Chebychev's Inequality

Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r>0$

$$P\left(g(x) \geq r\right) \leq \frac{E(g(x))}{r}$$

Proof.

$$E(g(x)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

$$\geq \int_{\{x:\, g(x) \geq r\}} g(x)\, f_X(x)\, dx \qquad \left\{ g \text{ is non negative} \right.$$

$$\geq r \int_{\{x:\, g(x) \geq r\}} f_X(x)\, dx$$

$$= r\, P\left(g(x) \geq r\right)$$

$$\therefore \quad \frac{E(g(x))}{r} \geq P\left(g(x) \geq r\right)$$

or

$$P\left(g(x) \geq r\right) \leq \frac{E(g(x))}{r}$$

Now if we consider

$$g(x) = \left(\frac{x-\mu}{\sigma}\right)^2$$

where $\mu = E(x)$ and $\sigma^2 = Var(x)$

for convenience, take $\lambda = t^2$

Then

$$P\left(\left(\frac{x-\mu}{\sigma}\right)^2 \geq t^2\right) \leq \frac{1}{t^2} E\left(\frac{x-\mu}{\sigma}\right)^2$$

$$P\left((x-\mu)^2 \geq \sigma^2 t^2\right) \leq \frac{1}{t^2}\frac{1}{\sigma^2} Var(x)$$

$$P\left(|x-\mu| \geq t\sigma\right) \leq \frac{1}{t^2}$$

or

$$P\left(|x-\mu| < t\sigma\right) \geq 1 - \frac{1}{t^2}$$

which gives a universal bound on the deviation $|x-\mu|$ in terms of σ.

for example, taking $t = 2$, we get

$$P(|x-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = 0.25$$

There is atleast 75% chance that a r.v. will be within 2σ of its mean

no matter what distn of X

Q. A random variable X has a mean $\mu = 8$ and variance $\sigma^2 = 9$, and an unknown probability distribution. Then find

a) $P(-4 < X < 20)$

b) $P(|X-8| \geq 6)$

Solⁿ.

a) $P(-4 < X < 20) = P(-4-8 < X-8 < 20-8)$

$= P(-12 < X-8 < 12)$

$= P\left(\frac{-12}{3} < \frac{X-8}{3} < \frac{12}{3}\right)$

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

$\mu = 8$

$\sigma^2 = 9 \Rightarrow \sigma = 3$

$t = 4$

$= P\left(-4 < \frac{X-8}{3} < 4\right)$

$= P\left(\left|\frac{X-8}{3}\right| < 4\right)$

$\geq 1 - \frac{1}{4^2}$

$P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}$

$t = 2$

$= \frac{15}{16}$

$\therefore P(-4 < X < 20) \geq \frac{15}{16}$

$t = 2$

Likewise, (b) $P(|X-8| \geq 6) = 1 - P(|X-8| < 6) \leq \frac{1}{4}$

Q. A random variable X has a mean $\mu = 10$ and a variance $\sigma^2 = 4$. Use Chebyshev's Theorem find

a) $P(|X-10| \geq 3)$ b) $P(|X-10| \geq c) \leq 0.04$ what is c?

Soln.

a) $P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2} = \frac{4}{9}$

Here $\mu = 10$, $\sigma = 2$

$\therefore t\sigma = 3 \Rightarrow t = 3/2 = 1.5$

b) $t\sigma = c \Rightarrow t = \frac{c}{\sigma}$

$\therefore \frac{1}{t^2} = \frac{\sigma^2}{c^2} = 0.04$

$\Rightarrow c^2 = \frac{1}{0.04} \cdot 4$

$c^2 = 25 \times 4$

$c^2 = 100$

$c = \sqrt{100} = \pm 10$

Reject $c = -10$

we get $c = 10$

$\ast$ $P(g(x) \geq \lambda) \leq \frac{E(g(x))}{\lambda}$ for any $\lambda > 0$

Q. Determine The minimum percentage of The houses That should sell for prices between 150,000 and 450,000.

Solⁿ: given That standard deviation of 50,000.

$\mu + t\sigma = 450,000$ — (1)

$\mu - t\sigma = 150,000$ — (2)

The interval in Question must be symmetrical around The mean — The mean needs to be in The middle of The range.

(1) + (2) ⇒ $2\mu = 600,000$

$\mu = 300,000$

(1) − (2) ⇒ $2t\sigma = 300,000$

$t\sigma = 150,000$

$\therefore t = \frac{150,000}{50,000} = 3$

Therefore

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

$$\geq \left(1 - \frac{1}{9}\right)$$

$$= 88.9\%$$

of The observations

∴ Minimum percentage of The houses That should sell for prices between 150,000 & 450,000 is 88.9%.

The number of home runs hit by the leaders in this category for the National League during the 2001 Major League Baseball season. The mean of the data is 37.9 and the standard deviation is 11. Verify that Chebyshev's Theorem holds true for two standard deviations around the mean.

Sorted National League Home Run leaders

73	64	57	49	49	45	41	39	38	38
37	37	37	36	36	34	34	34	34	34
33	31	31	30	30	29	27	27	27	25

Total — 30

$\mu + k\sigma = 37.9 + 2.11$ & $\mu - k\sigma = 37.9 - 2.11$

$= 59.9$ & $= 15.9$

28 observ in range 15.9–59.9

Percentage of data in this interval $\frac{28}{30} = 93.3\%$.

Chebyshev Theorem states that at least 75% of the player's record will fall within two standard deviations of the mean. Therefore, Chebyshev's Theorem holds true in this example.