198

~~Module I~~

CHAPTER 5

Tchebycheff's Inequality, Law of Large Numbers and Central Limit Theorem

5.1 TCHEBYCHEFF'S INEQUALITY

If X is any random variable having mean m and finite variance σ^2, then for any $\varepsilon > 0$,

$$P(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Proof: Case I. When X is a continuous random variable

$$P(|X - m| \geq \varepsilon) = \int_{|x-m|\geq\varepsilon} f(x)\,dx \leq \frac{1}{\varepsilon^2}\int_{|x-m|\geq\varepsilon} (x-m)^2 f(x)\,dx \qquad \left[\begin{array}{l}\because\ |x-m|\geq\varepsilon \\ \Rightarrow\ 1 \leq (x-m)^2/\varepsilon^2\end{array}\right]$$

$$\leq \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-m)^2 f(x)\,dx = \frac{\sigma^2}{\varepsilon^2} \qquad (\because \text{ integrand is non-negative})$$

Alternative

$$\sigma^2 = \int_{-\infty}^{\infty}(x-m)^2 f(x)\,dx = \int_{|x-m|\geq\varepsilon}(x-m)^2 f(x)\,dx + \int_{|x-m|<\varepsilon}(x-m)^2 f(x)\,dx$$

$$\geq \int_{|x-m|\geq\varepsilon}(x-m)^2 f(x)\,dx \geq \varepsilon^2\int_{|x-m|\geq\varepsilon} f(x)\,dx$$

$$[\because\ |x-m|\geq\varepsilon \ \Rightarrow\ (x-m)^2 \geq \varepsilon^2]$$

$$\Rightarrow \qquad \sigma^2 \geq \varepsilon^2 P(|X-m|\geq\varepsilon) \ \Rightarrow\ P(|X-m|\geq\varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Case II. When X is a discrete random variable

$$P(|X - m| \geq \varepsilon) = \sum_{|x_i-m|\geq\varepsilon} p_i \leq \frac{1}{\varepsilon^2}\sum_{|x_i-m|\geq\varepsilon}(x_i-m)^2 p_i \qquad \left[\begin{array}{l} p_i = P(X = x_i) \text{ and} \\ \text{since } |x_i - m| \geq \varepsilon \\ \Rightarrow\ 1 \leq (x_i-m)^2/\varepsilon^2\end{array}\right]$$

$$\leq \frac{1}{\varepsilon^2} \sum_{i=-\infty}^{\infty} (x_i - m)^2 p_i = \frac{\sigma^2}{\varepsilon^2}$$

Observation: (*i*) To determine the probability of an event described by a random variable, its distribution or density is required. Tchebycheff's inequality gives a bound for the probability of an event which depends on mean and variance but does not depend on the distribution of the random variable.

(*ii*) $P(|X - m| > \varepsilon) < \sigma^2/\varepsilon^2$.

Notes: (*i*) This inequality was given by the Russian probabilist P.L. Tchebycheff (1821 – 1894) in 1867.

(*ii*) Existence of the variance $\Rightarrow$ existance of the mean.

(*iii*) This inequality brings out the significance of the variance as a measure of dispersion about the mean somewhat quantitatively. It states that, for a given $\varepsilon > 0$, the amount of probability mass outside the interval $(m - \varepsilon, m + \varepsilon)$ is less than or equal to σ^2/ε^2 which is obviously small if the variance is small.

(*iv*) Tchebycheff's inequality can also be written as

(*a*) $P(|X - m| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$, for a given $\varepsilon > 0$,

Or

(*b*) $P(|X - m| \geq \tau\sigma) \leq \frac{1}{\tau^2}$, for a given $\tau > 0$.

(*v*) Let $X \sim N(m, \sigma)$. Using Tchebycheff's inequality, we have

$$P(|X - m| \geq 3\sigma) \leq \frac{1}{9} = 0.111.$$

But $$Z = \frac{X - m}{\sigma} \sim N(0, 1). \text{ So, we have}$$

$$P(|X - m| \geq 3\sigma) = P(|Z| \geq 3) = 0.0026.$$

It indicates that the Tchebycheff's inequality gives a rather poor bound for the probability in question.

ILLUSTRATIVE EXAMPLES – I

Example 1: *A random variable X has mean m = 12 and variance $\sigma^2 = 9$. Prove that*

$$P(6 < X < 18) \geq \frac{3}{4}.$$

Solution: Using Tchebycheff's inequality, we have

$$P(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \text{ for given } \varepsilon > 0.$$

$$\Rightarrow \quad P(|X - m| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

$$\Rightarrow \quad P(m - \varepsilon < X < m + \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

Taking $m = 12$ and $\sigma^2 = 9$, we get

$$P(12 - \varepsilon < X < 12 + \varepsilon) \geq 1 - \frac{9}{\varepsilon^2}$$

Putting $\varepsilon = 6$, we get

$$P(6 < X < 18) \geq 1 - \frac{9}{36}, \quad i.e., P(6 < X < 18) \geq \frac{3}{4} \quad \textbf{(Proved)}$$

Example 2: *Can we find a random variable X for which P (m – 2σ < X < m + 2σ) = 0.7?*

Solution: We have

$$P(m - 2\sigma < X < m + 2\sigma) = P(|X - m| < 2\sigma) \geq 1 - \frac{\sigma^2}{4\sigma^2}$$

(By Tchebycheff's inequality)

$$\Rightarrow \qquad P(m - 2\sigma < X < m + 2\sigma) \geq \frac{3}{4} = 0.75$$

Hence we conclude that there does not exist a random variable X satisfying the given condition.

Example 3: *Using Tchebycheff's inequality, find a lower bound for the probability of getting 64 to 184 driving licences issued by Road Transport Authority in a specific month. It is given that the number of driving licences issued per month be a random variable having mean m = 124 and standard deviation σ = 7.5.*

Solution: By Tchebycheff's inequality, we have $P(|X - m| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}$, *i.e.,* $P(|X - 124| < \varepsilon)$ $\geq 1 - \frac{56.25}{\varepsilon^2}$, for a given $\varepsilon > 0$ [since it is given that $m = 124$ and $\sigma = 7.5$].

$$\therefore \qquad P(124 - \varepsilon < X < 124 + \varepsilon) \geq 1 - \frac{56.25}{\varepsilon^2}$$

Putting $\varepsilon = 60$, we get

$$P(64 < X < 184) \geq 1 - \frac{56.25}{3600} = 0.984375,$$

this is the required lower bound.

Example 4: *The distribution of a random variable X is given by* $P(X = -1) = \frac{1}{8}$, $P(X = 0) = \frac{3}{4}$ $P(X = 1) = \frac{1}{8}$. *Verify Tchebycheff's inequality for the distribution.*

Solution: Here, $m = E(X) = (-1) \times \frac{1}{8} + 0 \times \frac{3}{4} + 1 \times \frac{1}{8} = 0$

and $\sigma^2 = E(X - m)^2 = E(X^2) = (-1)^2 \times \frac{1}{8} + 0^2 \times \frac{3}{4} + 1^2 \times \frac{1}{8} = \frac{1}{4}$.

For a given $\varepsilon > 0$, the Tchebycheff's inequality is

$$P(|X - m| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \; i.e., \quad P(|X| \geq \varepsilon) \leq \frac{1}{4\varepsilon^2}$$

Consider the following two cases.

Case 1. $0 < \varepsilon \leq 1$

$$\begin{aligned} P(|X| \geq \varepsilon) &= P\{(X = -1) \cup (X = 1)\} \\ &= P(X = -1) + P(X = 1) \end{aligned}$$

201

$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \leq \frac{1}{4\varepsilon^2} \qquad \left(\because \; 0 < \varepsilon \leq 1 \;\Rightarrow\; 0 < \varepsilon^2 \leq 1 \;\Rightarrow\; 1 \leq \frac{1}{\varepsilon^2}\right)$$

Case 2. $\varepsilon > 1$

Here X assumes the values $-1, 0, 1$ and so $|X| \geq \varepsilon$ is an impossible event for $\varepsilon > 1$.

$$\therefore \qquad P(|X| \geq \varepsilon) = 0 < \frac{1}{4\varepsilon^2}$$

Hence Tchebycheff's inequality is verified.

Remember: For a discrete random variable X:

$$E(X) = \Sigma x_i P(X = x_i); \; E(X^2) = \Sigma x_i^2 P(X = x_i).$$

Example 5: *A discrete random variable X assumes the values $-1, 0, 1$ with respective probabilities $\frac{1}{8}, \frac{3}{4}, \frac{1}{8}$. Evaluate $P(|X - m| \geq 2\sigma)$ and compare it with the upper bound obtained by Tchebycheff's inequality, where the mean and standard deviation of the random variable X are m and σ respectively.*

Solution: Here, $\qquad m = E(X) = (-1) \times \frac{1}{8} + 0 \times \frac{3}{4} + 1 \times \frac{1}{8} = 0$

and $\qquad \sigma^2 = E(X - m)^2 = E(X^2) = (-1)^2 \times \frac{1}{8} + 0^2 \times \frac{3}{4} + 1^2 \times \frac{1}{8} = \frac{1}{4}.$

$$\begin{aligned} \therefore \qquad P(|X - m| \geq 2\sigma) &= P(|X| \geq 1) \\ &= P\{(X = -1) \cup (X = 1)\} \\ &= P(X = -1) + P(X = 1) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}. \end{aligned}$$

By Tchebycheff's inequality, we have

$$P(|X - m| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4}$$

Here the two values coincide.

Example 6: *If the random variable X represents the sum of the numbers obtained when 2 fair dice are thrown, determine an upper bound for $P(|X - 7| \geq 3)$ and compare it with the exact probability.*

Solution: Let the random variables X_1, X_2 denote the outcomes of the first and second dice respectively.

$$\therefore \qquad E(X_1) = E(X_2) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}$$

and $\qquad E(X_1^2) = E(X_2^2) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6}.$

$$\therefore \qquad \text{Var}(X_1) = \text{Var}(X_2) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

Here the random variables X_1, X_2 are independent and therefore,

$$m = E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = \frac{7}{2} + \frac{7}{2} = 7,$$

$$\sigma^2 = \text{Var}(X) = \text{Var}(X_1 + X_2) = E\{(X_1 + X_2) - m\}^2 = E\left\{\left(X_1 - \frac{7}{2}\right) + \left(X_2 - \frac{7}{2}\right)\right\}^2$$

Place after the Remark of Example 14 in p. ~~207~~ 207.

Example 15. A random sample of size $n=100$ is taken from an infinite population with the mean $\mu=75$ and the variance $\sigma^2=256$. Based on Chebyshev's theorem with what probability can we assert that the value we obtain for $\bar{X}$ will fall between 67 and 83.5.
(WBUT-2014)

Solution Here sample mean

$\bar{X}=\frac{1}{n}(X_1+X_2+\cdots+X_n)$, where $X_1, X_2, \cdots, X_n$ are mutually independent random variables each having the same distribution of the parent random variable X.

$$\therefore E(\bar{X})=\frac{1}{n}E(X_1+X_2+\cdots+X_n)=\frac{1}{n}\{E(X_1)+E(X_2)+\cdots+E(X_n)\}$$

$$=\frac{1}{n}\cdot n\mu=\mu.$$

$$Var(\bar{X})=\frac{1}{n^2}\{Var(X_1)+Var(X_2)+\cdots+Var(X_n)\}$$

$$=\frac{1}{n^2}\cdot n\sigma^2=\frac{\sigma^2}{n}.$$

Using Chebyshev's inequality, we have

$$P(|\bar{X}-\mu|<\epsilon)\geq 1-\frac{\sigma^2}{n^2\epsilon^2}, \text{ for given } \epsilon>0.$$

$$\Rightarrow P(\mu-\epsilon<\bar{X}<\mu+\epsilon)\geq 1-\frac{\sigma^2}{n^2\epsilon^2}$$

Taking $\mu=75$, $\sigma^2=256$ and $n=100$, we get

$$P(75-\epsilon<\bar{X}<75+\epsilon)\geq 1-\frac{256}{100\epsilon^2}.$$

Putting $\epsilon=8.5$, we get

$$P(66.5<\bar{X}<83.5)\geq 0.965$$

Hence the probability that the value for $\bar{X}$ will fall between 67 and 83.5 is at least 0.965.