Frequently Asked Questions

1. What do you mean by Tchebyscheff's inequality?

Answer:

Tchebysheff's inequality is one of the powerful tools that we can use in statistical analysis. The inequality is named after Russian mathematician Pafnuty Chebyshev, although his friend and colleague Irénée-Jules Bienaymé first formulated it.

Chebyshev allows us to estimate the deviation of the random variable from its mean. In Probability Theory **Chebyshev's inequality** (also spelled as **Tchebysheff's inequality**) guarantees that in any probability distribution ,"nearly all" values are close to the mean — the precise statement being that no more than $1/k^2$ of the distribution's values can be more than *k* standard deviations away from the mean.

In this inequality, we remove the restriction that the random variable has to be non-negative. As a price, we now need to know additional information about the variable – (finite) expected value and (finite) variance

2. State and prove Tchebysheff's inequality for a continuous random variable.

Answer:

Statement:

Let X be a random variable for which E(X) and V(X) exists. Then for any positive number k,

$$P\{|x-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

OR

$$P\{|x-\mu| \leq k\sigma\} \geq 1-\frac{1}{k^2}$$

Proof:

Let X be a continuous random variable. Let f(x) be the probability function of x then

$$\sigma^2 = E(X-E(X))^2 = E(X-\mu)^2 = \int_{-\infty}^{\infty}(X-\mu)^2 f(x)dx$$

Range of values of X is divided into three intervals $(-\infty, \mu-k\sigma)$, $(\mu-k\sigma, \mu+k\sigma)$ and $(\mu+k\sigma, \infty)$

Therefore

$$\sigma^2 = \int_{-\infty}^{\mu-k\sigma} (X-\mu)^2 f(x)dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (X-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (X-\mu)^2 f(x)dx$$

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} (X-\mu)^2 f(x)dx + \int_{\mu+k\sigma}^{\infty} (X-\mu)^2 f(x)dx$$

(As the second integral, which is positive, is deleted)

In the case of first integral on the right hand side

$$x \leq \mu - k\sigma \Rightarrow (x-\mu) \leq -k\sigma \Rightarrow -(x-\mu) > k\sigma \Rightarrow (x-\mu)^2 > k^2\sigma^2$$

And in the case of the second integral

$$x \geq \mu + k\sigma \Rightarrow (x-\mu) \geq k\sigma \Rightarrow (x-\mu)^2 > k^2\sigma^2$$

Therefore,

$$\sigma^2 \geq k^2\sigma^2 [\int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx]$$

$$\sigma^2 \geq k^2\sigma^2 [P(-\infty < X \leq \mu - k\sigma) + P(\mu + k\sigma \leq X < \infty)]$$

$$\sigma^2 \geq k^2\sigma^2 [P(-\infty < X - \mu \leq -k\sigma) + P(+k\sigma \leq X - \mu < \infty)$$

$$\Rightarrow 1 \geq k^2 [P(-\infty < X - \mu \leq -k\sigma) + P(+k\sigma \leq X - \mu < \infty)$$

$$\Rightarrow 1 \geq k^2 [P|X-\mu| \geq k\sigma)$$

Therefore,

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

Since $P(A) + P(A^c) = 1$

$P(A^c) = 1 - P(A)$

Then

$$P\{|x-\mu| \le k\sigma\} \ge 1-\frac{1}{k^2}$$

3. Deduce Tchebysheff's inequality for a discrete random variable.

Answer:

Let X be a discrete random variable. Let p(x) be the probability mass function of x then

$$\sigma^2 = E(X-E(X))^2 = E(X-\mu)^2 = \sum_{-\infty}^{\infty}(X-\mu)^2 p(x)$$

Range of values of X is divided into three intervals (-∞, μ-kσ), (μ-kσ, μ+kσ) and (μ+kσ, ∞)

Therefore, $\sigma^2 = \sum_{-\infty}^{\mu-k\sigma}(X-\mu)^2 p(x) + \sum_{\mu-k\sigma}^{\mu+k\sigma}(X-\mu)^2 p(x) + \sum_{\mu+k\sigma}^{\infty}(X-\mu)^2 p(x)$

$$\sigma^2 \ge \sum_{-\infty}^{\mu-k\sigma}(X-\mu)^2 p(x) + \sum_{\mu+k\sigma}^{\infty}(X-\mu)^2 p(x)$$

(As the second summation, which is positive, is deleted)

In the case of first summation on the right hand side

$$x \le \mu-k\sigma \Rightarrow (x-\mu) \le -k\sigma \Rightarrow -(x-\mu) > k\sigma \Rightarrow (x-\mu)^2 > k^2\sigma^2$$

And in the case of the second summation

$$x \ge \mu+k\sigma \Rightarrow (x-\mu) \ge k\sigma \Rightarrow (x-\mu)^2 > k^2\sigma^2$$

Therefore,

$$\sigma^2 \ge k^2\sigma^2[\sum_{-\infty}^{\mu-k\sigma} p(x) + \sum_{\mu+k\sigma}^{\infty} p(x)]$$

$$\sigma^2 \ge k^2\sigma^2[P(-\infty < X \le \mu-k\sigma) + P(\mu+k\sigma \le X < \infty)]$$

$$\sigma^2 \ge k^2\sigma^2[P(-\infty < X-\mu \le -k\sigma) + P(+k\sigma \le X-\mu < \infty)]$$

$$\Rightarrow 1 \ge k^2[P(-\infty < X-\mu \le -k\sigma) + P(+k\sigma \le X-\mu < \infty)]$$

$$\Rightarrow 1 \ge k^2[P|X-\mu| \ge k\sigma)$$

Therefore, $[P|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

4. Write a note on Upper Bound and Lower Bound in Tchebysheff's inequality.

Answer:

The inequality $[P|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$ provides an upper bound for the probability of the difference between a variable and its mean to be more than constant kσ.

Therefore, $1/k^2$ is the upper bound

The inequality

$P\{|x-\mu| \leq k\sigma\} \geq 1-\frac{1}{k^2}$ provides an lower bound for the probability of the difference between a variable and its mean to be less than a constant kσ.

Therefore $1-\frac{1}{k^2}$ is the lower bound

5. Compare bounds obtained by Tchebysheff's inequality with another powerful tool Samuelson's inequality.

Answer:

Although Chebyshev's inequality is the best possible bound for an arbitrary distribution, this is not necessarily true for finite samples. Samuelson's inequality states that all values of a sample will lie within √(n - 1) standard deviations of the mean. Chebyshev's bound improves as the sample size increases.

When n = 10, Samuelson's inequality states that all members of the sample lie within 3 standard deviations of the mean: in contrast Chebyshev's states that 95% of the sample lies within 13.5789 standard deviations of the mean.

When n = 100, Samuelson's inequality states that all members of the sample lie within approximately 9.9499 standard deviations of the mean: Chebyshev's states that 99% of the sample lies within 140.0 standard deviations of the mean.

When n = 500, Samuelson's inequality states that all members of the sample lie within approximately 22.3383 standard deviations of the mean: Chebyshev's states that 99% of the sample lies within 11.1620 standard deviations of the mean.

Tchebysheff's inequality is likely to get better bounds for as the sample size increases.

6. Interpret Tchebysheff's rule for different values of the constant k.

Answer: According to Tchebysheff's rule

At least $1-\frac{1}{k^2}$ of the data values must be within k standard deviation of the mean where k is any value greater than 1

For k =2, 3, and 4, these are what the theorem states:

- k = 2, *at least* 75% of the data values must be within 2 standard deviations of the mean.
- k = 3, *at least* 89% of the data values must be within 3 standard deviations of the mean.
- k = 4, *at least* 94% of the data values must be within 4 standard deviations of the mean.

7. What do you mean by standard scores?

Answer:

- A *standard score* or *z score* is used when direct comparison of raw scores is impossible.
- A standard score or *z* score for a value is obtained by subtracting the mean from the value and dividing the result by the standard deviation.

8. Make a comparative study of the Tchebyscheff's inequality with the Empirical Rule for Normal Distributions

Answer:

Empirical Rule for Normal Distributions

The following apply to a bell-shaped distribution.

- Approximately 68% of the data values fall within one standard deviation of the mean.
- Approximately 95% of the data values fall within two standard deviations of the mean.
- Approximately 99.75% of the data values fall within three standard deviations of the mean.

Chebyshev's Inequality, allows us to extend this idea to any distribution: even if that distribution isn't normal. The theorem states that for a population or sample, the proportion of observations is no less than $(1 - (1 / k^2))$, as long as the standard scores or z score's absolute value is less than or equal to k. You can only use Chebyshev's Theorem to get results for **standard deviations over 1.**

Chebyshev's Inequality, produces the following:

- No information can be obtained on the fraction of values falling within 1 standard deviation of the mean
- At least 75% will fall within 2 standard deviations of the mean

- At least 88.8 % will fall within 3 standard deviations
- The proportion of values from a data set that will fall within *k* standard deviations of the mean will be at least $1 - 1/k^2$; where *k* is a number greater than 1.

9. What are the applications of Tchebyscheff's inequality?

Answer:

There are some cool applications of Tchebyscheff's inequality

- Using Tchebyscheff's inequality, we can get tighter bounds using higher moments without using complex inequalities.
- Inequality is very much used to estimate confidence interval.
- Using Tchebyscheff's inequality, we can prove that the median is at most one standard deviation away from the mean.
- Tchebyscheff's inequality also provides the simplest proof for weak law of large numbers.

Hence, clever application of the inequality provides very useful bounds without knowing anything about the distribution of the random variable. Chebyshev's inequality, on the other hand, bounds the probability that a random variable deviates from its expected value by any multiple of its standard deviation. Chebyshev does not expect the variable to non-negative but needs additional information to provide a tighter bound. Chebyshev inequality is tight – This means with the information provided, the inequalities provide the most information they can provide.

10. What are the points to be observed in Tchebyscheff's inequality as compared to other inequalities?

Answer:

There are few interesting things to observe in Tchebyscheff's inequality:

- In contrast to other inequalities, Chebyshev inequality allows you to bound the deviation on both sides of the mean.
- The length of the deviation is $k\sigma$ on both sides, which is usually (but not always) tighter than the bound k E[X].
- Intuitively, if the variance of X is small, then Chebyshev inequality tells us that X is close to its expected value with high probability.
- Using Chebyshev inequality, we can claim that at most one fourth of the values that X can take is beyond 2 standard deviation of the mean.

11. Suppose we randomly select a magazine article from a source with an average of 1000 words per article, with a standard deviation of 200 words. What is the probability that it has between 400 and 1600 words?

Answer:

Here $\mu=1000$ and $\sigma=200$ and

$\mu-k\sigma=400$ then $1000-k\,(200)=400$ which implies $k=3$

Similarly $\mu+k\sigma=1600$ also gives $k=3$

By Tchebyscheff's inequality $[P|X-\mu| \leq k\sigma) \geq 1-\frac{1}{k^2}$

$$\Rightarrow [P|X-1000| \leq 3(200)) \geq 1-\frac{1}{9} = 0.89$$

Hence, there is less than $1/_{k2}$ = 1/9 chance to be outside that range, by Chebyshev's inequality. Alternatively, we can then infer that the probability that it has between 400 and 1600 words (i.e. within *k* = 3 SDs of the mean) must be more than 89%.

12. X is a random variable with mean 10 and variance 9. Find a lower bound to P{Ix – 10I <9}.

Answer:

Solution:

By Tchebycheff's inequality

$$P\{|x-\mu| \leq k\sigma\} \geq 1-\frac{1}{k^2}$$

Given μ=10 and σ^2 =9

Therefore

$$P\{|x-10| \leq k.3\} \geq 1-\frac{1}{k^2}$$

Puttting 3k=9, k=3

$$1-\frac{1}{k^2} = 1-\frac{1}{9} = \frac{8}{9}$$

The lower bound of the probability $= 1-\frac{1}{k^2} = \frac{8}{9}$ =0.89

13. If X is a random variable with mean =6 and variance =2 find t such that P{Ix – 6I <t}≥0.96

Solution:

Given μ=6 and σ^2 =2

Therefore,

By Tchebycheff's inequality

$$P\{|x-\mu| \le k\sigma\} \ge 1-\frac{1}{k^2}$$

Substituting

$$P\{|x-6| \le k.\sqrt{2}\} \ge 1-\frac{1}{k^2}$$ ----------------------(1)

Put $1-\frac{1}{k^2}$=0.96

Then $\frac{1}{k^2}$ =1-0.96=0.04 which implies k =5

Substituting in (1)

$$P\{|x-6| \le 5.\sqrt{2}\} \ge 0.96$$

But given that

$$P\{|x-6| < t.\} \ge 0.96$$

Therefore t= 5 √2 = 7.07

14. Let X be a Random variable taking values -1,0,1 with probabilities 1/8,6/8 and 1/8 respectively. Find using Tchebyscheff's inequality the upper bound of the probability P{IxI≥1}. Compare this with the actual probability

Answer:

Here X takes values -1, 0, and 1 with probabilities 1/8, 6/8 and 1/8 respectively

Then E(X)= ∑xp(x) = 0

V(X) = E(X^2) – [E(X)]^2 = ¼ -0 = ¼

Therefore,

Given μ=0 and σ^2 =1/4

Applying Tchebycheff's inequality $[P|X-\mu| \ge k\sigma) \le \frac{1}{k^2}$

Put μ=0 and σ=1/2 then $[P|X-0| \geq k\frac{1}{2}) \leq \frac{1}{k^2} \Rightarrow [P|X| \geq \frac{k}{2}) \leq \frac{1}{k^2}$

Put k/2 = 1 then k=2 Therefore $1/k^2 = 1/4$

Therefore $[P|X| \geq 1) \leq \frac{1}{4}$

The upper bound =1/4

To find the actual probability $|X| \geq 1$ means x ≥1 or x≤ -1

But X takes value only -1,0, 1. Therefore $|X| \geq 1$ means x =1 and x=-1

Therefore $P(|X| \geq 1) = P[x=1,-1] = P[x=1] + P[X=-1] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$

The upper bound is ¼ and the actual probability is also ¼

15. For the number X obtained when a die is thrown prove that Tchebyscheff's inequality gives $[P|X-3.5| > 2.5) < 0.47$. Compare this with the actual probability.

Answer:

p(x) = ½ for x = 1,2,3,..,6

Then E(X)= ∑xp(x) = (1X1/6)+(2X1/6)+……+(6X1/6)=3.5

V(X) = $E(X^2) - [E(X)]^2 = 15.167 - (3.5)^2 = 2.917$

Σ=√2.917=1.71

Applying Tchebycheff's inequality $[P|X-\mu| > k\sigma) < \frac{1}{k^2}$

$$[P|X-3.5| > k(1.71) < \frac{1}{k^2}$$

Take k. (1.71) = 2.5 Then k = 2.5/1.71= 1.46

Therefore $1/k^2 = 0.47$

Therefore $[P|X - 3.5| > 2.5) < 0.47$

To find the actual probability $[P|X - 3.5| > 2.5) = 1 - P|X - 3.5| \leq 2.5$

$$= 1 - P[-2.5 + 3.5 \leq X \leq 2.5 + 3.5] = 1 - P[1 \leq X \leq 6] = 1 - [\frac{1}{6} + \ldots + \frac{1}{6}]$$

$$= 1 - \frac{6}{6} = 0$$