



MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS ENGINEERS

Chebyshev's Inequality

Dr. Mamatha H.R

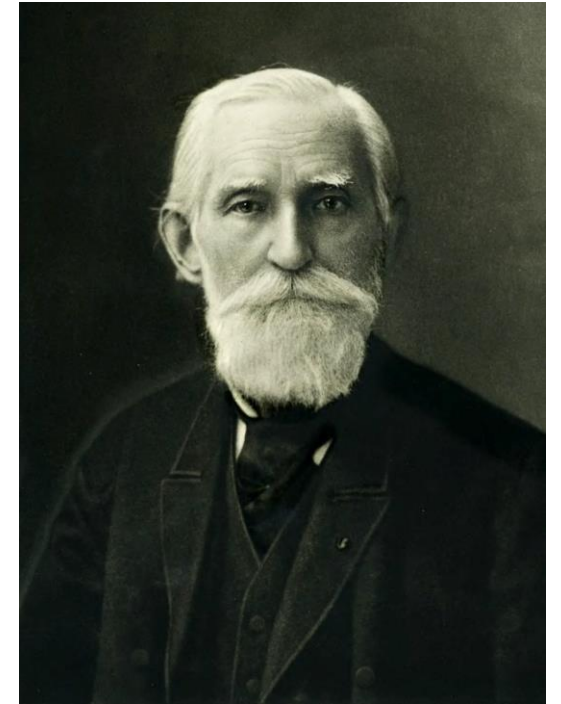
Department of Computer Science and Engineering

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Chebyshev's Inequality



- The mean of a random variable is a measure of the centre of its distribution, and the standard deviation is a measure of the spread.
- *Chebyshev's inequality relates the mean and the standard deviation by providing a bound on the probability that a random variable takes on a value that differs from its mean by more than a given multiple of its standard deviation.*
- Specifically, the probability that a random variable differs from its mean by k standard deviations or more is never greater than $1/k^2$.



Theorem by Russian mathematician Pafnuty Chebyshev

Statement of Chebyshev's Inequality

Chebyshev's inequality states that at least $1 - 1/K^2$ of data from a sample must fall within K standard deviations from the mean, where K is any positive real number greater than one.

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

Chebyshev's Inequality is used to describe the percentage of values in a distribution within an interval centered at the mean.

Formula

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

X = random variable

μ = expected value

σ = standard deviation

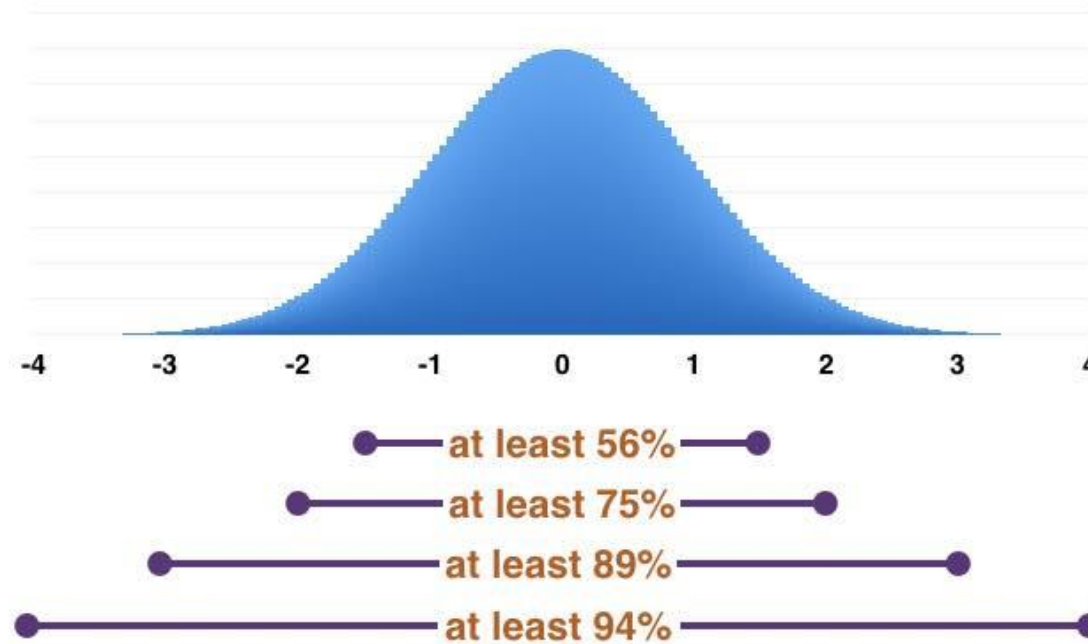
k = number of standard deviations from the mean

Only the case $k > 1$ is useful. When $k \leq 1$ the right hand $1/k^2 \geq 1$ and the inequality is trivial as all probabilities are ≤ 1 .

Chebyshev's Inequality

The proportion of observations within k standard deviations is at least $1 - 1/k^2$

No. Std Devs	$1 - 1/k^2$
1.5	0.56
2.0	0.75
3.0	0.89
4.0	0.94



Problem 1: The length of a rivet manufactured by a certain process has mean $\mu_X = 50$ mm and standard deviation $\sigma_X = 0.45$ mm. What is the largest possible value for the probability that the length of the rivet is outside the interval 49.1–50.9 mm?

Solution

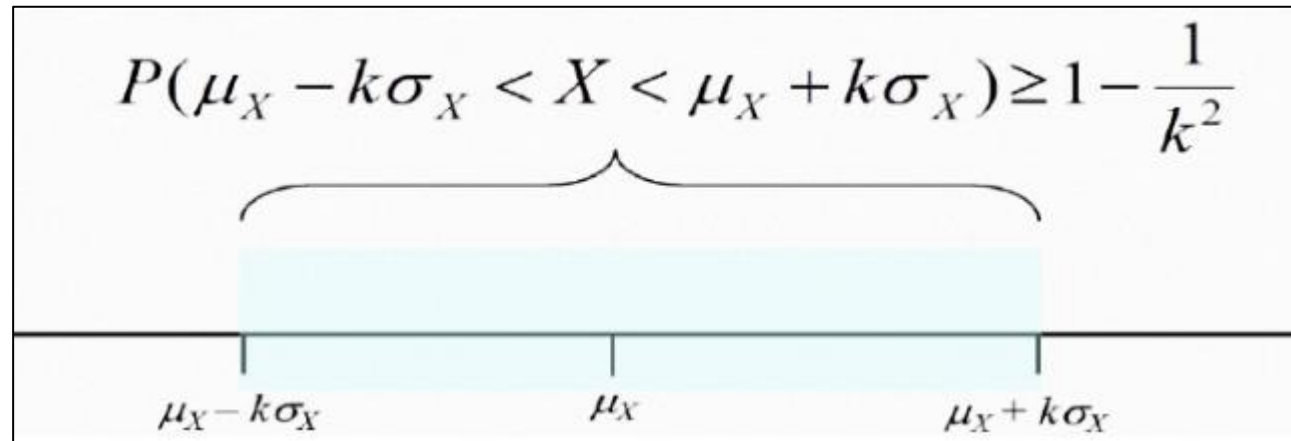
Let X denote the length of a randomly sampled rivet. We must find $P(X \leq 49.1 \text{ or } X \geq 50.9)$. Now

$$P(X \leq 49.1 \text{ or } X \geq 50.9) = P(|X - 50| \geq 0.9) = P(|X - \mu_X| \geq 2\sigma_X)$$

Applying Chebyshev's inequality with $k = 2$, we conclude that

$$P(X \leq 49.1 \text{ or } X \geq 50.9) \leq \frac{1}{4}$$

Chebyshev's inequality states that at least $1-1/K^2$ of data from a sample must fall within K standard deviations from the mean, where K is any positive real number greater than one.



Problem-2:

Assume that the probability density function for X , the length of a rivet in Example 2.46, is

$$f_X(x) = \begin{cases} [477 - 471(x - 50)^2]/640 & 49 \leq x \leq 51 \\ 0 & \text{otherwise} \end{cases}$$

It can be verified that $\mu_X = 50$ and $\sigma_X = 0.45$. Compute the probability that the length of the rivet is outside the interval 49.1–50.9 mm. How close is this probability to the Chebyshev bound of $1/4$?

Solution

$$\begin{aligned}P(X \leq 49.1 \text{ or } X \geq 50.9) &= 1 - P(49.1 < X < 50.9) \\&= 1 - \int_{49.1}^{50.9} \frac{477 - 471(x - 50)^2}{640} dx \\&= 1 - \left. \frac{477x - 157(x - 50)^3}{640} \right|_{49.1}^{50.9} \\&= 0.01610\end{aligned}$$

The actual probability is much smaller than the Chebyshev bound of $1/4$.

- Because the Chebyshev bound is generally much larger than the actual probability, it should only be used **when the distribution of the random variable is unknown.**
- When the distribution is known, then the probability density function or probability mass function should be used to compute probabilities.

- To illustrate the inequality, we will look at it for a few values of K:
 - For $K = 2$ we have $1 - 1/K^2 = 1 - 1/4 = 3/4 = 75\%$.
Chebyshev's inequality says that at least 75% of the data values of any distribution must be within two standard deviations of the mean.
 - For $K = 3$ we have $1 - 1/K^2 = 1 - 1/9 = 8/9 = 89\%$.
So Chebyshev's inequality says that at least 89% of the data values of any distribution must be within three standard deviations of the mean.

Problem 1

Computers from a particular company are found to last on average for three years without any hardware malfunction, with standard deviation of two months. At least what percent of the computers last between 31 months and 41 months?

Use Chebyshev's inequality.

- Mean $\mu = 36$ months, standard deviation $\sigma = 2$ months.
- Interval: 31 to 41 months \rightarrow distance from mean = 5 months.
- Hence $k = \frac{5}{2} = 2.5$.

Chebyshev:

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{(2.5)^2} = 1 - \frac{1}{6.25} = 0.84.$$

At least 84% of the computers last between 31 and 41 months.

Problem 2

What is the smallest number of standard deviations from the mean that we must go if we want to ensure that we have at least 50% of the data of a distribution?

Use Chebyshev's inequality:

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

Require at least 50% inside:

$$1 - \frac{1}{k^2} \geq 0.5 \Rightarrow \frac{1}{k^2} \leq 0.5 \Rightarrow k^2 \geq 2 \Rightarrow k \geq \sqrt{2} \approx 1.414.$$

Smallest number: $\boxed{\sqrt{2} \text{ standard deviations } (\approx 1.414\sigma)}$.

MCQ

1. According to Chebyshev's inequality, the proportion of data within **k standard deviations** of the mean is at least:
 - A) $1/k$
 - B) $1 - 1/k^2$
 - C) $1 - k$
 - D) $1 + 1/k^2$

2. Using Chebyshev's inequality, what proportion of data lies within 3 standard deviations of the mean?
 - A) At least 66.7%
 - B) At least 75%
 - C) At least 88.9%
 - D) At least 95%

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

MCQ

1. Option B
2. Option C



Do It Yourself !!!

The length of a metal pin manufactured by a certain process has mean 50 mm and standard deviation 0.45mm.

What is the largest possible value for the probability that the length of the metal pin is outside the interval $[49.1, 50.9]$ mm?



THANK YOU

Dr. Mamatha H.R

Department of Computer Science and Engineering