March 30, 2020
Dr. SUSHIL KUMAR AZAD

Limit Theorems – Continued ...

In this lecture, we shall discuss some applications of limit theorems (Chebyschev's inequality, CLT & strong LLNs).

In the last lecture, we saw the proof of Chebyschev's inequality based on the Markov's inequality. We shall now give an alternative proof of Chebyschev's inequality for the continuous case, that is, if X is a continuous r.v. with mean, μ and variance, σ^2, then for any +ve constant k,

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

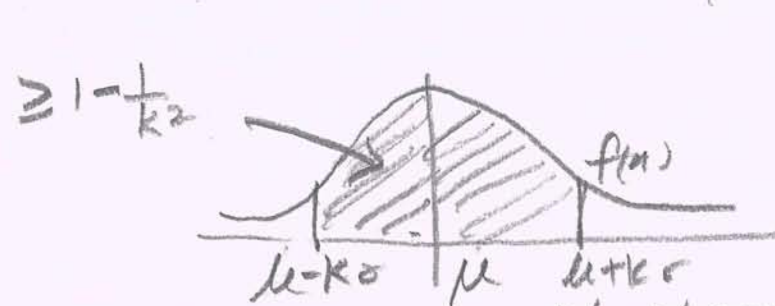
Proof (for the continuous case):

We have

$$\sigma^2 = E[(x-\mu)^2]$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 \underbrace{f(x)}_{\text{pdf of } X} \, dx$$

$$= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx \quad \text{--- (i)}$$



Since $f(x) \geq 0$, the 2^{nd} integral on the R.H.S of (i) is non-negative, therefore

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx. \quad \text{--- (ii)}$$

continued .

SUSHIL KUMAR AZAD

Now, in the 1st integral on the R.H.S. of (ii), we have $x < \mu - k\sigma$

$$\Rightarrow k\sigma < \mu - x \Rightarrow (\mu - x)^2 > k^2\sigma^2$$

$$\therefore \int_{-\infty}^{\mu - k\sigma} (x-\mu)^2 f(x)\,dx > \int_{-\infty}^{\mu - k\sigma} k^2\sigma^2 f(x)\,dx \quad \text{---(iii)}$$

and in the 2nd integral on the R.H.S. of (ii), we have

$$x > \mu + k\sigma \Rightarrow x - \mu > k\sigma$$

$$\Rightarrow (x-\mu)^2 > k^2\sigma^2 \quad \text{so}$$

$$\int_{\mu + k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx > \int_{\mu + k\sigma}^{\infty} k^2\sigma^2 f(x)\,dx \quad \text{---(iv)}$$

From (i), (iii) & (iv), we get

$$\sigma^2 \geq k^2\sigma^2 \left[\int_{-\infty}^{\mu - k\sigma} f(x)\,dx + \int_{\mu + k\sigma}^{\infty} f(x)\,dx \right]$$

or $$\frac{1}{k^2} \geq P[X \leq (\mu - k\sigma)] + P[X \geq (\mu + k\sigma)]$$

i.e. $$\frac{1}{k^2} \geq P[|X - \mu| \geq k\sigma] \quad \text{---(v)}$$

$$\Rightarrow P[|X-\mu| < k\sigma] = 1 - P[|X-\mu| \geq k\sigma]$$

$$\geq 1 - \frac{1}{k^2} \quad \text{from (v)}$$

i.e. $P[|X-\mu| < k\sigma] \geq 1 - \frac{1}{k^2}$, which is the result. //

SUSHIL KUMAR AZAD

Example 32. Let X be the number of items produced in a factory during a week, and $E(X) = 50$.

(a) What can be said about the probability that this week's production will exceed 75?

(b) If $\mathrm{var}(X) = 25$, then what can be said about the probability that this week's production will be between 40 and 60?

SUSHIL KUMAR AZAD

Solution. (a) Using the Markov's inequality, we get

$$P[X > 75] \le \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3}$$

(b) $P(40 < X < 60)$

$$= P(-10 < X - 50 < 10) = P(|X - 50| < 10)$$

By Chebyschev's inequality, we get

$$P(|X - 50| < 10)$$

$$= P(|X - 50| < 5 \times 2) \quad (\text{as } \sigma^2 = 25, \ \sigma = 5, \ k = 2)$$

$$\ge 1 - \frac{1}{k^2} = 1 - \frac{1}{4} = \frac{3}{4}$$

$\Rightarrow$ the probability that this week's production will be between 40 and 60 is at least 0.75.

//

Corollary (Chebyschev's inequality).

$$P((X - \mu) \ge k) \le \frac{\sigma^2}{k^2} \quad \text{when } k > 0$$

This follows from Chebyschev's inequality, since

$$X - \mu \ge k > 0 \Rightarrow |X - \mu| \ge k, \text{ so}$$

$$P((X - \mu) \ge k) \le P(|X - \mu| \ge k) \le \frac{\sigma^2}{k^2}. \quad //$$

Theorem 4 (One-sided Chebyshev's inequality)

If X is a r.v. with mean, 0 and finite variance, σ^2, then for any $k > 0$,

$$P[X \geq k] \leq \frac{\sigma^2}{\sigma^2 + k^2}$$

Proof. Let $b > 0$. Then

$$X \geq k \iff X + b \geq k + b$$

Hence, $P[X \geq k] = P[X + b \geq k + b]$

$$\leq P[(X+b)^2 \geq (k+b)^2],$$

Since $k + b > 0$, $(X+b) \geq (k+b) \Rightarrow (X+b)^2 \geq (k+b)^2$

Using Markov's inequality, we get,

$$P[X \geq k] \leq \frac{E[(X+b)^2]}{(k+b)^2} = \frac{\sigma^2 + b^2}{(k+b)^2}$$

Letting $b = \sigma^2/k$, we get

$$P[X \geq k] \leq \frac{\sigma^2 + \frac{\sigma^4}{k^2}}{(k^2+\sigma^2)^2/k^2} = \frac{\sigma^2(k^2+\sigma^2)}{(k^2+\sigma^2)^2}$$

$$= \frac{\sigma^2}{k^2+\sigma^2},$$ which is the result //

Example 33. Let X be the number of items produced in a factory during a week and $E(X) = 100$, and $Var(X) = 400$. Find an upper bound on the probability that this week's production will be at least 120, i.e. find $P[X \geq 120]$.

© SHIL KUMAR

SUSHIL KUMAR AZAD

Solution. We have

$$P[X \geq 120]$$
$$= P[X - 100 \geq 20]$$
$$= P[Y \geq 20] \quad (\text{where } E(Y) = E(X) - 100 = 0, \text{ and } \mathrm{var}(Y) = \mathrm{var}(X) = 400)$$

Hence, using Theorem 4, we have

$$\Rightarrow P[X \geq 120] = P[Y \geq 20]$$
$$\leq \frac{\mathrm{Var}(Y)}{\mathrm{var}(Y) + (20)^2} = \frac{\mathrm{Var}(X)}{\mathrm{var}(X) + 400}$$
$$= \frac{400}{800} = \frac{1}{2}$$
$$\Rightarrow P[X \geq 120] \leq \frac{1}{2} = 0.50$$

$\Rightarrow$ the probability that this week's production will be at least 120 is at most 0.50.

Example 34. Let X_i, $i = 1, \dots, 10$, be independent random variables each uniformly distributed over $(0, 1)$. Calculate an approximation to

$$P\left\{\sum_{i=1}^{10} X_i > 6\right\}$$

Solution. Since $X_i \sim U(0, 1)$, we get

$$E(X_i) = \frac{1+0}{2} = \frac{1}{2}, \text{ and}$$
$$\mathrm{var}(X_i) = \frac{(1-0)^2}{12} = \frac{1}{12} \quad \forall i = 1, \dots, 10$$

Now $P\{\sum_{i=1}^{10} X_i > 6\}$

$$= P\left\{\left(\sum_{i=1}^{10} X_i\right) - 5 > 1\right\} \quad (\mu = 0.5,\ n=10 \Rightarrow n\mu = 5)$$

$$= P\left\{\frac{X_1+\cdots+X_{10}-5}{\sqrt{10}\times\sqrt{\frac{1}{12}}} > \frac{1}{\sqrt{10}\sqrt{\frac{1}{12}}}\right\} \quad \left(\sigma = \sqrt{\tfrac{1}{12}} \text{ \& } n=10\right).$$

$$= P\left\{\frac{X_1+\cdots+X_{10}-5}{\sqrt{5/6}} > \sqrt{\frac{6}{5}} = \sqrt{1.2}\right\}$$

Using the CLT, $\left(\sum_{i=1}^{n} X_i - n\mu\right)/\sigma\sqrt{n} \longrightarrow$ the standard normal variate, as $n \to \infty$) we get

$$P\left\{\sum_{i=1}^{10} X_i > 6\right\} \approx 1 - \Phi(\sqrt{1.2})$$

$$\approx 1 - \Phi(1.095)$$

$$\approx 0.1367$$

$$\approx 0.14 \text{ (approx.)}$$

$\Rightarrow$ only 14 % of the time will $\sum_{i=1}^{10} X_i$ be greater than 6. //

SUSHIL KUMAR AZAD

Example 35 Let X_i, $i = 1, 2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) Poisson variates with mean, $E(X_i) = 2 = var(X_i)$. Find $P[190 < X_1 + X_2 + \cdots + X_{100} < 210]$.

Solution. Let $Y_{100} = X_1 + X_2 + \cdots + X_{100}$. Then,

$$E[Y_{100}] = \sum_{i=1}^{100} E(X_i) = 100 \times 2 = 200, \text{ and}$$

$$var(Y_{100}) = \sum_{i=1}^{100} var(X_i) = 200, \text{ that is,}$$

$$\mu_Y = 200 \text{ and } \sigma_Y^2 = 200.$$

By CLT, Y_{100} is approximately $N(200, 10\sqrt{2})$

Hence,
$$P(190 < Y_{100} < 210)$$
$$= P\left(-0.707 < \frac{Y_{100}-200}{10\sqrt{2}} < 0.707\right)$$
$$= P\left(\left|\frac{Y_{100}-200}{10\sqrt{2}}\right| < 0.707\right)$$
$$\simeq 0.52, \text{ (using the normal table)}$$

Example 36. Let X be the number of times that a fair coin, flipped 40 times, lands heads. Find the probability that $X = 20$. Use the normal approximation and then compare it to the exact solution.

Solution. We have, $n = 40$, θ, the probability that head occurs $= \frac{1}{2}$. Clearly $X \sim Bin(n, \theta)$

Hence, $\mu = E(X) = n\theta = 20$, and
$$\sigma^2 = var(X) = n\theta(1-\theta) = 10$$

Now, the binomial variable X is discrete, whereas the normal variable is continuous, so approximating X as normal variate, we have

$$P(X = 20) = P(19.5 < X < 20.5)$$
$$= P\left(\frac{19.5-20}{\sqrt{10}} < \frac{X-20}{\sqrt{10}} < \frac{20.5-20}{\sqrt{10}}\right)$$
$$= P\left(-0.16 < \frac{X-20}{\sqrt{10}} < 0.16\right)$$
$$\simeq \Phi(0.16) - \Phi(-0.16),$$

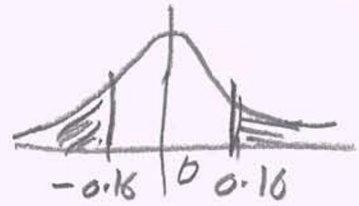
SUNIL KUMAR AZAD

We know, $\Phi(x) = P(Z \le x)$ ← standard normal variate

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$$

$$\therefore \Phi(-0.16) = P(Z > 0.16)$$

$$= 1 - \Phi(0.16)$$



SUSHIL KUMAR AZAD

Therefore, $P(X=20) \approx 2\,\Phi(0.16) - 1$.

$$= 2 \times 0.5636 - 1, \quad \text{(using the normal table)}$$

$$= 0.1272$$

$\Rightarrow P(X=20) \simeq 0.1272$.

Now the exact solution is given by, binomial distribution,

$$P(X=20) = {}^{40}C_{20}\left(\frac{1}{2}\right)^{40}$$

$\simeq 0.1268$, close to the normal approximate solution, (note: Error $= |0.1268 - 0.1272| = 0.0004) < 0.0005$,

i.e., Correct upto 3 decimal places. //

Example 37. Two unbiased dice are thrown. If X is the sum of the numbers showing up, then find an upper bound to the probability $P(|X - \mu_X| \ge 3)$.

Solution. The sample space S is

$$S = \{(i, j) \mid i, j = 1, 2, \cdots, 6\}.$$

Continued

Now, X denotes the sum of numbers showing up.
so X takes values

$$2, 3, \ldots, 12.$$

Then, prob. distribution of X is,

$X = x$	2	3	4	5	6	7	8	9	10	11	12
$P_X(x)$	$1/36$	$2/36$	$3/36$	$4/36$	$5/36$	$6/36$	$5/36$	$4/36$	$3/36$	$2/36$	$1/36$

Hence, $\mu = E(X) = \sum_{x=2}^{12} x\, p_X(x)$

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

$$E(X^2) = \sum_{x=2}^{12} x^2\, p_X(x)$$

$$= \frac{329}{6}, \text{ and}$$

$$\sigma^2 = \operatorname{var}(X) = E(X^2) - \mu^2$$

$$= \frac{35}{6}$$

Now, by Chebyschev's inequality, we have

$$P\left[\,|X - \mu_X| \geq k\sigma_X\right] \leq \frac{1}{k^2} \text{ for every } k > 0.$$

Take $k\sigma_X = 3$, that is, $k = \frac{3}{\sigma_X} = 3\sqrt{\frac{6}{35}}$

Then, $P\left[\,|X - \mu_X| > 3\right] = P\left[\,|X - \mu_X| \geq 3\right]$

$$= \leq \frac{1}{9} \times \frac{35}{6} = \frac{35}{54} \simeq 0.65.$$

$\Rightarrow$ the upper bound to the probability
$P(|X - \mu_X| \geq 3)$ is $\simeq \underline{0.65}$. //

SUSHIL KUMAR AZAD

Example 38. The lifetime of a special type of battery is a random variable with mean 40 hours and standard deviation 20 hours. A battery is used until it fails at which point it is replaced by a new one. Assuming a stockpile of 20 such batteries, the lifetimes of which are independent, approximate the probability that over 1100 hours of use can be obtained.

SUSHIL KUMAR AZAD

Solution: Let X_i denote the lifetime of the ith battery to be put in use.

Then, we want to find

$$P\{X_1 + X_2 + \cdots + X_{20} > 1100\}$$

Now

$$P\left\{\sum_{i=1}^{20} X_i > 1100\right\}$$

$$= P\left\{\frac{\sum_{i=1}^{20} X_i - 20\times 40}{20\sqrt{20}} > \frac{1100 - 20\times 20}{20\sqrt{20}}\right\}$$

(where $n = 20$, $\mu = 40$ and $\sigma = 20$) given.

$$= P\left\{\frac{\sum_{i=1}^{20} X_i - 800}{20\sqrt{20}} > \frac{15}{2\sqrt{5}}\right\}$$

$$= P\left\{\frac{\sum X_i - 800}{20\sqrt{20}} > 3.41 \text{ (approx)}\right\}$$

$\approx 1 - \Phi(3.41)$, as by CLT, $\left(\sum_{i=1}^{n} X_i - n\mu\right)/\sigma\sqrt{n} \to$ a standard normal variate as $n \to \infty$.

$\approx \underline{0.0003}$ (as $\Phi(3.41) = 0.9997$)