Module 3 [18MAB203T]

Limit Theorems

Laws of large numbers	**Central Limit Theorems**
concerned with stating conditions under which the average of a sequence of RV's converge to the expected average.	**concerned with determining conditions under which the sum of a large number of RV's has a probability distribution that is approximately normal.**

Markov's inequality: If X is a random variable that takes only non-negative values, then for any a>0, $P\{X \geq a\} \leq \dfrac{E(X)}{a}$.

Chebyshev's inequality: If X is a random variable with finite mean μ and variance σ^2, then for any value c>0, $P\{|X-\mu| \geq c\} \leq \dfrac{\sigma^2}{c^2}$.

Proof:

Since $(X-\mu)^2$ is a non-negative random variable we apply Markov's inequality by taking $a = c^2$.

$$P\{(X-\mu)^2 \geq c^2\} \leq \frac{E(X-\mu)^2}{c^2}$$

But $(X-\mu)^2 \geq c^2$ iff $|X-\mu| \geq c$

$$\Rightarrow P\{|X-\mu| \geq c\} \leq \frac{\sigma^2}{c^2}$$

Hence the proof.

Note:

1. Both the inequalities enable us to derive bounds on probabilities.
2. In Markov inequality mean of the probability distribution is known.
3. In Chebyshev's inequality both the mean and the variance of the probability distribution are known.

Proposition:

If Var(X) = 0, then P{X=E(X)} = 1. (i.e., the only RV's having variances equal to 0 are those that are constant with probability 1).

Proof:

By Tchebycheff inequality

$$P\{|X-\mu| \geq c\} \leq \frac{\sigma^2}{c^2}$$

And $P\{|X-\mu| \leq c\} \geq 1-\frac{\sigma^2}{c^2}$

Now $P\{|X-\mu| \geq c\} \leq 0 \ (\because \sigma^2 = 0) \,\&\, P\{|X-\mu| \leq c\} \geq 1$ ...(I)

(I) holds even for small values of c.

$\therefore$ when $c \to 0$

$$P\{|X-\mu| = 0\} = 1$$

(i.e.) $P\{X = \mu\} = 1$

(i.e.) $P\{X = E(X)\} = 1$

Hence the proof.

The weak law of large numbers statement:

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed RV's each having finite mean $E(X_i) = \mu$. Then for any $\varepsilon > 0, P\left\{\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0$ as $n \to \infty$.

Proof:

The result is proved by taking the additional assumption that the random variables have a finite variance σ^2.

$$E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{n\mu}{n} \text{ and } Var\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Using Tchebycheff's inequality

$$P\left\{\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}.$$

As $n \to \infty, P\left\{\left|\frac{X_1 + X_2 + ... + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0$

The Central Limit Theorem:

1. Liapounoff's form
2. Lindberg Levy's form and Corollary

Refer T. Veerarajan Book.

It states that the sum of large number of independent RV's has a distribution that is approximately normal.

Uses:

1) It provides a simple method for computing approximate probabilities for sums of independent RV's.
2) It explains the remarkable fact that the empirical frequencies of so many natural populations exhibit bell-shaped curves.

The strong law of large numbers

Let $X_1, X_2, ... X_n, ...$ be a sequence of independent and identically distributed RV's each having a finite mean $\mu = E(X_i)$. Then with pbt 1, $\frac{X_1 + X_2 + ... + X_n}{n} \to \mu$ as $n \to \infty$.

One Sided Chebyshev Inequality

(To obtain an upper bound for a probability of the form $P\{X - \mu \geq a\}$ where 'a' is some positive value and when only the mean $\mu = E(X)$ and variance $\sigma^2 = Var(X)$ of the distribution of X are known)

If X is a random variable with mean 0 and finite variance σ^2 then for any $a > 0$, $P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$

.

Corollary:

If $E(X) = \mu, Var(X) = \sigma^2$, then for a > 0,

$P\{X \geq \mu + a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$ (i.e.) $P\{X - \mu \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$

and $P\{X \leq \mu - a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$ (i.e.) $P\{\mu - X \geq a\} \leq \frac{\sigma^2}{\sigma^2 + a^2}$.

Example:

If the number of items produced in a factory during a week is a RV with mean 100 and variance 400, compute an upper bound on the probability that this week's production will be atleast 120.

Solution:

Using one-sided Chebyshev's inequality

$$P(X \geq 120) = P(X - 100 \geq 20) \leq \frac{400}{400 + 20^2}$$
$$\leq \frac{1}{2}$$

Note:

1) Using Markov's inequality

$P(X \geq 120) \leq \frac{E(X)}{120} = \frac{5}{6}$ which is a far weaker bound.

2) When moment generating function of the RV X is known, we can obtain even more effective bounds on $P(X \geq a)$.

Moment Generating Function of a RV X (mgf)

$M_X(t) = \mathrm{E}(\mathrm{e}^{tx})$ is the mgf of the random variable X.

Chernoff Bounds

$$\begin{aligned} P(X \geq a) &\leq \mathrm{e}^{-at} M_X(t) \forall t > 0 \\ P(X \leq a) &\leq \mathrm{e}^{-at} M_X(t) \forall t < 0 \end{aligned} \qquad \text{(II)}$$

Equation (II) is known as Chernoff inequality.

The Chernoff's bound can be obtained by minimizing the right hand side

$$P(X \geq a) \leq \min \mathrm{e}^{-at} M_X(t), \ \forall t \geq 0$$
$$\& \, P(X \leq a) \leq \min \mathrm{e}^{-at} M_X(t), \ \forall t \leq 0$$

(i.e., Chernoff bounds hold for all t in either the positive or negative quadrant. We obtain the best bound on $P(X \geq a)$ by using the 't' that minimizes $\mathrm{e}^{-at} M_X(t)$)

Chernoff Bounds for the Standard Normal RV

Let $Z = \frac{X - \mu}{\sigma}$ be a standard normal random variable.

Then $M_Z(t) = e^{\frac{t^2}{2}}$.

Now $P(Z \geq a) \leq \mathrm{e}^{-at} e^{\frac{t^2}{2}}, \forall t > 0$.

To minimize R.H.S. (i.e.) $\mathrm{e}^{-at} e^{\frac{t^2}{2}}$.

(i.e.) to minimize $\frac{t^2}{2} - at$.

$\frac{2t}{2} - a = 0 \rightarrow t = a$.

Thus for a>0.

$P(\mathrm{Z} \geq \mathrm{a}) \leq e^{\frac{-a^2}{2}}$ and for a<0, $P(\mathrm{Z} \leq \mathrm{a}) \leq e^{\frac{-a^2}{2}}$.

Chernoff Bounds for the Poisson RV

X is a Poisson random variable.

$M_X(t) = e^{\lambda(e^t - 1)}$.

$\therefore$ Chernoff bound on $P(X \geq i)$ is $P(X \geq i) \leq e^{\lambda(e^t-1)} e^{-it}, t > 0$.

Minimizing R.H.S. (i.e.) minimizing $\lambda(e^t - 1) - it$.

We get $\lambda e^t - i = 0 \Rightarrow e^t = \frac{i}{\lambda}$ provided $\frac{i}{\lambda} > 1$.

$$\text{Now } P(X \geq i) \leq \mathrm{e}^{\lambda\left(\frac{i}{\lambda}-1\right)} \left(\frac{i}{\lambda}\right)^{-i}$$
$$\leq \frac{e^i e^{-\lambda} \lambda^i}{i^i} = \frac{(e\lambda)^i e^{-\lambda}}{i^i}$$

State and prove Jensen's inequality

If f(x) is a convex function then $E[f(x)] \geq \mathrm{f[E(x)]}$ provided that the expectation exist and are finite.

[$f(x)$ is convex if $f''(x) \geq 0, \forall\, \mathrm{x}$]

Proof:

Expanding $f(x)$ in a Taylor's series expansion about $\mu = E(X)$ yields

$f(x) = f(\mu) + f'(\mu)(x-\mu) + \frac{f''(\xi)(x-\mu)^2}{2!}$ where ξ is some value between x & μ .

Since $f''(\xi) \geq 0$, we obtain

$f(x) \geq f(\mu) + f'(\mu)(x-\mu)$

Taking expectation yields

$E[f(x)] \geq f(\mu) + f'(\mu)E(x-\mu) = f(\mu)$

and hence $E[f(X)] \geq f[E(X)]$.

State and prove Cauchy Schwartz inequality

Statement:

For any two random variables X & Y, then $\{E(XY)\}^2 \leq E(X^2).E(Y^2)$.

Proof:

$E\{(X-tY)\}^2 \geq 0$ where 't' is real.

$E(X^2) - 2tE(XY) + t^2E(Y^2) \geq 0$

$(\text{i.e.})\, t^2E(Y^2) - 2tE(XY) + E(X^2) \geq 0$

L.H.S. is a quadratic expression in t, (i.e.) discriminant of L.H.S. is less than or equal to 0.

$4[E(XY)]^2 \leq 4E(X^2).E(Y^2)$

$\Rightarrow \{E(XY)\}^2 \leq E(X^2).E(Y^2)$

Tchebycheff Inequality:

Introduction: If we know the prob. distribution of a r.v X, we may compute $E(X)$ and $Var(X)$. Conversly, If $E(X)$ and $Var(X)$ are known, we cannot construct the prob. distribution of X and hence compute quantities such as $P\{|X-E(X)| \le k\}$. Although we cannot evaluate such probabilities from a knowledge of $E(X)$ and $V(X)$, several approximation techniques have been developed to upper and/or lower bounds to such probabilities. The most important of such techniques is Tchebycheff inequality.

Statement: If X is a RV with $E(X)=\mu$ and $Var(X)=\sigma^2$, then $P\{|X-\mu| \ge c\} \le \frac{\sigma^2}{c^2}$ where $c>0$.

or $P\{|X-\mu| \le c\} \ge 1-\frac{\sigma^2}{c^2}$ where $c>0$.

Alternative forms: If we put $c=k\sigma$, where $k>0$, then Tchebycheff inequality takes the form

$$P\left\{\left|\frac{X-\mu}{k}\right| \ge \sigma\right\} \le \frac{1}{k^2}$$

(or) $$P\left\{\left|\frac{X-\mu}{k}\right| \le \sigma\right\} \ge 1-\frac{1}{k^2}$$

1) A r.v X has mean $\mu = 12$ and variance $\sigma^2 = 9$ and an unknown prob. distribution. Find $P(6 < X < 18)$.

Sol:

By Tchebycheff's inequality

$$P\{|X-\mu| \le c\} \ge 1 - \frac{\sigma^2}{c^2}$$

$$\Rightarrow P\{-c \le X-\mu \le c\} \ge 1 - \frac{\sigma^2}{c^2}$$

$$\Rightarrow P\{\mu - c \le X \le \mu + c\} \ge 1 - \frac{\sigma^2}{c^2}$$

~~Taking~~ Given $\mu = 12$, $\sigma^2 = 9$

$$P\{12 - c < X < 12 + c\} \ge 1 - \frac{9}{c^2}$$

putting $c = 6$, we get

$$P\{6 < X < 18\} \ge 1 - \frac{9}{36}$$

$$\Rightarrow P\{6 < X < 18\} \ge 3/4$$

2.) If the RV X is uniformly distributed over $(-\sqrt{3}, \sqrt{3})$, Compare $P\{|X-\mu| \ge 3\sigma/2\}$ and compare it with the upper bound obtained by Tchebycheff's inequality.

Sol:

Uniform distribution pdf: $f(x) = \frac{1}{b-a}$, $a < x < b$

$$\text{Mean} = \frac{1}{2}(b+a)$$

$$\text{Variance} = \frac{1}{12}(b-a)^2$$

Scanned with CamScanner

Mean $= \mu = \frac{\sqrt{3}-\sqrt{3}}{2} = 0$, Variance $= \sigma^2 = \frac{1}{12}(\sqrt{3}+\sqrt{3})^2 = 1$

ie, $\mu = 0$, $\sigma^2 = 1$

$$\therefore P\left\{|X-\mu| \geq \frac{3\sigma}{2}\right\} = P\{|X| \geq 3/2\}$$

$$= 1 - P\{-3/2 < X < 3/2\}$$

$$= 1 - \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}}\,dx \qquad \because f(x) = \frac{1}{b-a}, \; a<x<b$$

$$= 1 - \frac{1}{2\sqrt{3}}\left[x\right]_{-3/2}^{3/2}$$

$$= 1 - \frac{1}{2\sqrt{3}}\left[\frac{3}{2}+\frac{3}{2}\right]$$

$$= 1 - \frac{\sqrt{3}}{2} = 0.134$$

By Tchebycheff's inequality

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

$$\therefore \quad P\left\{|X-\mu| \geq \frac{3}{2}\sigma\right\} \leq \frac{4}{9} = 0.444$$

which is a poor upper bound.

Can we find a RV X for which $P\{\mu-2\sigma \leq X \leq \mu+2\sigma\} = 0.6$?

Sol:

$$P\{\mu-2\sigma \leq X \leq \mu+2\sigma\} = P\{|X-\mu| \leq 2\sigma\}$$

By Tchebycheff's inequality $P\{|X-\mu| \leq c\} \geq 1-\frac{\sigma^2}{c^2}$

(or) $P\left\{\left|\frac{X-\mu}{k}\right| \leq \sigma\right\} \geq 1-\frac{1}{k^2}$

$$\Rightarrow P\{\mu-2\sigma \leq X \leq \mu+2\sigma\} \geq 1-\frac{1}{2^2} \quad \because k = \frac{1}{2}\; 2$$

$$\geq 0.75$$

$\therefore$ there does not exist a RV X satisfying the given condition.

A discrete RV X takes the values $-1, 0, 1$ with probabilities $\frac{1}{8}, \frac{3}{4}, \frac{1}{8}$ respectively. Evaluate $P\{|X-\mu| \geq 2\sigma\}$ and Compare it with the upper bound given by Tchebycheff's inequality.

Sol:

Given

X	-1	0	1
$P(X=x)$	1/8	3/4	1/8

$$E(X) = \left(-1\times\frac{1}{8}\right) + \left(0\times\frac{3}{4}\right) + \left(1\times\frac{1}{8}\right) = 0$$

$$E(X^2) = \left((-1)^2\times\frac{1}{8}\right) + \left(0\times\frac{3}{8}\right) + \left(1^2\times\frac{1}{8}\right) = \frac{2}{8} = \frac{1}{4}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{4} - 0 = \frac{1}{4}$$

$$\Rightarrow P\{|X-\mu| \geq 2\sigma\} = P\{|X-0| \geq 2\cdot\frac{1}{2}\}$$

$$= P\{|X| \geq 1\}$$

$$= 1 - P\{|X| < 1\}$$

$$= 1 - P\{-1 < X < 1\} = 1 - P\{X=0\}$$

$$= 1 - \frac{3}{4} = 1/4$$

By Tchebycheff's inequality

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

$$\Rightarrow P\{|X-\mu| \geq 2\sigma\} \leq \frac{1}{2^2} = \frac{1}{4}$$

The two values Coincide.

5.) A fair die is tossed 720 times. Use Tchebycheff's inequality to find a lower bound for the prob. of getting 100 to 140 sixes.

Sol: Let X be the number of sixes

$p = P\{\text{getting '6' in a single toss}\} = 1/6$

$q = \frac{5}{6}$ and $n = 720$

X follows a binomial distribution

$\therefore$ Mean $= np$ and Variance $= npq$

$\Rightarrow$ mean $= 720 \times \frac{1}{6}$ and Variance $= 720 \times \frac{1}{6} \times \frac{5}{6} = 100$

ie, $\mu = 120$ and $\sigma = 10$.

By Tchebycheff's inequality

$$P\{|X-\mu| \leq K\sigma\} \geq 1 - \frac{1}{K^2}$$

$$\Rightarrow P\{|X-120| \leq 10K\} \geq 1 - \frac{1}{K^2}$$

$$\Rightarrow P\{-10K \leq X - 120 \leq 10K\} \geq 1 - 1/K^2$$

$$\Rightarrow P\{120 - 10K \leq X \leq 120 + 10K\} \geq 1 - \frac{1}{K^2}$$

putting $K = 2$, we get

$$P\{100 \leq X \leq 140\} \geq 1 - \frac{1}{4} = 3/4$$

$\therefore$ the lower bound for the prob. $= 0.75$

6) Use Tchebycheff's inequality to find how many times a fair coin must be tossed in order that the prob. that the ratio of the number of heads to the number of tosses will lie between 0.45 and 0.55 will be at least 0.95.

Sol: Given $P\left\{0.45 \le \frac{X}{n} \le 0.55\right\} \geq 0.95 \longrightarrow (1)$

Let X be the number of heads

Then X follows a binomial distribution.

ie, $X \sim B(np, \sqrt{npq})$ where $p = q = 1/2$

$\therefore \frac{X}{n} \sim B\left(\frac{1}{2}, \frac{1}{2\sqrt{n}}\right)$

By Tchebycheff's inequality

$$P\left\{\left|\frac{X}{n} - \frac{1}{2}\right| \le c\right\} \geq 1 - \frac{\sigma^2}{c^2}$$

$$\Rightarrow P\left\{0.5 - c \le \frac{X}{n} \le 0.5 + c\right\} \geq 1 - \frac{1}{4nc^2} \quad \left[\because \sigma^2 = \frac{1}{4n}\right]$$

Putting $c = 0.05$, we get

$$P\left\{0.45 \le \frac{X}{n} \le 0.55\right\} \geq 1 - \frac{100}{n} \longrightarrow (2)$$

Comparing (1) and (2), we get

$$1 - \frac{100}{n} = 0.95$$

$$\Rightarrow 1 - 0.95 = \frac{100}{n}$$

$$\Rightarrow n = \frac{100}{0.05}$$

$$\boxed{n = 2000}$$

Scanned with CamScanner

A RV X is exponentially distributed with parameter 1. Use Tchebycheff's inequality to show that $P(-1 \leq X \leq 3) \geq 3/4$. Find the actual probability also.

Sol:

For Exponential distribution:

$$\text{pdf}: f(x) = \lambda e^{-\lambda x}, \; x \geq 0$$

$$\text{Mean} = \frac{1}{\lambda} \quad \text{and} \quad \text{variance} = \frac{1}{\lambda^2}$$

Here $\lambda = 1$

i.e., $\mu = 1$ and $\sigma = 1$

By Tchebycheff's inequality,

$$P\{|X-1| \leq 2\} \geq 1 - 1/4$$

$$\Rightarrow P\{-1 \leq X \leq 3\} \geq 3/4$$

$\because$ ~~$P(-1 \leq X \leq 3)$~~ $= P\{|X-\mu| \leq k\sigma\} \geq 1 - \frac{1}{k^2}$

To find Actual probability:

$$P(-1 \leq X \leq 3) = \int_0^3 e^{-x}\,dx$$

$$= \left(-e^{-x}\right)_0^3 = 1 - e^{-3} = 0.9502$$

Markov's inequality: If X is a r.v that takes only non-negative values, then for any $a>0$,

$$P\{X \geq a\} \leq \frac{E(X)}{a}$$

Note: (i) Both the inequalities (Markov's and Chebyshev's) enable us to derive bounds on probabilities.

(ii) In Markov inequality mean of the probability distribution is known.

(iii) In Chebyshev's inequality both the mean and the variance of the prob. dist. are known.

1) If the number of items produced in a factory during a week is a RV with mean 100 and variance 400, Compute an upper bound on the prob. that this week's production will be atleast 120.

Sol:

Using Markov's inequality $P(X \geq a) \leq \frac{E(X)}{a}$

$$\Rightarrow P(X \geq 120) \leq \frac{100}{120} = \frac{5}{6}$$

2) Suppose that the average grade on Maths exam is 70%. Give an upper bound on the proportion of students

who score atleast 90%.

Sol:

Given $E(X) = \frac{70}{100} = \frac{7}{10}$

Markov's inequality : $P(X \geq k) \leq \frac{E(X)}{k}$

$\Rightarrow P(X \geq 90\%) \leq \frac{70\%}{90\%} = 77.8\%$

ie, atmost 77.8% of students can possible score this high.

3.) A coin is weighted so that its probability of landing on heads is 20%. Suppose the coin is flipped 20 times. Find the bound for the prob. If lands on heads atleast 16 times. ~~and~~ also ~~find actual~~.

Sol:

Given $n = 20$, $p = 20\% = \frac{20}{100} = \frac{1}{5}$

$\therefore$ Mean $= np = 20 \times \frac{1}{5} = 4$

$P(X \geq 16) \leq \frac{E(X)}{16} = \frac{4}{16} = \frac{1}{4}$

Corollary: If $E(X)=\mu$, $Var(X)=\sigma^2$, then for $a>0$

$$P\{X \geq \mu+a\} \leq \frac{\sigma^2}{\sigma^2+a^2} \quad \text{or} \quad P\{X-\mu \geq a\} \leq \frac{\sigma^2}{\sigma^2+a^2}$$

$$P\{X \leq \mu-a\} \leq \frac{\sigma^2}{\sigma^2+a^2} \quad \text{or} \quad P\{\mu-X \geq a\} \leq \frac{\sigma^2}{\sigma^2+a^2}$$

One sided Chebyshev Inequality:

If X is a r.v with mean 0 and finite variance σ^2 then for any $a>0$, $P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2+a^2}$.

1.) If the number of items produced in a factory during a week is a Rv with mean 100 and variance 400, compute an upper bound on the prob. that this week's production will be atleast 120.

Sol:

Using one-sided Chebyshev's inequality

$$P(X \geq 120) = P(X-100 \geq 20) \leq \frac{400}{400+20^2}$$

$$\leq \frac{1}{2}$$

— x —

1.) If X denotes the sum of the numbers obtained when 2 dice are thrown, obtain an upper bound for $P\{|X-7| \geq 4\}$. Compare with the exact probability.

Sol:

$S = \{(1,1)\ (1,2)\ (1,3)\ (1,4)\ (1,5)\ (1,6)$

$(2,1)\ (2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6)$

$(3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6)$

(4,1) (4,2) (4,3) (4,4) (4,5) (4,6)

(5,1) (5,2) (5,3) (5,4) (5,5) (5,6)

(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) }

X – the sum of the numbers obtained.

X	2	3	4	5	6	7	8	9	10	11	12
$P(X=x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

$$E(X) = \frac{2}{36} + \frac{6}{36} + \frac{12}{36} + \frac{20}{36} + \frac{30}{36} + \frac{42}{36} + \frac{40}{36} + \frac{36}{36} + \frac{30}{36} + \frac{22}{36} + \frac{12}{36}$$

$$= \frac{252}{36} = 7$$

$$E(X^2) = \frac{4}{36} + \frac{18}{36} + \frac{48}{36} + \frac{100}{36} + \frac{180}{36} + \frac{294}{36} + \frac{320}{36} + \frac{324}{36}$$

$$+ \frac{300}{36} + \frac{242}{36} + \frac{144}{36} = \frac{1974}{36} = \frac{329}{6}$$

$$Var(X) = E(X^2) - [E(X)]^2 = \frac{329}{6} - 49 = \frac{35}{6}$$

By Tchebycheff's inequality

$$P\{|X-\mu| \geq c\} \leq \frac{\sigma^2}{c^2}$$

$$\therefore P\{|X-7| \geq 4\} \leq \frac{35}{96}$$

Now;

$$P\{|X-7| \geq 4\} = P\{X = 2, 3, 11 \text{ or } 12\}$$

$$= \frac{1}{36} + \frac{2}{36} + \frac{2}{36} + \frac{1}{36} = \frac{1}{6}$$

There is much difference between the two values.

2.) A discrete RV X can assume the values $x = 1, 2, 3, \ldots$ with probability 2^{-x}. Show that $P\{|X-2| \geq 2\} \leq 1/2$, while the actual probability is $1/8$.

Sol:

X	1	2	3	4	5	...
$P(X=x)$	$\frac{1}{2}$	$\frac{1}{2^2}$	$\frac{1}{2^3}$	$\frac{1}{2^4}$	$\frac{1}{2^5}$	...

$$E(X) = \sum_{x=1}^{\infty} x\, 2^{-x}$$

$$= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2^2} + 3 \cdot \frac{1}{2^3} + 4 \cdot \frac{1}{2^4} + \ldots$$

Since $(1-x)^{-2} = 1 + 2x + 3x^2 + 4x^3 + \ldots$

$$\therefore E(X) = \frac{1}{2}\left[1 + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2^2} + 4 \cdot \frac{1}{2^3} + \ldots\right]$$

$$= \frac{1}{2}(1 - 1/2)^{-2} = \frac{1}{2} \cdot 4 = 2$$

$$E(X^2) = \sum_{x=1}^{\infty} x^2\, 2^{-x}$$

$$= \sum_{x=1}^{\infty} [x^2 + x - x]\, 2^{-x}$$

$$= \sum_{x=1}^{\infty} x(x+1)\, 2^{-x} - \sum_{x=1}^{\infty} x\, 2^{-x}$$

$$= \left[1 \cdot 2 \cdot \frac{1}{2} + 2 \cdot 3 \cdot \frac{1}{2^2} + 3 \cdot 4 \cdot \frac{1}{2^3} + \ldots\right] - 2$$

$$= \frac{1}{2}\left[1 \cdot 2 + 2 \cdot 3 \cdot \frac{1}{2} + 3 \cdot 4 \cdot \frac{1}{2^2} + \ldots\right] - 2$$

$$= \frac{1}{2} \cdot 2\left[1 + 3 \cdot \frac{1}{2} + 6 \cdot \frac{1}{2^2} + 10 \cdot \frac{1}{2^3} + \ldots\right] - 2$$

$$= \left(1 - \frac{1}{2}\right)^{-3} - 2 = 6$$

$$\therefore Var(X) = 6 - 4 = 2$$

Scanned with CamScanner

By Tchebycheff's inequality,

$$P\{|X-\mu| \geq c\} \leq \sigma^2/c^2$$

$$\Rightarrow P\{|X-2| \geq 2\} \leq \frac{2}{4} = 1/2$$

Now, $P\{|X-2| \geq 2\} = 1 - P\{|X-2| < 2\}$

$$= 1 - P\{-2 < X-2 < 2\}$$

$$= 1 - P\{0 < X < 4\}$$

$$= 1 - \left[\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3}\right] = 1/8$$

3) If $E(X) = 8$ and $E(X^2) = 68$, find a lower bound for $P(5 < X < 11)$ using Tchebyshev's inequality.

<u>Sol</u>:

$$Var(X) = E(X^2) - [E(X)]^2 = 68 - 64 = 4$$

By Tchebycheff's inequality $P\{|X-\mu| \leq c\} \geq 1 - \sigma^2/c^2$

$$\Rightarrow P\{|X-8| \leq c\} \geq 1 - \frac{4}{c^2}$$

$$\Rightarrow P\{-c \leq X-8 \leq c\} \geq 1 - \frac{4}{c^2}$$

$$\Rightarrow P\{8-c \leq X \leq c+8\} \geq 1 - \frac{4}{c^2}$$

Put $c = 3$, we get

$$\Rightarrow P\{5 \leq X \leq 11\} \geq 1 - \frac{4}{9} = 5/9$$

H.W

4) A r.v X has mean 10 and variance 16. Find the lower bound for $P(5 < X < 15)$.

5.) Use Tchebycheff's inequality to prove that $P(X=\mu)=1$, if $Var(X)=0$.

Proof: By Tchebycheff's inequality

$P\{|X-\mu| \geq c\} \leq \frac{\sigma^2}{c^2}$ and $P\{|X-\mu| \leq c\} \geq 1-\sigma^2/c^2$

Now $P\{|X-\mu| \geq c\} \leq 0$ and $P\{|X-\mu| \leq c\} \geq 1$ $\because \sigma^2=0$ $\longrightarrow$ (1)

(1) holds even for small values of c

Hence in the limit when $c \to 0$

$P\{|X-\mu|=0\}=1 \Rightarrow P\{X=\mu\}=1$

6.) A r.v has the pdf $f_X(x)=3e^{-3x}$, $x>0$ obtain an upper bound for $P(X \geq 1)$.

Sol:

$$E(X)=\int_{-\infty}^{\infty} x f(x)\,dx=\int_0^{\infty} x \cdot 3e^{-3x}\,dx=\frac{1}{3}$$

$$E(X^2)=\int_{-\infty}^{\infty} x^2 f(x)\,dx=\int_0^{\infty} x^2 \cdot 3e^{-3x}=\frac{2}{9}, \quad Var(X)=E(X^2)-[E(X)]^2$$

$$=\frac{2}{9}-\frac{1}{9}=1/9$$

By one sided chebyshev inequality.

$\cancel{P\{X \geq a\} \leq \frac{\sigma^2}{\sigma^2+a^2}}$ $P\{X \geq \mu+a\} \leq \frac{\sigma^2}{\sigma^2+a^2}$

$\Rightarrow \cancel{P\{X \geq 1\} \leq \frac{1/9}{\frac{1}{9}+1}} \cancel{\frac{1/9}{10/9}}$ $P\{X \geq \frac{1}{3}+\frac{2}{3}\} \leq \frac{1/9}{\frac{1}{9}+\frac{4}{9}}$

$\leq \frac{1}{5}$

$\leq \cancel{\frac{1}{10}}$

Scanned with CamScanner

Central limit theorem

and

Transformation of r.v

Central limit theorem

Central limit theorem says that the prob. distribution of the sum of a large number of independent random variables approaches a normal distribution.

Theorem: If $X_1, X_2, \ldots X_n$ is a sequence of n independent and identically distributed (i.i.d) r.v's, each having mean μ and variance σ^2, and if $\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$, then the variate $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ has a distribution that approaches the standard normal distribution as $n \to \infty$, provided the m.g.f. exists.

Different forms of Central limit theorem

Central Limit Theorem (Liapounoff's Form)

If $X_1, X_2, \ldots X_n \ldots$ be a sequence of independent r.v's with $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$, $i = 1, 2, \ldots$ 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.

Central Limit Theorem (Lindeberg-Levy's Form)

If $X_1, X_2, \ldots X_n \ldots$ be a sequence of independent identically distributed r.v's with $E(X_i) = \mu$ and $var(X_i) = \sigma^2$, $i = 1, 2, \ldots$ and if $S_n = X_1 + X_2 + \ldots X_n$, then under general conditions, S_n follows a normal distribution with mean $n\mu$ and variance $n\sigma^2$ as n tends to infinity.

ie, $S_n \sim N(n\mu, \sigma\sqrt{n})$

Corollary: If $\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ then $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$.

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

The corollary says that if $\bar{X}$ is the mean of a sample of size n taken from a population with mean μ and Variance σ^2 then $z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ as $n \to \infty$ is a standard normal distribution.

1) The lifetime of a certain brand of an electric bulb may be considered a RV with mean 1200h and standard deviation 250h. Find the probability, using central limit theorem, that the average lifetime of 60 bulbs exceeds 1250h.

Sol: Let X_i represents the lifetime of the bulb.

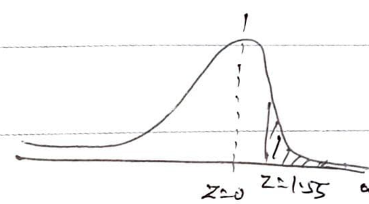
$E(X_i) = 1200$ and $Var(X_i) = (250)^2$, $n = 60$

Let $\bar{X}$ denote the mean lifetime of 60 bulbs.

By Corollary of Lindeberg - Levy form of CLT

$\bar{X} \sim N\left(1200, \frac{250}{\sqrt{60}}\right)$.

$$P(\bar{X} > 1250) = P\left(\frac{\bar{X} - 1200}{\frac{250}{\sqrt{60}}} > \frac{1250 - 1200}{\frac{250}{\sqrt{60}}}\right)$$



$$= P(Z > 1.55) = 0.5 - 0.4394 = 0.0606$$

$[\because P(Z > 1.55) = P(0 < Z < \infty) - P(0 < Z < 1.55)$

2) If $X_1, X_2, \ldots X_n$ are Poisson variates with parameter $\lambda = 2$, use the central limit theorem to estimate $P(120 \le S_n \le 160)$, where $S_n = X_1 + X_2 + \ldots + X_n$ and $n = 75$.

Sol:

$E(X_i) = \lambda = 2$ and $var(X_i) = \lambda = 2$

By CLT, S_n follows $N(n\mu, \sigma\sqrt{n})$ ie, $S_n \sim N(150, \sqrt{150})$

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

$$= P\{-2.45 \le Z \le 0.82\}$$

$$= P\{-2.45 \le Z \le 0\} + P\{0 \le Z \le 0.82\}$$

$$= P\{0 \le Z \le 2.45\} + P\{0 \le Z \le 0.82\}$$

$$= 0.4929 + 0.2939 = 0.7866$$

3) 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.

Sol: Given $E(X_i) = \mu$ and $var(X_i) = 1.5$

Let $\bar{X}$ denote the sample mean.

By Corollary under CLT, $\bar{X} \sim N\left(\mu, \frac{\sqrt{1.5}}{\sqrt{n}}\right)$

We have to find n such that

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

$$\Rightarrow P(-0.5 < \bar{X} - \mu < 0.5) \geq 0.95$$

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

$$\Rightarrow P\left(\frac{-0.5}{\sqrt{1.5/n}} < Z < \frac{0.5}{\sqrt{1.5/n}}\right) \geq 0.95$$

$$\Rightarrow P\left(\frac{-0.5\sqrt{n}}{\sqrt{1.5}} < Z < \frac{0.5\sqrt{n}}{\sqrt{1.5}}\right) \geq 0.95$$

$$\Rightarrow 2P\left(0 < Z < \frac{0.5\sqrt{n}}{\sqrt{1.5}}\right) \geq 0.95$$

$$\therefore P(0 < Z < 0.4082\sqrt{n}) \geq 0.475$$

$P(0 < Z < Z_1) \geq 0.475$ where $Z_1 = 0.4082\sqrt{n} \rightarrow (1)$

From the table $Z_1 = 1.96 \rightarrow (2)$

From (1) & (2), we have

$$0.4082\sqrt{n} = 1.96$$

$$\sqrt{n} = \frac{1.96}{0.4082} = 4.8016$$

$$\Rightarrow n = 23.055 \simeq 23$$

$\therefore$ the size of the sample must be at least 23

4.) If V_i, $i = 1, 2, \ldots 20$, are independent noise voltages received in an 'adder' and V is the sum of the voltages received, find the probability that the total incoming voltage V exceeds 105, using CLT. Assume that each of the r.v's V_i is uniformly distributed over $(0, 10)$.

Sol:

Given V_i, $i = 1, 2, \ldots 20$ follows uniform distribution over $(0,10)$

By defn: Uniform distribution: $f(x) = \frac{1}{b-a}$, $a < x < b$

Mean $= \frac{a+b}{2}$ and Variance $= \frac{(b-a)^2}{12}$

$\therefore$ p.d.f of $V_i = \frac{1}{10-0} = \frac{1}{10}$

$$\text{Mean} = E(V_i) = \frac{b+a}{2} = \frac{10+0}{2} = 5$$

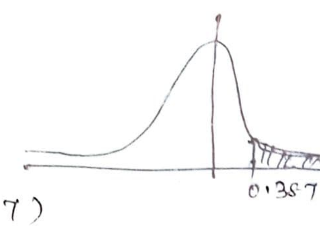
$$Var(V_i) = \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = \frac{(10)^2}{12}$$

The total voltage is $V = V_1 + V_2 + \cdots + V_n$

By CLT, $V : S_n \sim N(n\mu, \sigma\sqrt{n})$ ie, $S_n \sim N\left(100, \frac{10\sqrt{20}}{\sqrt{12}}\right)$

$$\therefore P(V > 105) = P\left(\frac{V-100}{\frac{10\sqrt{20}}{\sqrt{12}}} > \frac{105-100}{\frac{10\sqrt{20}}{\sqrt{12}}}\right)$$

$$= P(Z > 0.387)$$



$$= P(0 < Z < \infty) - P(0 < Z < 0.387)$$

$$= 0.5 - 0.1517 = 0.3483$$

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

Sol:

Given $E(X_i) = \mu = 60$, $Var(X_i) = \sigma^2 = 400$

Using CLT, $\bar{X}$ follows normal dist. with mean

$\mu = 60$ and Variance $= \frac{\sigma^2}{n} = 4$.

ie, $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \Rightarrow \bar{X} \sim N(60, \frac{20}{\sqrt{100}})$

To find $P(|\bar{X} - \mu|) \leq 4)$:

$$P(|\bar{X} - \mu| \leq 4) = P(|\bar{X} - 60| \leq 4)$$

$$= P(-4 \leq \bar{X} - 60 \leq 4)$$

$$= P(56 \leq \bar{X} \leq 64)$$

$$= P\left(\frac{56-\mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{64-\mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left(\frac{56-60}{20/\sqrt{100}} \leq Z \leq \frac{56-60}{20/\sqrt{100}}\right)$$

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

$$= 2P(0 \leq Z \leq 2)$$

$$= 2 \times 0.4772 = 0.9544$$

6 30 electronic devices $D_1, D_2, \dots D_{30}$ are used in the following manner. As soon as D_1 fails, D_2 becomes operative. when D_2 fails, D_3 becomes operative and so on, If the time to failure of D_i is an exponentially distributed r.v with parameter $\lambda = 0.1/h$ and T is the total time of operation of all the 30 devices, find the prob.

that T exceeds 350 hrs, using CLT.

Sol:

Exponential dist. mean $= \frac{1}{\lambda}$, Variance $= \frac{1}{\lambda^2}$

Given $\lambda = 0.1/hr$

$\therefore$ Mean $= \frac{1}{0.1} \times \frac{10}{10} = 10$, Variance $= \frac{1}{(0.1)^2} = 100$

$T : S_n \sim N(n\mu, \sigma\sqrt{n}) = N(300, 10\sqrt{30})$

$P(T > 350) = P\left(z > \frac{350-300}{10\sqrt{30}}\right) = P(z > 0.9128) = P(z > 0.91)$

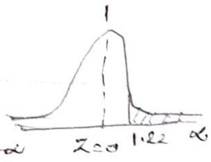
$= 0.5 - 0.3186 = 0.1814$

7) If X_i, $i = 1, 2, \ldots 50$, are independent RV's, each having a Poisson distribution with parameter $\lambda = 0.03$ and $S_n = X_1 + X_2 + \cdots + X_n$, evaluate $P(S_n \geq 3)$, using CLT. Compare your anwer with the exact value of the probability.

Sol:

Given $\lambda = 0.03$, $n = 50$, $X_i \sim$ Poisson distribution

Using CLT, $S_n \sim N(1.5, \sqrt{1.5})$



$P(S_n \geq 3) = P\left(z \geq \frac{3-1.5}{\sqrt{1.5}}\right) = P(z \geq 1.2247)$

$= 0.5 - P(0 < z < 1.22)$

$= 0.5 - 0.3888 = 0.1112$

Exact value: $P(S_n \geq 3) = P(S_{50} \geq 3) = P(X_1 + X_2 + \cdots + X_{50} \geq 3)$

$$= 1 - P(X_1 + X_2 + \cdots + X_{50} < 3) = 1 - \sum_{x=0}^{2} e^{-(50 \times 0.03)} \frac{(50 \times 0.03)^x}{x!}$$

$$= 1 - \sum_{x=0}^{2} e^{-1.5} \frac{(1.5)^x}{x!} = 0.191$$

Note:

If $X_1, X_2 \ldots X_n$ are 'n' poisson r.v's with parameters $\lambda_1, \lambda_2 \ldots \lambda_n$ then $X_1 + X_2 + \cdots + X_n$ is a poisson ~~varia~~ r.v with parameter $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

8) A fair coin is tossed 250 times. Find the probability that heads will appear between 120 and 140 times using central limit theorem.

Sol:

Let X denote the number of heads appear.

$\therefore$ X is a binomial r.v with parameter $n = 250$ and $p = 1/2$.

$\therefore q = 1 - p = 1/2$

$\therefore$ Mean $= np = 125$, Variance $= npq = 62.5$

$\therefore$ By CLT, $X \sim N(125, \sqrt{62.5})$

$$P(120 < X < 140) = P\left(\frac{120-125}{\sqrt{62.5}} < Z < \frac{140-125}{\sqrt{62.5}}\right)$$

$$= P(-0.63 < Z < 1.89)$$

$$= P(-0.63 < Z < 0) + P(0 < Z < 1.89)$$

$$= 0.2357 + 0.4706 = 0.7063$$

9) 20 dice are thrown. Find the approximate probability that the sum obtained is between 65 and 75

Sol:

RV X_i is the value on the i^{th} die.

$E(X_i) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 7/2$

$Var(X_i) = E(X_i^2) - [E(X_i)]^2 = \left(\frac{1^2+2^2+3^2+4^2+5^2+6^2}{6}\right) - \frac{49}{4} = 35/12$

To find ~~By CLT~~ $P(65 \le \sum_{i=1}^{20} X_i \le 75)$

By CLT, $S_n \sim N(n\mu, \sigma\sqrt{n}) = N(70, 7.6376)$

$\therefore P(65 \le S_n \le 75) = P\left(\frac{65-70}{7.6376} \le z \le \frac{75-70}{7.6376}\right)$

$= P(-0.655 \le z \le 0.655)$

$= 2P(0 \le z \le 0.655) = 2(0.2422)$

$= 0.4844$

<u>Proof of Central limit theorem</u>

Let $S_n = X_1 + X_2 + \dots + X_n$, where each X_i is distributed with mean μ and variance σ^2

Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$

Then $E(Z_n)$ $E\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)$