Lecture-30

Markov's inequality

Chebyshev's inequality

Weak law of large numbers.

$X_1, X_2, \cdots, X_n$ iid

Each has finite mean

$E[X_i] = \mu$,

for any $\epsilon > 0$

$$P\left\{ \left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq \epsilon \right\} \to 0$$

as $n \to \infty$

Proof: Additional assumption:

All X_i have the same finite variance

$$Var(X_i) = \sigma^2$$

$$Z = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$E[Z] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n} \cdot \mu \cdot n = \mu$$

$$Var(Z) = Var\left(\frac{\Sigma X_i}{n}\right)$$

$$= \frac{1}{n^2} Var(\Sigma X_i)$$

$\parallel$ independent

$$= \frac{1}{n^2} \Sigma Var(X_i) = \frac{\sigma^2}{n}$$

$$E[Z] = \mu$$

$$\text{Var}(Z) = \frac{\sigma^2}{n}$$

Chebyshev's inequality

$$P(|Z - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

$$P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

when $n \to \infty$,

$$\frac{\sigma^2}{n\epsilon^2} \to 0$$

Central Limit Thm.

4

Let $X_1, X_2, \dots, X_n$ be a sequence of i.i.d. random variables, each having mean μ & variance σ^2

$$E[X_i] = \mu, \quad Var(X_i) = \sigma^2$$

$$\frac{X_1 + X_2 \dots + X_n - n\mu}{\sqrt{n}\,\sigma} \longrightarrow N(0,1)$$

std. normal distribution

as $n \to \infty$

e.g.

black hole

Astronomer

$\mu = d$ light years

variance $= 4$ light years $= \sigma^2$

How many measurements does he need ~~in~~ to take in order to be 95% sure that the estimated distance is accurate within ± 0.5 light years.

6

X_1 = reading no. 1
$X_2 =$ 2
$\vdots$ $\vdots$
$X_n =$ n

$\left.\right\}$ distance

X_i are independent, same $\mu = d$,
same $\sigma^2 = 4$

$\cancel{Z_n}$

$$Z_n = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sqrt{n}\,\sigma} \to N(0,1)$$

as $n \to \infty$

$$Z_n = \frac{\sum X_i - nd}{2\sqrt{n}}$$

$$P\left(-0.5 \le \frac{\sum X_i}{n} - d \le 0.5\right) = 0.95$$

$$P\left(\left(-0.5\right)\frac{\sqrt{n}}{2} \le \left(\frac{\sum X_i - nd}{n}\right)\frac{\sqrt{n}}{2} \le \left(0.5\right)\frac{\sqrt{n}}{2}\right) = 0.95$$

$$P\left(-\frac{\sqrt{n}}{4} \leq Z_n \leq \frac{\sqrt{n}}{4}\right) = 0.95$$

$$\Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right)$$

$$2\,\Phi\left(\frac{\sqrt{n}}{4}\right) - 1 = 0.95$$

$$\Phi\left(\frac{\sqrt{n}}{4}\right) = \frac{1+0.95}{2} = 0.975$$

$$\frac{\sqrt{n}}{4} = 1.96$$

$$\sqrt{n} = 4 * 1.96$$

$$n \doteq 4^2 * (1.96)^2$$

$$\simeq 61 \ldots$$

$\lceil n \rceil = 62$ 62 observations

If the astronomer wants to be ~~as~~ more certain, he can use Chebyshev's Inequality. 8

$x_1, x_2 \ldots x_n$: n measurements

$$E\left[\sum_{i=1}^{n} \frac{x_i}{n}\right] = d$$

$$Var\left(\frac{\sum x_i}{n}\right) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$= \frac{4}{n}$$

$$P\left(\left|\frac{\sum x_i}{n} - d\right| \geq 0.5\right) \leq \frac{\sigma^2}{n\epsilon^2}$$ 9

$$\leq \frac{4}{n(0.5)^2}$$

$$\boxed{P\left(\left|\frac{\sum x_i}{n} - d\right| \geq 0.5\right)} \leq = \frac{16}{n}$$

$$\boxed{1 - P(\quad)} = 1 - \frac{16}{n}$$

$$0.95 = 1 - \frac{16}{n}$$

$$\frac{16}{n} = 0.05$$

$$n = \frac{16}{0.05} = 16 * 20$$

$$= \boxed{320}$$

e.g.

no. of students who register in Economics elective is a Poisson random variable with $\mu = 100$

if $P(X \geq 120)$ register, 2 batches.

if $X \leq 119$ register, it'll be one class.

What is the probability that she'll be teaching two batches?

$$P(X \geq 120) = \sum_{i=120}^{\infty} \frac{e^{-100} (100)^i}{i!}$$

Use C.L.T.

$P(X \geq 120)$

continuity correction

$P(X \geq 119.5)$

$$P\left(\frac{X-\mu}{\sigma} \geq \frac{119.5-\mu}{\sigma}\right)$$

$$P\left(Z \geq \frac{19.5}{10}\right)$$

$P(Z \geq 1.95)$

$= 1 - \Phi(1.95)$

$= 0.0256$

e.g.: An instructor has to check 50 copies. Time required to check one copy, on an average, is $\mu = 20$ minutes, $\sigma = 4$ minutes. Compute the probability that the instructor will evaluate at least 25 copies within the first 450 minutes?

X_i = time required to evaluate i^{th} copy. 13

$$P\left(\sum_{i=1}^{25} X_i < 450\right)$$

$X = \sum X_i \qquad P(X \leq 450)$

$E[X] = \sum E[X_i] = 25 * 20 = 500$

$Var(X) = \sum Var(X_i) = 25 * 16 = 400$

Assume that X_i are independent.

Z = std. normal.

$P(X \leq 450) = \cancel{P(}$

$$P\left(\cancel{X=450}\ \frac{X-500}{\sqrt{400}} \leq \frac{450-500}{\sqrt{400}}\right)$$

$$\simeq P(Z \leq -2.5)$$

$$= P(Z \geq 2.5)$$

$$= 1 - \Phi(2.5)$$

$$= 1 - 0.9938$$

$$= 0.0062$$

Need to take continuity correction.

$$P\left(Z \leq \frac{450.5 - 500}{20}\right)$$

$$P(Z \leq -2.475)$$

$$= 1 - \Phi(2.475)$$

$$= 1 - 0.9932$$

$$= \boxed{0.0068}$$