF TECHNOLOGY
West Tambaram, Chennai - 44

SAIRAM DIGITAL RESOURCES

YEAR II
SEM IV

MA8391

PROBABILITY AND STATISTICS

Department of Information Technology

UNIT II

TWO DIMENSIONAL RANDOM VARIABLES

2.6 CENTRAL LIMIT THEOREM

SCIENCE & HUMANITIES

CENTRAL LIMIT THEOREM

STATEMENT:

If $X_1, X_2, \ldots\ldots X_n$ is a sequence of n independent and identically distributed (i,i,d) random variables, each having mean μ and variance σ^2, and if $\overline{X} = \dfrac{X_1 + X_2 + \ldots\ldots + X_n}{n}$, then the variate $Z = \dfrac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ has a distribution that approaches the standard normal distribution as $n \to \infty$, provided the m.g.f. exist

Proof:

M.G.F of Z about the origin is

$$M_Z(t) = E\left(e^{tZ}\right)$$

$$= E\left[e^{t\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)}\right] = E\left[e^{\frac{t\sqrt{n}\,\overline{X}}{\sigma}} \cdot e^{-\frac{t\mu\sqrt{n}}{\sigma}}\right]$$

$$= e^{-\frac{\mu t\sqrt{n}}{\sigma}} . E\left[e^{\frac{t\sqrt{n}}{\sigma}\left[\frac{X_1+X_2+.....+X_n}{n}\right]}\right]$$

$$= e^{-\frac{t\mu\sqrt{n}}{\sigma}} . E\left[e^{\frac{tX_1}{\sigma\sqrt{n}}} . e^{\frac{tX_2}{\sigma\sqrt{n}}} .....e^{\frac{tX_n}{\sigma\sqrt{n}}}\right]$$

[Since $X_1, X_2, ...X_n$ are independent,

$E(X_1, X_2, ..., X_n) = E(X_1)E(X_2)....E(X_n)$]

Hence $$M_Z(t) = e^{-\frac{t\mu\sqrt{n}}{\sigma}} E\left(e^{\frac{tX_1}{\sigma\sqrt{n}}}\right) E\left(e^{\frac{tX_2}{\sigma\sqrt{n}}}\right) \dots E\left(e^{\frac{tX_n}{\sigma\sqrt{n}}}\right)$$

The variables $X_1, X_2, \dots X_n$ have the same M.G.F

$$\therefore \quad M_Z(t) = e^{-\frac{\mu t\sqrt{n}}{\sigma}} \left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

Where $M_X\left(\frac{t}{\sigma\sqrt{n}}\right)$ is the m.g.f of $X = X_i$, $i = 1, 2, \dots, n$.

Taking log on both sides.

$$\log M_Z(t) = \log\left[e^{-\frac{t\mu\sqrt{n}}{\sigma}}\right] + n\log\left[M_X\left(\frac{t}{\sigma\sqrt{n}}\right)\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\log\left[E\left(e^{\frac{tX}{\sigma\sqrt{n}}}\right)\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\log\left[E\left(1+\left(\frac{t}{\sigma\sqrt{n}}\right)X + \frac{1}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 X^2 + \dots\right)\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\log\left[1+\left(\frac{t}{\sigma\sqrt{n}}\right)\mu_1' + \frac{1}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 \mu_2' + \dots\right]$$

$$= -\frac{t\mu\sqrt{n}}{\sigma} + n\left[\left(\frac{t}{\sigma\sqrt{n}}\mu_1' + \frac{\mu_2'}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \dots\right) - \frac{1}{2}\left(\mu'\frac{t}{\sigma\sqrt{n}} + \dots\right)^2 + \dots\right]$$

Put $\mu' = \mu = mean$

$$\log M_Z(t) = -\frac{\mu\sqrt{n}}{\sigma} + \frac{\sqrt{n}\mu t}{\sigma} + \frac{t^2}{2\sigma^2}\left[\mu_2' - (\mu_1')^2\right]$$

+ terms containing n in the denominator

$$\log M_Z(t) = \frac{t^2}{2\sigma^2}\sigma^2 +$$ terms containing n in the denominator

$$\log M_Z(t) = \frac{t^2}{2}$$

i.e., $M_Z(t) = e^{\frac{t^2}{2}}$ as $n \to \infty$

The M.G.F of Z is the m.g.f of $N(0,1)$ i.e., as $n \to \infty$ the distribution of Z tends to the standard normal distribution.

Different form of central limit theorem.

i) Liapounoff's form of CLT

If $X_1, X_2, \ldots\ldots, X_n$ is sequence of independent random variables with $E[X_i] = \mu_i$ and $V[X_i] = \sigma_i^2$, $i = 1, 2, 3, \ldots$ and if $S_n = X_1 + X_2 + \ldots. + X_n$, then under certain general conditions, S_n follows a normal distribution with mean $\mu = \sum_{i=1}^{n} \mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$ as n tends to infinity.

Note:

The sampling distribution of the sample mean approaches a normal distribution irrespective of the

The sampling distribution of the sample mean approaches a normal distribution irrespective of the distribution of the population from where the sample is taken and approximation to the normal distribution becomes very close with increase in sample size.

i.e., $\overline{X} = \frac{1}{n}(X_1 + X_2 + \ldots.. + X_n)$, then $E(\overline{X}) = \frac{n\mu}{n} = \mu$ and $V(\overline{X}) = \frac{1}{n^2}(n\sigma^2) = \frac{n}{\sigma^2}$.

$\therefore$ $\overline{X}$ follows $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ as $n \to \infty$.

ii) Lindeberg – Levy C.L.T

If the random variables $X_1, X_2, \ldots, X_n$ have the same p.d.f with mean μ and variance σ^2 then the random variable $S_n = X_1 + X_2 + \ldots + X_n$ follows normal distribution with mean $n\mu$ and variance $n\sigma^2$ that is $S_n \sim N(n\mu\ ,\ n\sigma^2)$.

1. A distribution with unknown mean μ has variance equal to 1.5. Use central limit theorem to find how large a sample should be taken from the distribution in order that the probability will be at least 0.95 that the sample mean will be within 0.5 of the population mean.

distribution in order that the probability will be at least 0.95 that the sample mean will be within 0.5 of the population mean.

Solution:

Given $E[X_i] = \mu_i$ and $V[X_i] = 1.5$

Let $\overline{X}$ denote the sample mean. Then

$$\overline{X} : N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

i.e., $$\overline{X} : N\left(\mu, \frac{\sqrt{1.5}}{\sqrt{n}}\right)$$

We have to find n such that

$$P\left(-0.5 < \overline{X} - \mu < 0.5\right) \geq 0.95$$

$$P\left[-\frac{0.5}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{0.5}{\frac{\sigma}{\sqrt{n}}}\right] \geq 0.95$$

$$P\left[-\frac{0.5\sqrt{n}}{\sqrt{1.5}} < Z < \frac{0.5\sqrt{n}}{\sqrt{1.5}}\right] \geq 0.95$$

$$2 \times P\left[0 < Z < \frac{0.5\sqrt{n}}{\sqrt{1.5}}\right] \geq 0.95$$

$$2 \times P\left[0 < Z < 0.4082\sqrt{n}\right] \geq 0.95$$

$$\therefore \quad P\left[0 < Z < 0.4082\sqrt{n}\right] \geq 0.475$$

From the table of areas under normal curve

$P[0 < Z < 1.96] = 0.475$

$\therefore$ Least value of n is given by $0.4082\sqrt{n} = 1.96$

i.e., $n = 24$

$\therefore$ The size of the sample must be at least 24.

2. Let $X_1, X_2, \ldots, X_n$ be Poisson variates with parameter $\lambda = 2$. Let $S_n = X_1 + X_2 + \ldots + X_n$ where $n = 75$, Find $P[120 \le S_n \le 160]$.

Solution:

Given X_i, $i = 1, 2, 3, \ldots, 75$ are Poisson variates with $\lambda = 2$.

Now $E[X_i] = \lambda = 2$ and $V[X_i] = \lambda = 2$, $i = 1, 2, 3, \ldots, 75$

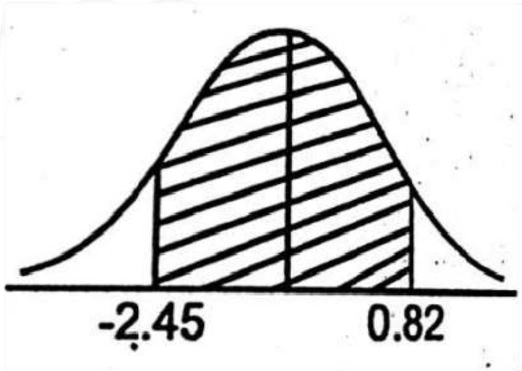
Let $S_n = X_1 + X_2 + \ldots + X_n$, $n = 75$

By central limit theorem S_n follows normal distribution with

mean $E[S_n] = n\mu = 2\times 75 = 150$ and variance $V[S_n] = n\sigma^2 = 150$.

To find $P[120 \le S_n \le 160]$

$$= P\left[\frac{120-150}{\sqrt{150}} \le \frac{S_n - 150}{\sqrt{150}} \le \frac{160-150}{\sqrt{150}}\right]$$



$$= P[-2.45 \le Z \le 0.82]$$

$$= P[-2.45 \le Z \le 0] + [0 \le Z \le 0.82]$$

$$= P[0 \le Z \le 2.45] + [0 \le Z \le 0.82]$$

$$= 0.4929 + 0.2939$$

$$= 0.7868 .$$

3. The life time of a particular variety of electric bulbs may be considered as a random variable with mean 1200 hours and standard deviation 250 hours. Using central limit theorem find the probability that the average life time of 60 bulbs exceeds 1250 hours.

Solution:

Let $X_1, X_2, \ldots\ldots, X_{60}$ be the life time of the bulbs

$\mu = E[X_i] = 1200$ hours $\qquad i = 1, 2, \ldots., 60$

$\sigma = S.D.(X_i) = 250$ hours

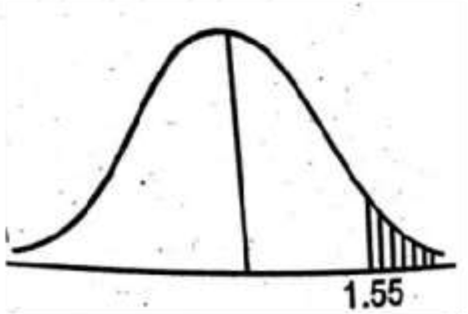
Let $\overline{X}$ denote the average life time of 60 bulbs.

By central limit theorem $n = 60$ and $\overline{X}$ follows $N\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right)$.

The standard normal variate is $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\overline{X} - 1200}{250 / \sqrt{60}}$

Now, $P\left[\overline{X} > 1250\right]$

$$= P\left[\frac{\overline{X} - 1200}{250 / \sqrt{60}} > \frac{1250 - 1200}{250 / \sqrt{60}}\right]$$



$$= P\left[Z > 1.55\right]$$

$= 0.5 - P[0 < Z < 1.55]$

$= 0.5 - 0.4394 = 0.0606$

4. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random variables with mean = 2 and variance $\sigma^2 = \frac{1}{4}$. Find $P(192 < S_n < 210)$ where $S_n = X_1 + X_2 + \ldots + X_{100}$.

Solution:

Given mean = 2, variance $\sigma^2 = \frac{1}{4}$ for each random variable X_i $n = 100$

S_n follows normal distribution with mean $= n\mu$ $= 100(2) = 200$ and $S.D = \sigma\sqrt{n} = \frac{1}{2} \times 10 = 5$.

Now, $P(192 < S_n < 210) = P\left(\frac{192 - 200}{5} < \frac{S_n - 200}{5} < \frac{210 - 200}{5}\right)$

$$= P(-1.6 < Z < 2)$$
$$= P(0 < Z < 1.6) + P(0 < Z < 2)$$

$$= 0.4452 + 0.4772 = 0.9224$$

5. A sample of size 100 is taken from a population whose mean is 60 and variance is 400. Using central limit theorem, with what probability can we assert that the mean of the sample will not differ from $\mu = 60$ by more than 4?

Solution:

Let $X_1 + X_2 + ... + X_{100}$ be the sample values from the given population.

Given $E[X_i]=60$, $Var[X_i]=400,\ i=1,2,3,\ldots,100$

Using central limit theorem, the sample mean $\overline{X}$ follows normal distribution with mean $\mu=60$ and variance $\frac{\sigma^2}{n}=\frac{400}{100}=4$. $\therefore \frac{\sigma}{\sqrt{n}}=2$

$$Z=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}=\frac{\overline{X}-60}{2}$$

$$P\left[\left|\overline{X}-\mu\right|\le 4\right]=P\left[\left|\overline{X}-60\right|\le 4\right]$$

$$=P\left[-4\le \overline{X}-60\le 4\right]$$

$$=P\left[56\le \overline{X}\le 64\right]$$

$$= P\left[\frac{56-\mu}{\sigma/\sqrt{n}} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le \frac{64-\mu}{\sigma/\sqrt{n}}\right]$$

$$= P\left[\frac{56-60}{2} \le \frac{\overline{X}-60}{2} \le \frac{64-60}{2}\right]$$

$$= P\left[-2 \le Z \le 2\right]$$

$$= 2P\left[0 \le Z \le 2\right]$$

$$= 2 \times 0.4772 = 0.9544$$