Data Science

Statistics:

Moments of a Distribution

Satishkumar L. Varma

Professor, Department of Information Technology
PCE, New Panvel
Scopus | Web of Science | Google Scholar | Google Site | Website

Moments of a Distribution: Outline

- Moment: Measure of central location
- Moment: Measure of dispersion
- Moment: Measure of asymmetry
- Moment: Measure of peakedness

"Well planned milestone is a key to success."

Moments of a Distribution: Learning Objectives & Outcomes

- **Learning Objectives:** Course Instructor or Faculty aims
  - Statistics — Moments of a distribution
  - Moments in statistics are popularly used to describe the characteristic of a distribution.

Moments of a Distribution: Introduction

- Statistics — Moments of a distribution
- Moments in statistics are popularly used to describe the characteristic of a distribution.

Moments of a Distribution: Introduction

- First moment- Mean
- Measure the location of the central point.

$$\overline{x} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Moments of a Distribution: Second Moment

- Second moment- Standard Deviation (SD, σ(Sigma)):
- Measure the spread of values in the distribution OR how far from the normal.

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N}$$

Moments of a Distribution: Second Moment

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N}$$

- Small SD : Numbers are close to mean
- High SD : Numbers are spread out
- **For normal distribution:**
  - Within 1 SD: 68.27% values lie
  - Within 2 SD: 95.45% values lie
  - Within 3 SD: 99.73% values lie
- **Advantages over Mean Absolute Deviation(MAD):**
- 1. Mathematical properties- Continuous, differentiable.
- 2. SD of a sample is more consistent estimate for a population
  - When drawing repeated samples from a normally distributed population, the standard deviations of samples are less spread out as compare to mean absolute deviations.

Moments of a Distribution: Third Moment

- Third moment- Skewness
- Measure the symmetry in the distribution.

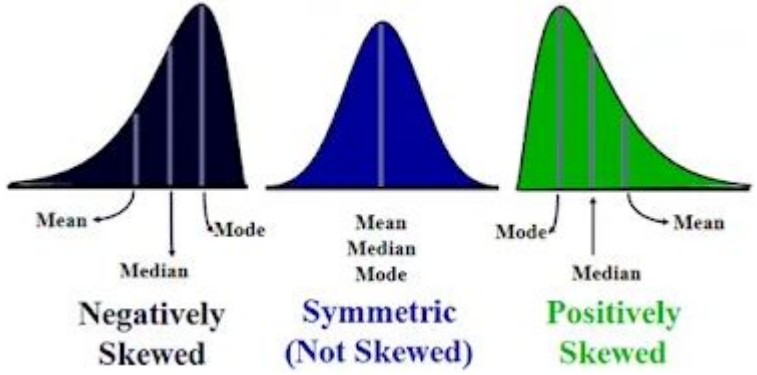
$$Skew = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^3$$

Moments of a Distribution: Third Moment

- Third moment- Skewness: Measure the symmetry in the distribution.
  - Skewness=0 [Normal Distribution, Symmetric]

$$Skew = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^3$$

- **Other Formulas:**
  - 1. Skewness = (Mean-Mode)/SD
  - 2. Skewness = 3*(Mean-Median)/SD
  - (Mode = 3*Median-2*Mean)
- **Transformations** (to make the distribution normal):
- a. Positively skewed (right): Square root, log, inverse
- b. Negatively skewed (left) : Reflect and square[sqrt(constant-x)],
- reflect and log, reflect and inverse



Moments of a Distribution: Fourth Moment

- Fourth moment- Kurtosis:
- Measure the amount in the tails.

$$Kurt = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^4$$

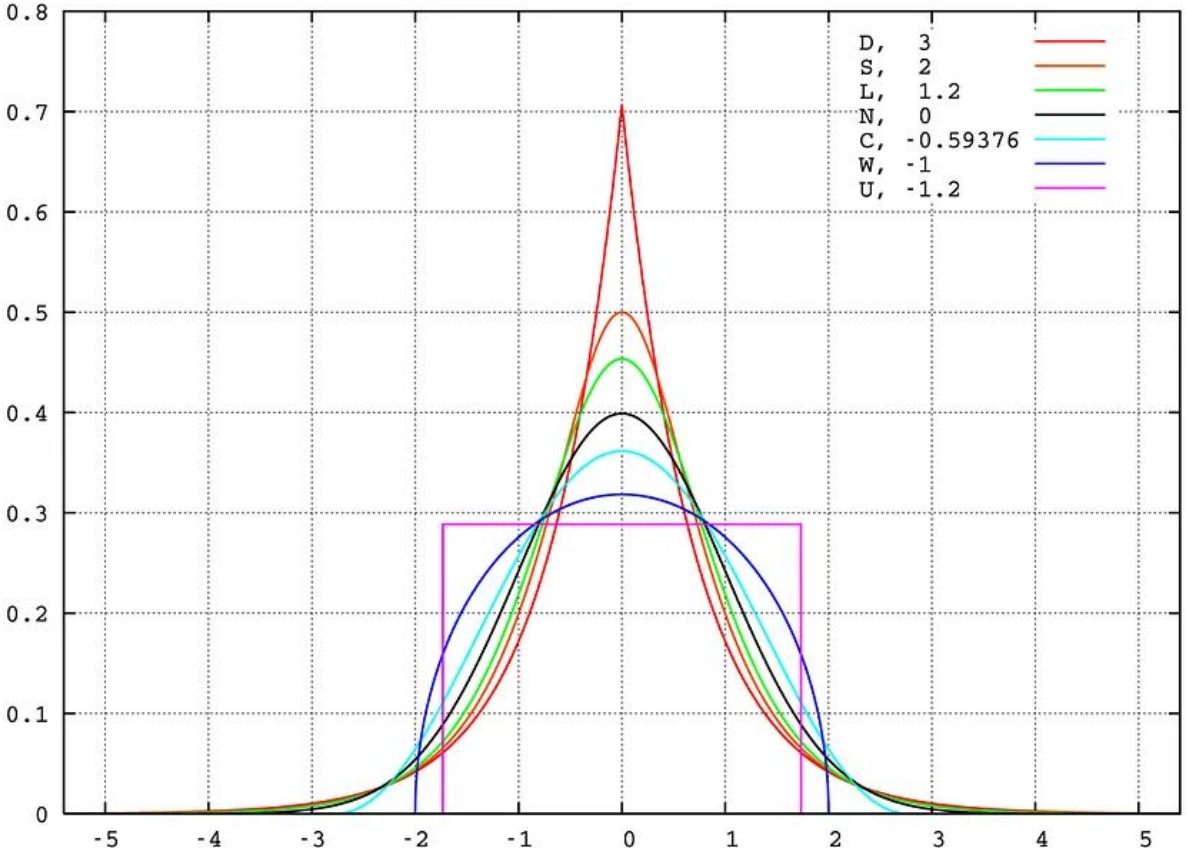
Moments of a Distribution: Fourth Moment

- Kurtosis=3 [**Normal Distribution**]
- Kurtosis<3 [Lighter tails]
- Kurtosis>3 [Heavier tails]
- **Other Formulas:**
  - Excess Kurtosis = Kurtosis - 3
- **Understanding:**
  - Kurtosis is the average of the standardized data raised to fourth power.
  - Any standardized values less than |1| (i.e. data within one standard deviation of the mean) will contribute petty to kurtosis.
  - The standardized values that will contribute immensely are the outliers.
  - High Kurtosis alerts about attendance of outliers.

$$Kurt = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(X_i - \bar{X})}{\sigma}\right]^4$$

Moments of a Distribution: Fourth Moment

- Excess Kurtosis for Distributions [Laplace (D)ouble exponential; Hyperbolic (S)ecant; (L)ogistic; (N)ormal; ©osine; (W)igner semicircle; (U)niform]



Summary

- Statistical moments can be introduced as features of (the probability distribution of) a random variable (RV).
- Interpretation of Moment Statistics
- Mean (M1) - 1st moment about the origin - central tendency measure.
- Variance (m2) - 2nd moment about the mean - dispersion measure.
- Skewness (a3) - 3rd standardize moment - skewness measure.

References

A. Text Books:

1. Hand-book on STATISTICAL DISTRIBUTIONS for experimentalists by Christian Walck, Particle Physics Group, Fysikum, University of Stockholm, 2007.
2. Probability and Statistics: The Science of Uncertainty by Michael J. Evans and Jeffrey S. Rosenthal, University of Toronto.
3. Moments and Their Applications in Ordered Statistics by Faizan Mohammad, 2012.

B. References:

4. SD and variance: https://www.mathsisfun.com/data/standard-deviation.html
5. Advantages of the mean deviation: http://www.leeds.ac.uk/educol/documents/00003759.htm

Thank You.