4a. CLT (IIT-M)

24 April 2024 07:22

Looking at the sample mean and the distribution mean,

$$E(\bar{x}) = \mu, \quad V(\bar{x}) = \frac{\sigma^2}{n}$$

Variance goes to 0 as n becomes larger

But can we say something more precise to quantify the relationship b/w $\bar{x}$ and μ?

WEAK LAW OF LARGE NUMBERS

Since $\bar{x}$ is a random variable, we try to find a relation in terms of probability

$P(|\bar{x}-\mu| > \delta) \longrightarrow$ likelihood $\bar{x}$ deviates from μ by some δ.

Using Chebyshev's inequality,

$$P\left((\bar{x}-\mu)^2 > \delta^2\right) \leq \frac{E\left[(\bar{x}-\mu)^2\right]}{\delta^2}$$

$$P\left(|\bar{x}-\mu| > \delta\right) \leq \frac{E\left[(\bar{x}-\mu)^2\right]}{\delta^2}$$

$$\boxed{P\left(|\bar{x}-\mu| > \delta\right) \leq \underbrace{\frac{\sigma^2}{n\delta^2}}_{\text{upper bound}} \xrightarrow[\text{fixed } \delta]{n \to \infty} 0}$$

- With probability more than $1-\frac{\sigma^2}{n\delta^2}$, $\bar{x} \in [\mu-\delta, \mu+\delta] \longrightarrow$ works for $\delta > \frac{\sigma}{\sqrt{n}}$
- Chebyshev's inequality is a very "weak" bound

MOMENT GENERATING FUNCTION (MGF)

$X \longrightarrow$ random variable with 0 mean

Note: If X has a non-zero mean, simply subtract the mean value from all X

MGF of X, denoted as $M_X(\lambda): \mathbb{R} \to \mathbb{R}$

$$\boxed{M_X(\lambda) = E\left[e^{\lambda X}\right]}$$

- X: Discrete with pmf f_X

  $X \in \{x_1, x_2, \dots, x_n\}$

$$M_X(\lambda) = f_X(x_1)e^{\lambda x_1} + \dots + f_X(x_n)e^{\lambda x_n}$$

$$= \sum_{i=1}^{n} f_X(x_i)e^{\lambda x_i}$$

- X: Continuous with PDF f_X and support T_X

$$M_X(\lambda) = \int_{T_X} f_X(x) e^{\lambda x} dx$$

- $X \sim$ Normal $(0, \sigma^2)$

$$\boxed{M_X(\lambda) = e^{\frac{\lambda^2 \sigma^2}{2}}}$$

IMPORTANT

$$\int_{-\infty}^{\infty} e^{\lambda x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}} dx$$

Why moment generating function?

$$E[e^{\lambda X}] = E\left[1 + \lambda X + \frac{\lambda^2}{2!} X^2 + \frac{\lambda^3}{3!} X^3 + \cdots\right]$$

$$= 1 + \lambda \underbrace{E[X]}_{\text{first moment}} + \frac{\lambda^2}{2!} \underbrace{E[X^2]}_{\text{second moment}} + \frac{\lambda^3}{3!} \underbrace{E[X^3]}_{\text{third moment}} \cdots$$

Eg: Value of the 6^{th} moment of Normal $(0, 3)$.

$$M_X(\lambda) = e^{\frac{\lambda^2 \sigma^2}{2}} = e^{\frac{3\lambda^2}{2}}$$

$$1 + \lambda E[X] + \ldots + \frac{\lambda^6}{6!} E[X^6] \ldots = 1 + \frac{\lambda^2}{2!}\sigma^2 + 0 + \frac{\lambda^4}{4!}(3\sigma^4)$$

$$\cdots + \frac{\lambda^6}{6!}(5\sigma^6)$$

$$\Rightarrow E[X^6] = 5\sigma^6$$

$$= 5(3)^3$$

$$= \underline{135}$$

CENTRAL LIMIT THEOREM → how does this work???

$\underbrace{X_1, X_2, \ldots, X_n}_{\text{samples}} \sim$ iid X with $E(X) = 0$, $V(X) = \sigma^2$

X can be of any distribution

$Y = \dfrac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$ → sample mean but divide by $\sqrt{n}$

$$\boxed{M_Y(\lambda) \longrightarrow e^{\frac{\lambda^2 \sigma^2}{2}} = \text{MGF of Normal } (0, \sigma^2)}$$

Contrasting with Weak Law of Large Numbers

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

$$M_{\bar{X}}(\lambda) \longrightarrow E(X)$$

Using CLT to approximate probability

$X_1, \ldots, X_n \sim X$ (iid samples)

$\mu = E(X)$, $\sigma^2 = V(X)$

$Y = X_1 + \cdots + X_n$; $E(Y) = n\mu$

Let us try to find the probability that Y deviates from its mean by a factor more than some δ

$P(Y - n\mu > \delta n\mu) = ?$

Using CLT, we know that

$$\frac{Y - n\mu}{\sqrt{n}} \approx \text{Normal}(0, \sigma^2)$$

$$\frac{Y - n\mu}{\sqrt{n}\,\sigma} \approx \text{Normal}(0, 1)$$

Let $F_Z(z)$ be CDF of Normal $(0,1)$

$$P(Y - n\mu > \delta n\mu) = P\Bigg(\underbrace{\frac{Y - n\mu}{\sqrt{n}\,\sigma}}_{\text{Normal }(0,1)\,:\,Z} > \frac{\delta\sqrt{n}\,\mu}{\sigma}\Bigg)$$

$$\boxed{P(Y - n\mu > \delta n\mu) \approx 1 - F_Z\left(\frac{\delta\sqrt{n}\,\mu}{\sigma}\right)}$$

QUESTIONS

1) The average life of a coffee machine is 5 years, with a standard deviation of 1 year. Assuming that the lives of these machines follow approximately a normal distribution, find the probability that the mean life of a random sample of 16 such machines falls between 4.6 and 5.4 years. Enter your answer correct to two decimals.

$\mu = 5$ years $= E(\bar{X})$

$\sigma = \sigma^2 = 1$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{16}} = \frac{1}{4} = 0.25$$

$$Z = \frac{X - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$Z_1 = \frac{4.6 - 5}{0.25} = -1.6$$

$$Z_2 = \frac{5.4 - 5}{0.25} = 1.6$$

$$P(-1.6 < Z < 1.6) = 0.9452 - 0.0548$$
$$= 0.8904$$

2) The random variable X representing the number of cherries in a cherry puff has the following probability distribution:

X	4	6	8	10
$P(X = x)$	0.3	0.1	0.4	0.2

Table 7.6.1: PMF of X

Find the probability that the average number of cherries in 64 cherry puffs will be less than 7. Enter your answer correct to two decimals.

$$E(x) = 0.3(4) + 0.1(6) + (0.4)8 + 0.2(10)$$
$$= 7 = \mu_{\bar{x}}$$

$$E(x^2) = 54$$

$$\sigma^2 = 54 - 49 = 5$$

$$\sigma = \sqrt{5}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{5}}{\sqrt{64}} = 0.078125$$

$$P(\bar{x} < 7) = P\left(Z < \frac{7-7}{\ldots}\right)$$

$$= P(Z < 0) = 0.5$$

3) An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 600 hours and a standard deviation of 50 hours. Find the probability that a random sample of 25 bulbs will have an average life of less than 615 hours. Enter your answer correct to two decimals.

$\mu = 600$ hours $= \mu_{\bar{x}}$

$\sigma = 50$ hours

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{5} = 10$$

$$P(\bar{x} < 615) = P\left(Z < \frac{615-600}{10}\right)$$

$$= P\left(Z < \frac{3}{2}\right)$$

$$= P(Z < 1.5)$$

$$= \underline{\underline{0.933}}$$

4) The number of accidents in a certain city is modeled by a Poisson random variable with an average rate of 9 accidents per day. Suppose that the number of accidents on different days are independent. Use the central limit theorem to find the probability that there will be more than 3300 accidents in a certain year. Assume that there are 365 days in a year. Enter your answer correct to two decimals.

$\lambda = 9$ accidents / day

$\mu = 9, \quad \sigma^2 = 9, \quad \sigma = 3$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{365}} = 0.157027$$

For more than 3300 accidents, you need more than 9 accidents a day

$$P(\bar{x} > 9) = 1 - P(\bar{x} \leq 9)$$

$$= 1 - P\left(Z \leq \frac{9-9}{\ldots}\right)$$

$$= 0.5?$$

X

$X \rightarrow$ no of accidents in a year

$\mu = 9$ per day

$= 3285$ per year

$$\sigma = \sqrt{3285} = 57.3149$$

$$P(X > 3300) = 1 - P(X \leq 3300)$$

$$= 1 - P\left(Z \leq \frac{3300 - 3285}{57.3149}\right)$$

$$= 1 - P(Z \leq 0.26)$$

$$= 1 - 0.6026$$

$= \underline{\underline{0.3974}}$