Lecture - 33

Recap:

Chebyshev's inequality

Weak law of large numbers

Central Limit Theorem

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

Then

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma}$$ tends to

standard normal as $n \to \infty$.

e.g.: An astronomer ②
is measuring distance from
his lab to a distant black hole,

$\mu = d$ light ~~ly~~ years.

variance $= 4$ light years.

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

3

$$= \frac{\sum X_i - nd}{\sqrt{n}\,\sigma} \quad \overset{d}{\approx}$$

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

as $n \to \infty$

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

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

$$|||$$

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

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

$$= 0.95$$

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

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

$$\phi\left(\frac{\sqrt{n}}{4}\right) - \phi\left(-\frac{\sqrt{n}}{4}\right) = 0.95$$

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

$$\phi\left(\frac{\sqrt{n}}{4}\right) = \frac{1+0.95}{2} = \frac{1.95}{2} = 0.975$$

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

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

$$n = 16 \times (1.96)^2 = 61. \ldots$$

$\lceil n \rceil = 62$ observations.

Also use Chebyshev's inequality to get a <u>better</u> estimate.

$X_1, \ldots X_n$: n measurements

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

$$d = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \cdot n \cdot d = d$$

$$Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \sum Var\left(\frac{X_i}{n}\right)$$

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

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

6

$$P\left(\left|\frac{\sum x_i}{n} - d\right| > k\right) \leq \frac{\sigma^2}{k^2}$$

$$P\left(\left|\frac{\sum x_i}{n} - d\right| \geq 0.5\right) \leq \frac{4}{n(0.5)^2}$$

$$P(\quad) \leq \frac{4 \cdot 4}{n} = \frac{16}{n}$$

$$1 - P(\,) \geq 1 - \frac{16}{n}$$

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

$$\frac{16}{n} \geq 1 - 0.95 = 0.05$$

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

$$= 320 \text{ observations}$$

e.g.

No. of students who enroll (X) in Economics elective is a Poisson random variable with $\mu = 100$.

$X \geq 120$	$X \leq 119$
2 batches.	One batch

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

Can be computed using a code on a PC

What if you use CLT?

8

$$P(X \geq 120)$$

$$P\left(\frac{X-\mu}{\sigma} \geq \frac{120-\mu}{\sigma}\right)$$

$$P\left(Z \geq \frac{120-100}{10}\right)$$

$$P(Z \geq 2) = 1 - \phi(2)$$

$$\simeq 0.02$$

9

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. (i.i.d)

Compute the probability that the instructor will evaluate at least 25 copies within 450 minutes.

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

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