భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Biostatistics
BT2023

Lecture 8
Shape distribution, Skewness, Kurtosis

Himanshu Joshi
1 September 2023

Start up notes

- Students should be made to think, to doubt, to communicate, to question, to l*earn from their mistakes,* and most importantly have fun in their learning.
- The best way I have found, always ready to adjust, to change, to unlearn.
- Try to make full of the resources available around you.
- Learn to use internet
- Role of social media in current age of information.
- Don't run out of energy

Measure of dispersion

15, 20, 17, 19, 21, 13, 12, 10, 17, 9, 12

Range: subtract two farthest data points

Interquartile range : Divide your data into four parts, the range between first and third quartile is you IQR.

Mean Deviation or average deviation: The arithmetical mean of the mode of all the deviations from the central values,

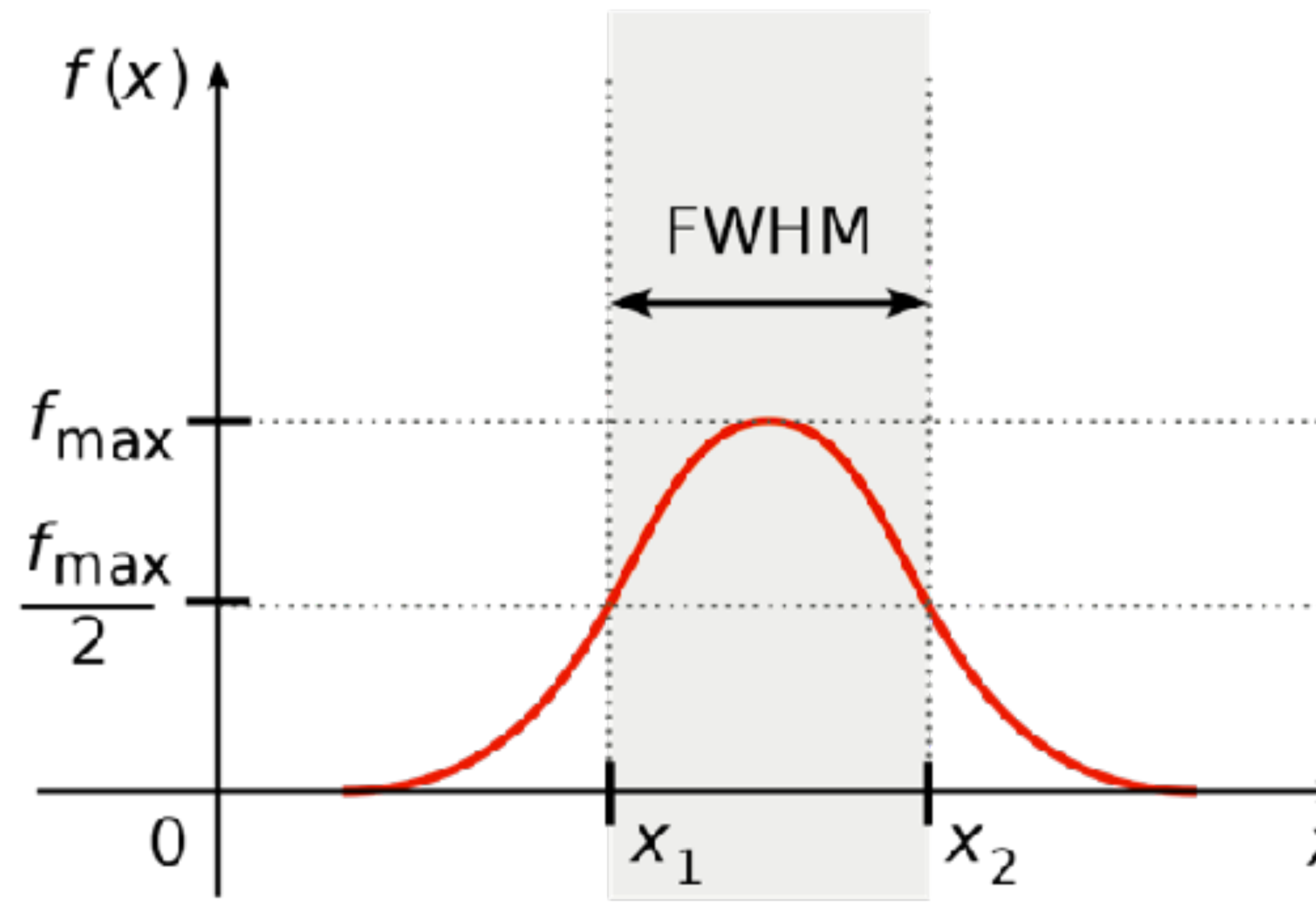
$$\frac{1}{n}\sum|x_i - \bar{x}| \qquad \frac{1}{n}\sum|x_i - M_d \qquad \frac{1}{n}\sum|x_i - Z|$$

From mean   From median   From mod

Coefficient of mean deviation $\dfrac{MD_x}{\bar{x}}$

The relative measure of dispersion corresponding to the mean deviation is called coefficient of mean deviation, be it the Mean, Mode or Median

Full width half maximum



Measurement of dispersion

Standard deviation

$$\sigma = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

Coefficient of standard deviation $\frac{\sigma}{\bar{x}}$

Variance σ^2

15, 20, 17, 19, 21, 13, 12, 10, 17, 9, 12

Standard deviation

Derivations ?

$$\sigma = \sqrt{\frac{\sum_n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_n x_i^2 - n\bar{x}^2}{n}}$$

n used for Population

$$s = \sqrt{\frac{\sum_n {x_i}^2 - n\bar{x}^2}{n-1}}$$

n-1 used for Sample

Variance = σ^2

Standard error in Mean

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

Z-score

$$Z_{score} = \frac{x - \bar{x}}{\sigma}$$

To find if the data points is to the outliers

Z_score > 3 are called outlier

Handy method of computing standard deviation

Y = X - A

Mean (X) = mean(Y) + A

$$\sigma = \sqrt{\frac{\sum {d_i}^2}{n} - \left(\frac{\sum d}{n}\right)^2}$$

48, 43, 65, 57, 31, 60, 37, 48, 59, 78 $\sigma = 13.26$

Algebraic identity

$$\sum_{i=1}^{N} (x_i - \bar{x})^2 = \sum_{i=1}^{N} x_i^2 - n\bar{x}^2$$

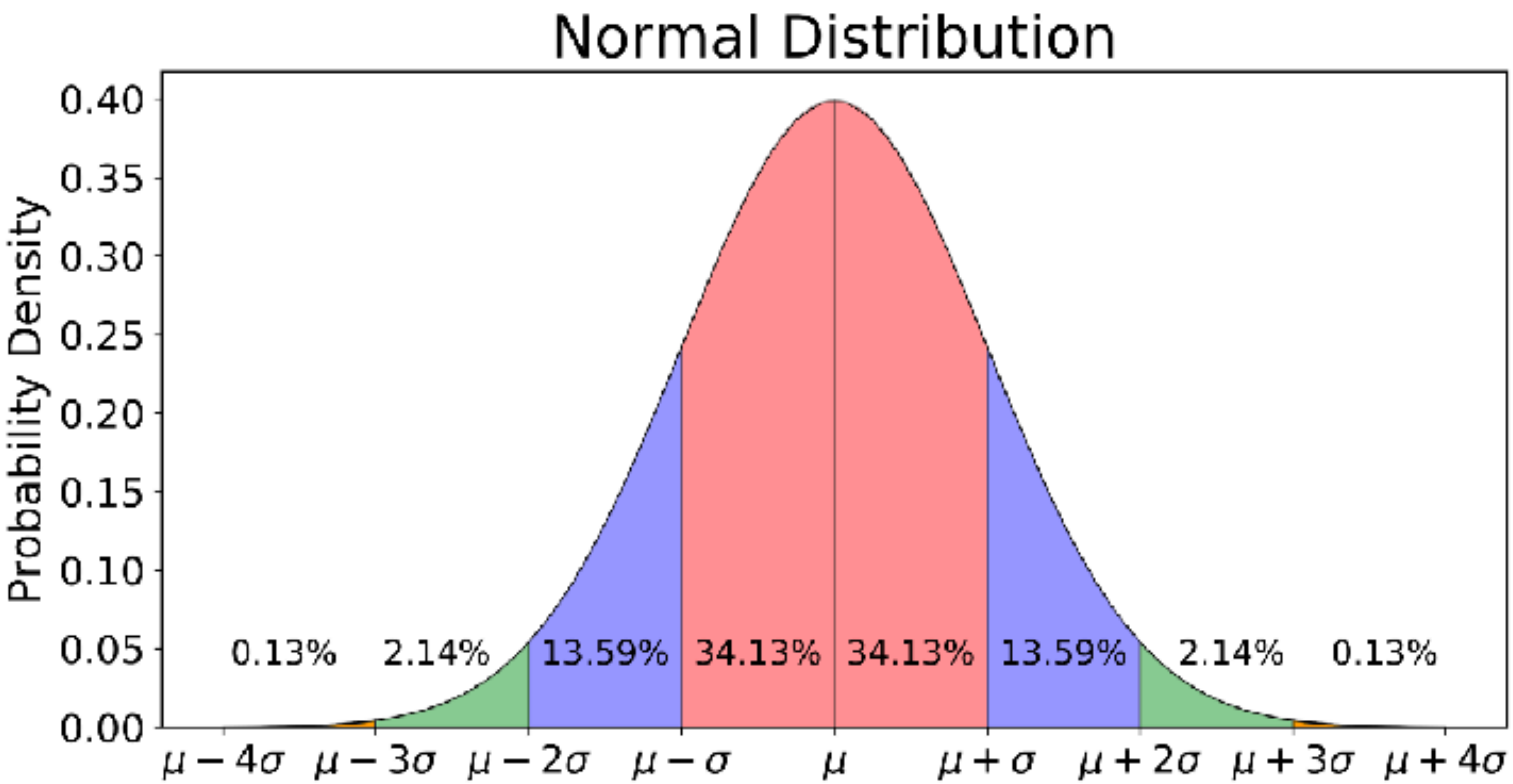
Expected value and standard deviation

For a random variable X having a density f(x)

Limits of Variability

68–95–99.7 rule

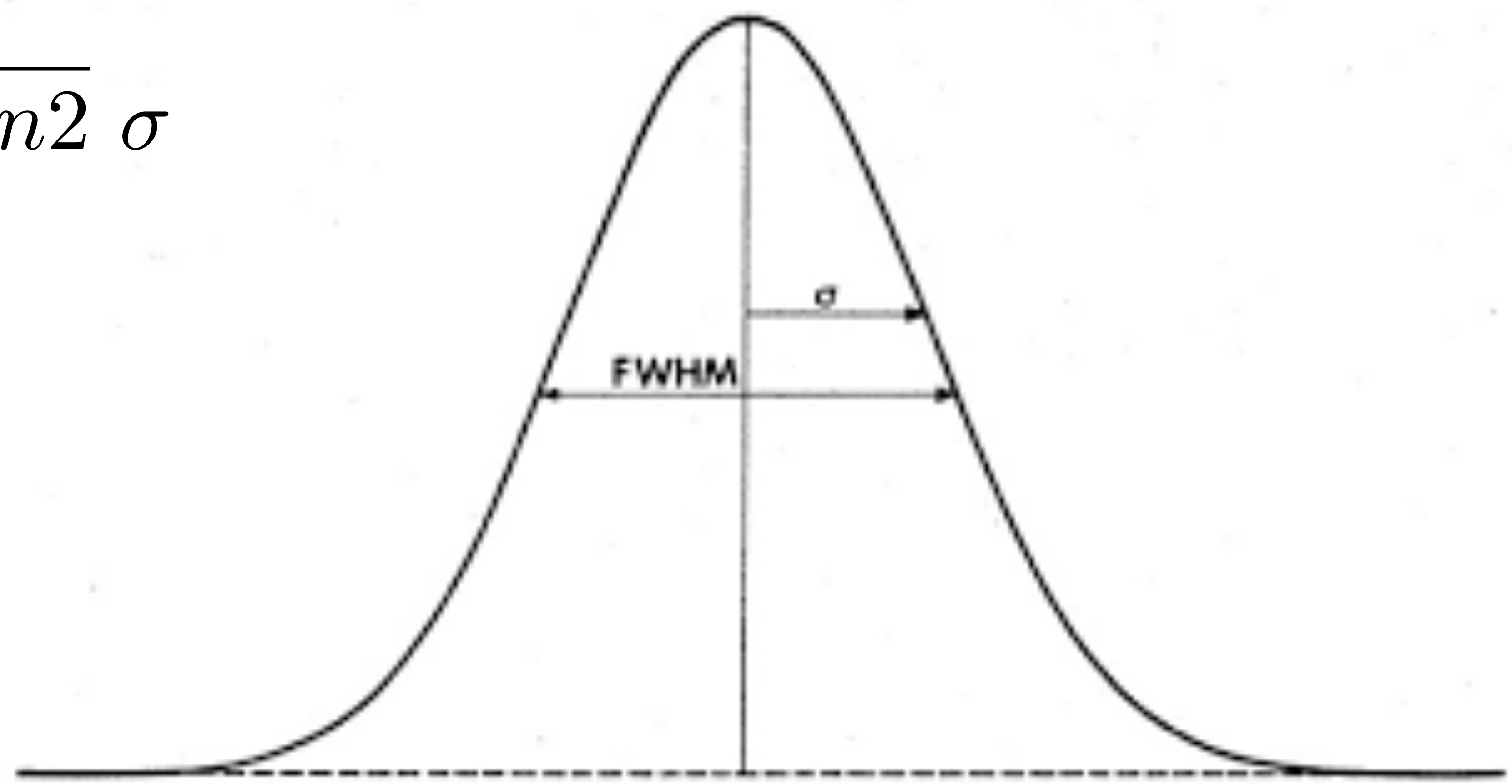
If the population of the interest is normally distributed the standard deviation tells about the proportion of the observation above or below entrain values



Normal or symmetrical distribution

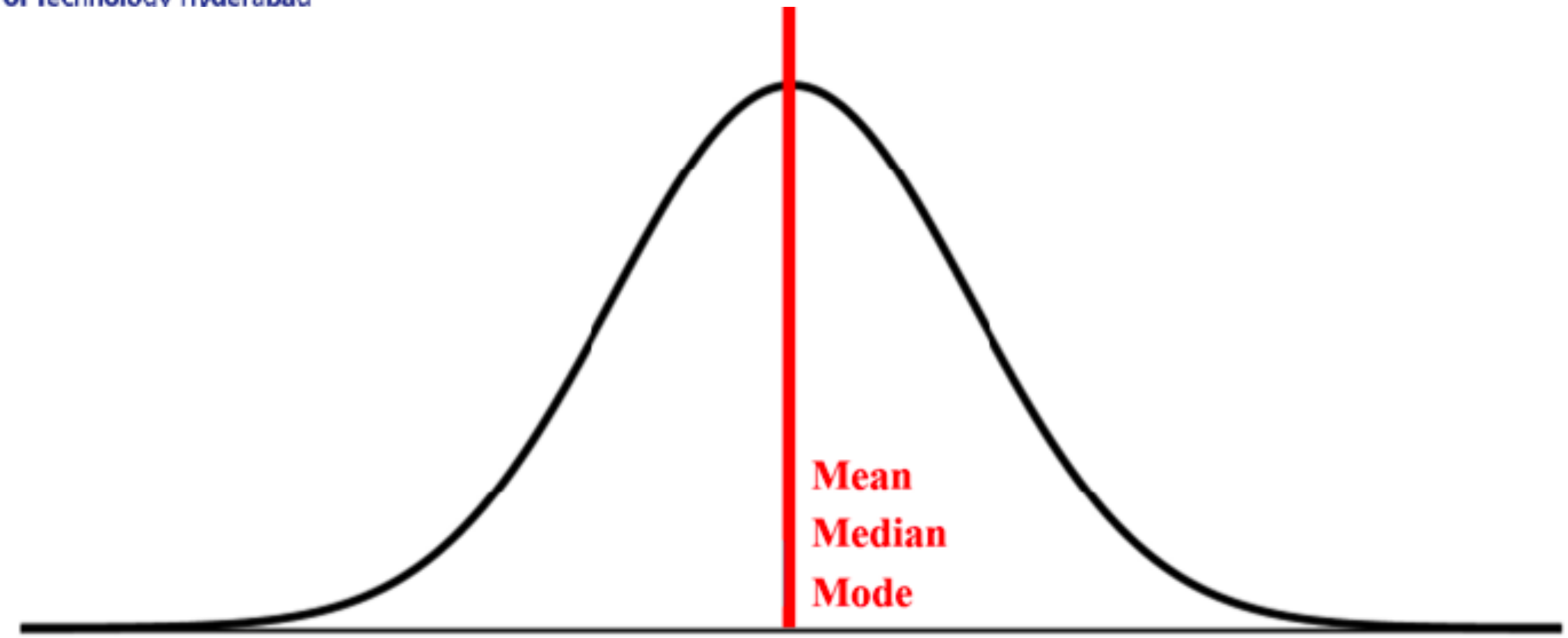
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp - \left[\frac{(x - x_0)^2}{2\sigma^2}\right]$$

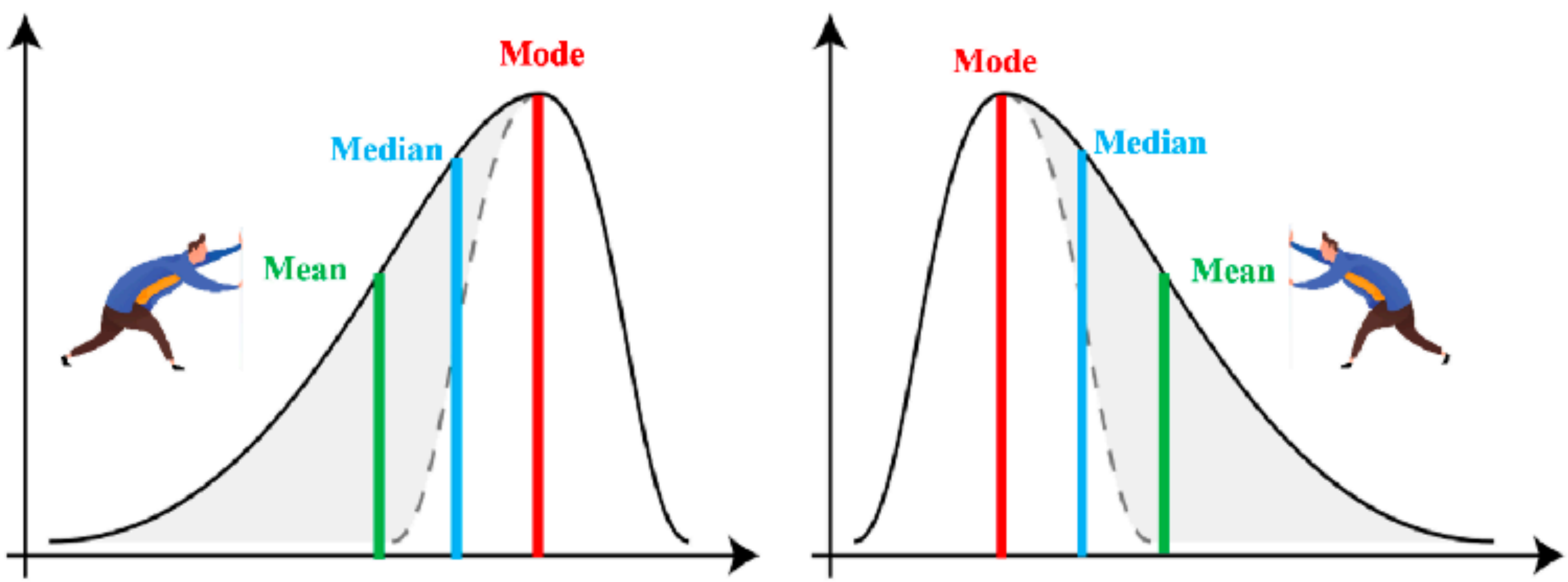
$$FWHM = 2\sqrt{2ln2}\ \sigma$$



Plotting this function

Skewness





Reference
[Towards Data Science]

Measures of Skewness

Absolute skewness Mean - Mode

Karl Pearson coefficient of skewness

$$S_k = \frac{Mean - Mode}{Stanadrd\ deviation} = \frac{3(Mean - Median)}{Standard\ deviation}$$

Bowley coefficient of skewness $= \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

Kelly's coefficient of skewness $\dfrac{P_{10} + P_{90} - 2Median}{P_{90} - P_{10}}$

Skewness coefficient based on moments

Moments in mechanics is refers to the rotating effect of a force, in statistics it is used to describe the peculiarities in the frequency distribution

	Individual series	Discrete series
$First\ moment\ :\mu_1$ **Always 0**	$$\frac{\sum_N (X-\bar{X})}{N}$$	$$\frac{\sum_N f(X-\bar{X})}{N}$$
$Second\ moment\ :\mu_2$ **Measure the variance**	$$\frac{\sum_N (X-\bar{X})^2}{N}$$	$$\frac{\sum_N f(X-\bar{X})^2}{N}$$
$Third\ moment\ :\mu_3$ **Measure skewness**	$$\frac{\sum_N (X-\bar{X})^3}{N}$$	$$\frac{\sum_N f(X-\bar{X})^3}{N}$$
$Forth\ moment\ :\mu_4$ **Measure Kurtosis**	$$\frac{\sum_N (X-\bar{X})^4}{N}$$	$$\frac{\sum_N f(X-\bar{X})^4}{N}$$

Skewness

Skewness coffcient $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$

adjusted Fisher-Pearson coefficient $\frac{\sqrt{N(N-1)}}{N} \frac{\mu_3^2}{\mu_2^3}$

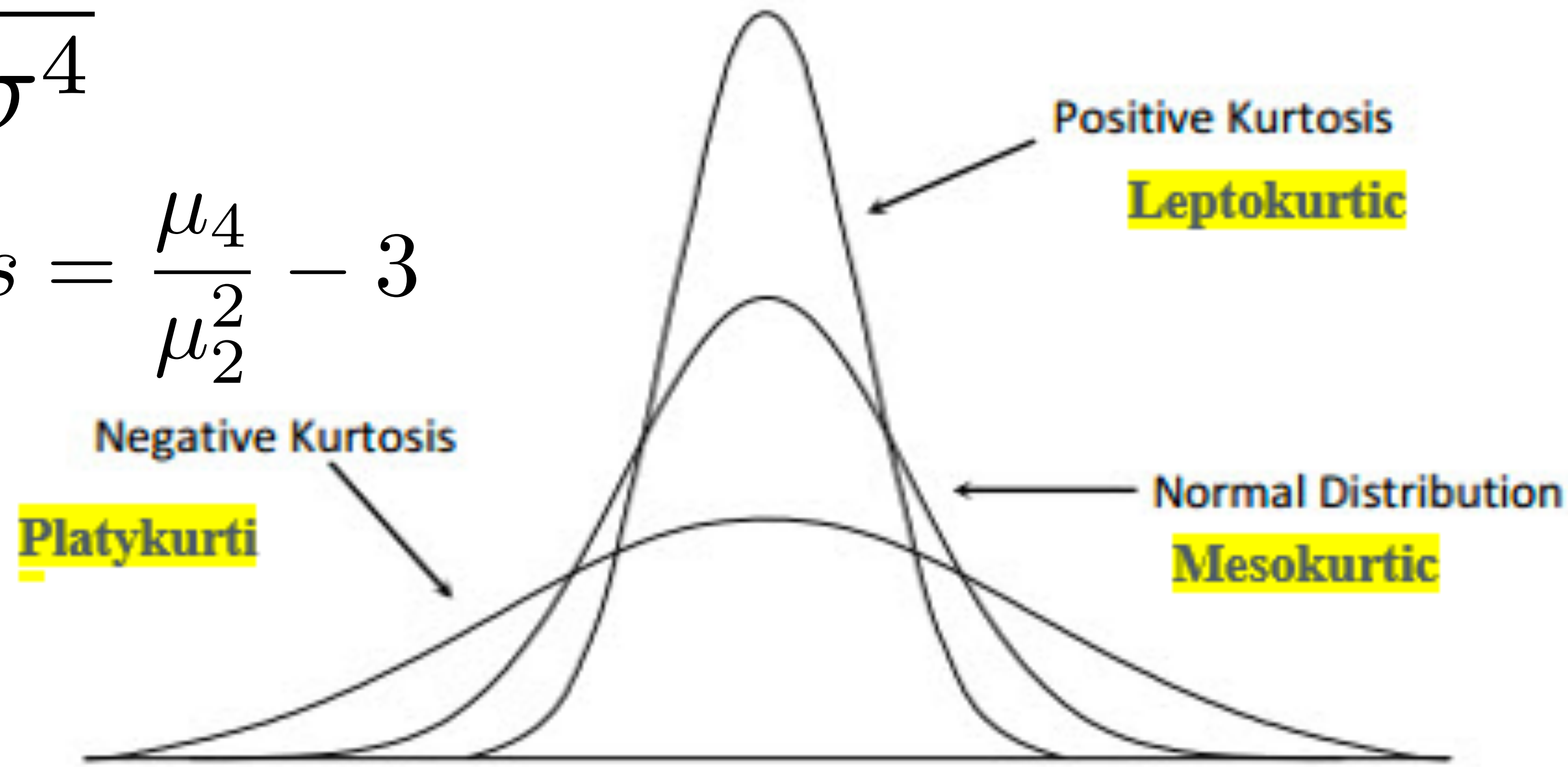
The pre factor approaches to 1 as N tends to large values

Kurtosis

Measure of “tailedness” or peakedness

$$K|X| = \frac{\mu_4}{\sigma^4}$$

$$Excess\ Kurtosis = \frac{\mu_4}{\mu_2^2} - 3$$



Positive kurtosis indicates a "heavy-tailed" distribution >> Leptokutic
Negative kurtosis indicates a "light tailed" distribution >> Platykurtic

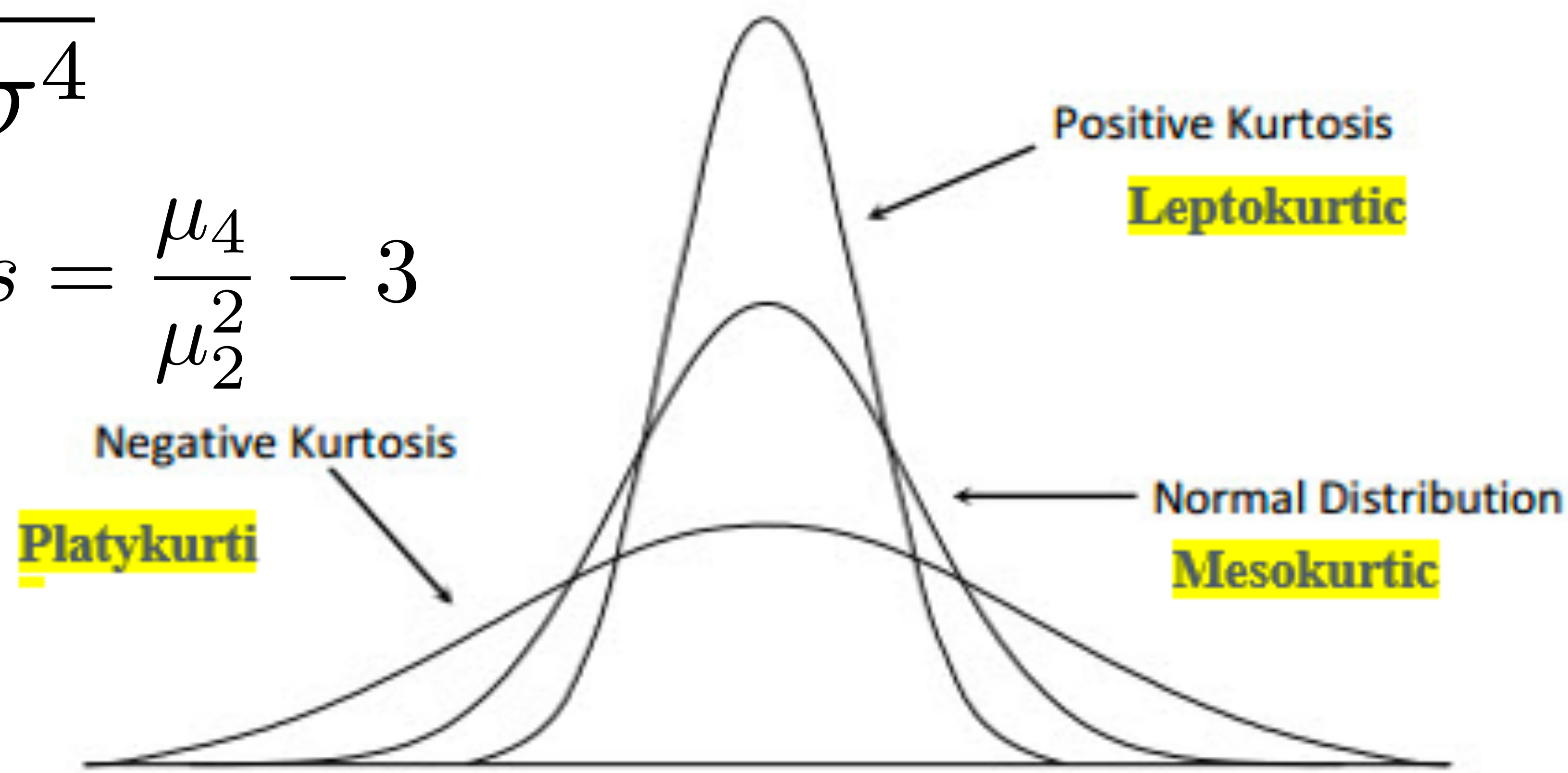
I want you to think about these coefficients, look the formula closely, see if you can relate to a mathematical formula you have seen before

Kurtosis

Measure of “tailedness” or peakedness

$$K|X| = \frac{\mu_4}{\sigma^4}$$

$$Excess\ Kurtosis = \frac{\mu_4}{\mu_2^2} - 3$$



Positive kurtosis indicates a "heavy-tailed" distribution >> Leptokutic
Negative kurtosis indicates a "light tailed" distribution >> Platykurtic

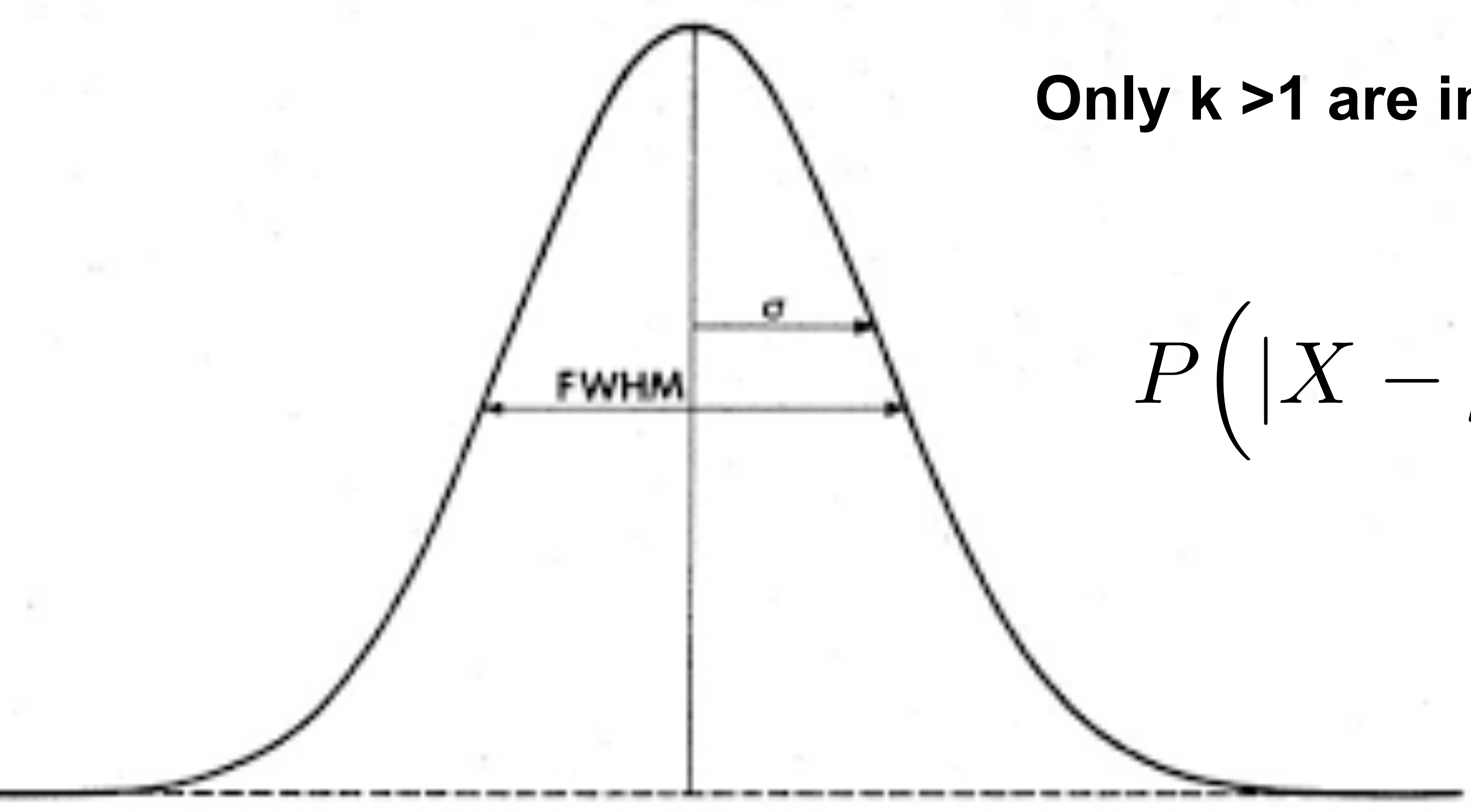
I want you to think about these coefficients, look the formula closely, see if you can relate to a mathematical formula you have seen before

Measure of dispersion

Chebyshev's inequality

The rule is often known as Chebyshev's theorem, tells about the range of standard deviations around the mean, in statistics.In a probability distribution, no more than a certain fraction of values can be more than a certain distance from the mean.

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



Only k >1 are interesting because for k<1 it trivial

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

Homework reading

- Chebyshev inequality
- Gamma Function
- Box plot and violin plot

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Next Class

2:30 PM Tuesday, 5 September 2023