1. Introduction

Welcome to the series of E-learning modules on Tchebychev's inequality, proof and applications. In this module, we are going to cover Tchebychev's inequality, its proof and applications, relationship with the empherical rule of normal distribution and sample examples to apply the inequality.

By the end of this session, you will be able:

- ExplainTchebychev's inequality
- Derive the inequality
- Explain the applications of the inequality
- Apply Tchebychev's inequality to the practical problems

The Theorem is named after Russian mathematician Pafnuty Chebyshev, although his friend and colleague Irenee-Jules Bienayme first formulated it. The theorem was first stated without proof by Bienayme in eighteen fifty three and later without proof by Chebyshev in eighteen seventy-four. His student Andrey Markov provided a proof in eighteen eighty-four in his PhD thesis.

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 by k square of the distribution's values can be more than k standard deviations away from the mean.

The inequality has great utility because it can be applied to completely arbitrary distributions (unknown except for mean and variance). For example, it can be used to prove the weak law of large numbers.
Tchebychev's inequality is usually stated for random variables, but can be generalized to a statement about the measure spaces.

Chebyshev's inequality is another powerful tool that we can use in statistical analysis. 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.

In contrast to Markov's inequality, Chebyshev allows you to estimate the deviation of the random variable from its mean. A common use of these inequalities is that it estimates the probability of the deviation from its mean in terms of its standard deviation.
Similar to Markov inequality, we can state two variants of Chebyshev. Let us look at the simplest version.

2. Tchebychev’s Inequality-Statement

Tchebychev’s inequality-statement
Let X be a Random variable for which Expected value of X and Variance of X exists. Then, for any positive number k,
Probability of modulus of (X minus mu) greater than or equal to k sigma is less than or equal to 1 by k square
Or
Probability of modulus of (X minus mu) less than or equal to k sigma is greater than or equal to 1 minus 1 by k square

Proof:
Case 1:
Let X be a continuous random variable. Let f of x be the probability function of x, then
Sigma square is equal to Expected value of X minus E of X whole square which is equal to Expected value of X minus mu whole square, which is equal to integral from minus infinity to plus infinity (X minus mu) whole square f of x dx
Range of values of X is divided into three intervals (minus infinity, mu minus k sigma), (mu minus k sigma, mu plus k sigma) and (mu plus k sigma, infinity)
Therefore,
Sigma square is equal to integral from minus infinity to mu minus k sigma (X minus mu) whole square f of x dx plus integral from mu minus k sigma to mu plus k sigma (X minus mu) whole square f of x dx plus integral from mu plus k sigma to infinity (X minus mu) whole square f of x dx

Sigma square is greater than or equal to integral from minus infinity to mu minus k sigma (X minus mu) whole square f of x dx plus integral from mu plus k sigma to infinity (X minus mu) whole square f of x dx
(as the second integral which is positive is deleted)
In the case of first integral on the right hand side
X less than or equal to mu minus k sigma implies (x minus mu) is less than or equal to minus k sigma which implies minus of (x minus mu) greater than k sigma implies (x minus mu) whole square is greater than k square sigma square
And in the case of the second integral
X greater than or equal to mu plus k sigma implies (x minus mu) is greater than or equal to k sigma which implies (x minus mu) whole square is greater than k square sigma square

Therefore,
Sigma square is greater than or equal to k square sigma square into [integral from minus infinity to mu minus k sigma f of x dx plus integral from mu plus k sigma to infinity f of x dx]
Sigma square is greater than or equal to k square sigma square into [Probability of (minus infinity less than X less than or equal to mu minus k sigma) plus Probability of (mu plus k sigma less than or equal to X less than infinity)]
Sigma square is greater than or equal to k square sigma square into [Probability of (minus

infinity less than X minus mu less than or equal to minus k sigma) plus Probability of (k sigma less than or equal to X minus mu less than infinity)]
Which implies one greater than or equal to k square into [Probability of (minus infinity less than X minus mu less than or equal to minus k sigma) plus Probability of (k sigma less than or equal to X minus mu less than infinity)]
Which implies one greater than or equal to k square into [Probability of modulus of (X minus mu) greater than or equal to k sigma]

Therefore,
Probability of modulus of (X minus mu) greater than or equal to k sigma is less than or equal to 1 by k square
Since P of A plus P of A complement is equal to 1
P of A complement is equal to 1 minus P of A
Then,
Probability of modulus of (X minus mu) less than or equal to k sigma is greater than or equal to 1 minus 1 by k square

Case 2:
When X is a discrete random variable
In the case of a discrete random variable, the proof is same as that of continuous random variable, but with the difference that the integration has to be replaced by summation.

Note:
Only the case *k* greater than or equal to 1 provides useful information (when *k* is less than 1, the right-hand side is greater than one, so the inequality becomes vacuous as the probability of any event cannot be greater than one). As an example, using *k* is equal to root 2 shows that at least half of the values lie in the interval (mu minus root two sigma, mu plus root two sigma).

Since it can be applied to complete arbitrary distributions (unknown except for mean and variance), the inequality generally gives a poor bound compared to what might be possible if something is known about the distribution involved.

3. Upper Bound and Lower Bound

Upper Bound and Lower Bound:
The inequality Probability of modulus of (X minus mu) greater than or equal to k sigma is less than or equal to 1 by k square provides an upper bound for the probability of the difference between a variable and its mean to be more than a constant k sigma. Therefore, 1 by k square is the upper bound.
The inequality Probability of modulus of (X minus mu) less than or equal to k sigma is greater than or equal to 1 minus 1 by k square provides a lower bound for the probability of the difference between a variable and its mean to be less than a constant k sigma. Therefore, 1 minus 1 by k square is the lower bound.

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

When n is equal to ten, Samuelson's inequality states that all members of the sample lie within 3 standard deviations of the mean. In contrast, Chebyshev states that ninety five percent of the sample lies within Thirteen point five seven eight nine standard deviations of the mean.
When n is equal to hundred, Samuelson's inequality states that all members of the sample lie within approximately nine point nine four nine nine standard deviations of the mean. Chebyshev states that ninety nine percent of the sample lies within one forty standard deviations of the mean.

When n is equal to five hundred, Samuelson's inequality states that all members of the sample lie within approximately twenty two point three three eight three standard deviations of the mean. Chebyshev states that ninety nine percent of the sample lies within eleven point one six two zero standard deviations of the mean.
It is likely to get better bounds for finite samples as 'n' increases.

Chebyshev's Theorem enables us to state that a proportion of data values must be within a specified number of standard deviation of the mean. The advantage of this theorem is that the theorem applies to any data set regardless of the shape of the distribution of the data.
Hence here is the Chebyshev's rule:
At least one minus 1 by k square of the data values must be within k standard deviation of the mean, where k is any value greater than 1.

For k is equal to 2, 3, and 4:
- k is equal to 2, *at least* **seventy five percent** of the data values must be within 2 standard deviations of the mean
- k is equal to 3, *at least* **eighty nine percent** of the data values must be within 3 standard deviations of the mean
- k is equal to 4, *at least* **ninety four percent** of the data values must be within 4 standard deviations of the mean

4. Empirical Rule for Normal Distributions and Applications of Chebyshev's Inequality

Empirical rule for normal distributions
The following points apply for a bell-shaped distribution:

- Approximately sixty eight percent of the data values fall within one standard deviation of the mean
- Approximately ninety five percent of the data values fall within two standard deviations of the mean
- Approximately ninety nine point seven five percent of the data values fall within three standard deviations of the mean

Standard scores

- 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.

Chebyshev’s Inequality allows us to extend this idea to any distribution even if that distribution is not normal. The theorem states that for a population or sample, the proportion of observations is no less than (1 minus 1 by k square), 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 sometimes called Chebyshev’s Theorem have the following useful rules:

- No information can be obtained on the fraction of values falling within 1 standard deviation of the mean
- At least seventy five percent will fall within 2 standard deviations of the mean
- At least eighty eight point eight percent 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 one minus 1 by k square, where k is a number greater than 1
- This theorem applies to any distribution regardless of its shape

There are few interesting things to observe here:

- In contrast to Markov inequality, 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 into Expected value of X. Similarly, the fraction 1 by k square is much more tighter than 1 by k that we get from Markov inequality.
- 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 deviations of the mean

Applications of Chebyshev Inequality
The following are some of the applications of Tchebycheff's inequality:

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

5. Examples

Examples:

1) Suppose we randomly select a journal article from a source with an average of one thousand words per article with a standard deviation of two hundred words. We can then infer that the probability that it has between six hundred and one thousand four hundred words (i.e. within *k* is equal to 2 Standard deviations of the mean) must be more than seventy-five percent.

Solution:
Here mu is equal to one thousand, sigma is equal to two hundred and k is equal to 2
By Tchebycheff's inequality,
Probability of modulus of (X minus mu) less than or equal to k sigma is greater than or equal to 1 minus 1 by k square
Which implies Probability of modulus of (X minus one thousand) less than or equal to 2 into two hundred is greater than or equal to 1 minus 1 by 4 which is equal to zero point seven five.
Hence, there is less than 1 by k square, which is equal to 1 by 4 chance to be outside that range by Chebyshev's inequality. On the other hand, we can then infer that the probability that it has between six hundred and one thousand and four hundred words (i.e. within *k* is equal to 2 Standard deviations of the mean) must be more than seventy-five percent.

2) X is a random variable with mean 8 and variance 4. Find a lower bound to Probability of modulus of x minus eight less than four.
Solution:
By Tchebycheff's inequality
Probability of modulus of (X minus mu) less than or equal to k sigma is greater than or equal to 1 minus 1 by k square
Given mu is equal to 8 and sigma square is equal to 4

Therefore, Probability of modulus of (X minus 8) less than or equal to k into 2 is greater than or equal to 1 minus 1 by k square
Putting 2k is equal to 4, we get k is equal to 2
Then, 1 minus 1 by k square is equal to 1 minus 1 by 4, which is equal to 3 by 4
The lower bound of the probability is equal to 1 minus 1 by k square is equal to 3 by 4, which is equal to point seven five

3) If X is a random variable with mean is equal to 3 and variance is equal to 2. Find t such that Probability of modulus of (x minus 3) less than t is greater than or equal to zero point nine nine
Solution:
Given mu is equal to 3 and sigma square is equal to 2
By Tchebycheff's inequality,
Probability of modulus of (X minus mu) less than or equal to k into sigma is greater than or equal to 1 minus 1 by k square
Substituting

Probability of modulus of (X minus 3) less than or equal to k into root 2 is greater than or equal to 1 minus 1 by k square. Call this as (1)

Put 1 minus 1 by k square is equal to zero point nine nine
Then, 1 by k square is equal to 1 minus zero point nine nine, which is equal to zero point zero one, which implies k is equal to ten
Substituting in (1)
Probability of modulus of (X minus 3) less than or equal to ten into root 2 is greater than or equal to zero point nine nine
But given that
Probability of modulus of (X minus 3) less than t is greater than or equal to zero point nine nine
Therefore, t is equal to ten into root 2, which is equal to fourteen point one four

Here's a summary of our learning in this session, where we understood:

- The concept of Tchebycheff's inequality
- The proof of Tchebycheff's inequality or theorem
- The applications of Tchebycheff's inequality
- The comparison with empirical rule of Normality
- How to apply the Tchebycheff's inequality and to get the lower and upper bounds