Week 02

1) Expected Value (mean)

*) The expected value is the long-run average outcome of a random variable if we repeated an experiment infinitely many times

*) Theoritical Mean = Expected Value

For Discrete =>

$$\boxed{E[x] = \mu = \sum_i x_i p_i}$$

x_i = values
p_i = probabilities

ex:- Toss a fair dice

$$E[x] = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$

*) 3.5 not an actual value, but the expected outcome over many trials

For Continuos =>

$$\boxed{E[x] = \mu = \int_{-\infty}^{\infty} x f(x)\, dx}$$

ex:- For a uniform distribution $[0,1]$

$$E[x] = \int_0^1 x \cdot 1\, dx = \frac{1}{2}$$

*) Mean / Balancing point
*) Defined for discrete and continuos variables
*) Weighted average of the PMF / PDF
*) Very high because of outliers

2] Central Tendency : Median and Mode

*) Central tendency is a statistical concept that describes the center point or typical value of a dataset or a probability distribution

Median

*) The median is the middle value when data is sorted
*) If n is odd → middle element
*) If n is even → average of the two middle elements
*) less affected by outliers (better for skewed data)

Mode

*) The mode is the value that occours more frequently
*) A distribution can be :
   .) Unimodal → 1mode
   .) Bimodal → 2 peaks
   .) Multimodal → more than 2 peaks

*) Works well for categorical data (fav color, product preference)

3) Expected value of a function

$$X \Rightarrow g(X)$$

variable function

Discrete case $\Rightarrow$

X takes values x_i, probabilities p_i
then

$$\boxed{E[g(X)] = \Sigma g(x_i) \cdot p_i}$$

example: Dice roll (1-6)

- mean of X = $E[X] = 3.5$
- $g[x] = x^2$

$$E[x^2] = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2)$$

$$= 15.17$$

$\Rightarrow$ Expected square of the roll is 15.17

Continuous case $\Rightarrow$

If X has PDF $f(x)$

$$\boxed{E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx}$$

ex: Uniform $(0,1)$, $g[x] = x^2$

$$E[x^2] = \int_0^1 x^2 (1)\, d(x) = \frac{1}{3}$$

In general:

$$E[aX+b] = aE[X]+b$$

$$E[aX] = aE[X]$$

$$E[b] = b$$

4] Sum of expectations

$$E[X_1+X_2+\dots+X_n] = E[X_1]+E[X_2]+\dots+E[X_n]$$

*) This works always, regardless of whether X and Y are independent or not

5] Variance

*) Variance measures how spread out the values of a random variable are around the mean.

- Mean → tell us the center of data
- Variance → tell us how far the data is spread from the mean

*) If data points are close to the mean (low variance)

*) If data points are far away (high variance)

For a random variable X with mean $\mu = E(X)$

$$Var(X) = E[(X-\mu)^2] = E[X^2]-(E[X])^2$$

1] Find X's mean
2] Find the deviation from that mean for every value of X
3] Square those ~~[illegible]~~ deviations
4] Average those squared deviations

"Average squared deviation"

6] Standard deviation

*) Standard deviation is a measure of how spread out the data is around the mean.

$$\boxed{\sigma = \sqrt{var(x)} = \sqrt{E[(X-\mu)^2]}}$$

*) Variance is in squared units (height in cm^2)
*) Standard deviation brings it back to the original units, making it easier to interpret

7] Sum of gaussians (Normal distributions)

*) If X and Y are independent guassians, then their sum is also guassian.

$X \sim N(\mu_1, \sigma_1^2)$
$Y \sim N(\mu_2, \sigma_2^2)$

$Z = X + Y$

$$\boxed{Z \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)}$$

*) A guassian curve is "Stable" under addition

*) Adding two independent guassians just shifts the center (Mean) and increase the spread (Variance)

*) This is a special case of the central Limit Theorem (CLT): adding many independent random variables tends toward a normal distribution.

8] Standardizing a distribution

*) Standardizing means transforming a random variable or dataset so that:
→ Its mean becomes 0
→ Its standard deviation becomes 1

$$Z\,(\text{standerized Variable}) = \frac{X - \mu}{\sigma}$$

⇒ Substracting μ = shifts the distribution so it's centered at 0
⇒ Dividing by σ = rescales the spread so Variance = 0

*) In ML standardizing can improve the convergence rate of optimization algorithms (gradient descent, SVM, PCA) and prevent some features from dominating others, leading to improve model.

9] moments of a distribution

*) Moments are numerical values that describe the shape and characteristics of a probability distribution.

*) They are essentially expected values of powers (or functions) of the random variable.

$$\mu_r' = E[x^r] = r^{th} \text{ raw moment}$$

$$E[x] = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$$
(1st moment)

$$E[x^2] = p_1 x_1^2 + p_2 x_2^2 + \cdots p_n x_n^2$$
(2nd moment)

⋮

$$E[x^k] = p_1 x_1^k + p_2 x_2^k + \cdots p_n x_n^k$$
(k^{th} moment)

10] Skewness

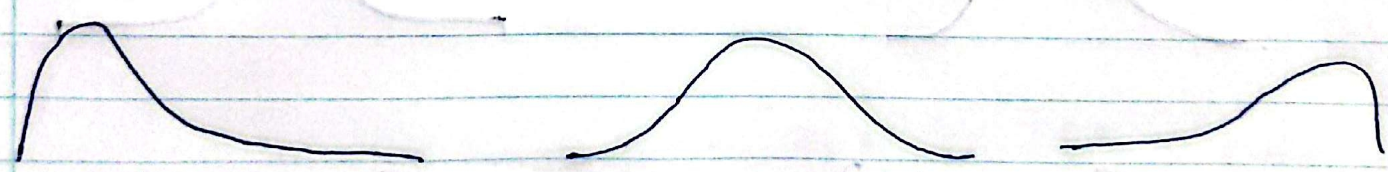
First moment $\Rightarrow$ $\mu_1 = E[x]$ (Mean)

Second moment $\Rightarrow$ $\mu_2 = E[(x-\mu)^2] = Var(x)$

*] Skewness is the third moment where we go if two distributions have same moment 1 and 2

*] Skewness is a measure of the asymmetry of a probability distribution around its mean.

$$\boxed{\text{Skewness} = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]}$$



$\rightarrow$ positively skewed

$$E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] > 0$$

$\leftrightarrow$ Not skewed

$$E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] = 0$$

$\leftarrow$ Negatively Skewed

$$E\left[\left(\frac{x-\mu}{\sigma}\right)^3\right] < 0$$

11] kurtosis

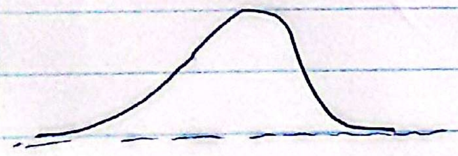
*) kurtosis measures the "tailedness" of a probability distribution

*) Basically how heavy or light the tails are Compared to a normal distribution

*) This is the fourth central moment

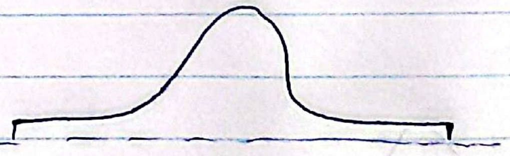
$$\boxed{\text{kurtosis} = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]}$$

Thin tails



$E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right] = \text{small}$

Thick tails



$E\left[\left(\frac{x-\mu}{\sigma}\right)^4\right] = \text{large}$

Even if they have same variances

12] Quantiles and Box plots

*) Quantiles are values that divide a dataset or probability distribution into equal parts.

25 % quantile (First quantile - Q_1)
50% quantile (median - Q_2)
75% quantile (Third quantile - Q_3)

$$\boxed{\text{Interquartile range (IQR)} = Q_3 - Q_1}$$

Ex :- $Q_1 = 18.35$
$Q_2 = 27.8$
$Q_3 = 52.95$

$IQR = 34.6$
$x_{min} = 8.7$
$x_{max} = 75$

⇓

Box plot