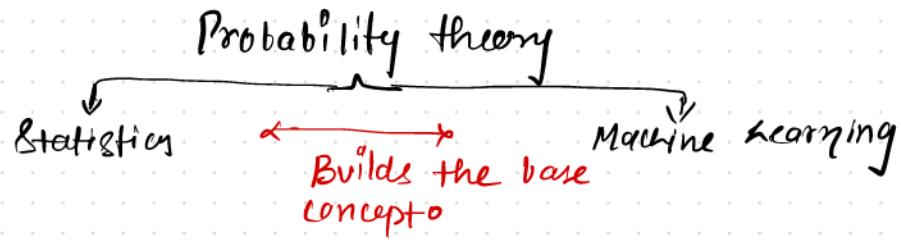


Probability & Statistics for Machine Learning

What is Probability theory?

- It's a mathematical framework that performs certain set of actions, where these actions contains some sort of uncertainty.
- Whether that person have cancer or not. **Example** It means you're not sure about something or not decided yet.
- She was uncertain of his true feelings for him.
- Usually probability is expressed over a range of 0 (will not happen) to 1 (will happen).
- It enables machine learning to predict future events based on historical data.



Terminologies :-

- * Sample Space :- It's a set of all possible outcomes of an experiment. It is denoted by Ω .
↳ symbol
- * Events :- An event is an outcome or defined collection of outcomes of an expt.
A subset of the Ω is called an Event.
- Example :- Determination of the Gender then $\Omega = \{g, b\}$

Degree of belief :-

- * If a doctor analyzes a patient & says that the patient has 40% chance of having flu.

$\frac{1}{0} \rightarrow$ absolute certainty
 $0 \rightarrow$ absolute certainty

that the patient don't have flu.
that the patient has flu.

Probability Axioms

1. (Nonnegativity) $P(A) \geq 0$, for every event A .
2. (Additivity) If A and B are two disjoint events, then the probability of their union satisfies

$$P(A \cup B) = P(A) + P(B).$$

Furthermore, if the sample space has an infinite number of elements and A_1, A_2, \dots is a sequence of disjoint events, then the probability of their union satisfies

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$$

3. (Normalization) The probability of the entire sample space Ω is equal to 1, that is, $P(\Omega) = 1$.

Some example :-



Example 1.8

In a presidential election, there are four candidates. Call them A, B, C, and D. Based on our polling analysis, we estimate that A has a 20 percent chance of winning the election, while B has a 40 percent chance of winning. What is the probability that A or B win the election?

Solution

$$\text{P(A wins or B wins)} = P(A \cup B)$$
$$= P(A) + P(B)$$
$$0.2 + 0.4 = 0.6$$

In summary, if A_1 and A_2 are disjoint events, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. The same argument is true when you have n disjoint events A_1, A_2, \dots, A_n :

$$P(A_1 \cup A_2 \cup A_3 \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n), \text{ if } A_1, A_2, \dots, A_n \text{ are disjoint.}$$

In fact, the third axiom goes beyond that and states that the same is true even for a countably infinite number of disjoint events. We will see more examples of how we use the third axiom shortly.

Note :- Review the concept of factorials & combinatorics

The Law of Large Numbers :-

Expt:- Tossing a coin; $P(\text{getting a head}) = \frac{1}{2}$ or 0.5 or 50%
Say that we've calculated that there's $30\frac{1}{10}0$ chance
that we'll get heads on every toss in a small five-toss
experiment.

The law of large numbers states that the more expts
we run, the closer we will tend to get to the expected
probability.

Gambler's fallacy:- Occurs when an individual believes
that a certain event is less likely or more likely to happen
based on past set of events.

Eg:- Tossing a fair coin

Random Variable :-

→ Given set of possible values from a random experiment.

→ Eg: Tossing a coin :-

$$\checkmark X = \begin{cases} 0 & (\text{heads}) \\ 1 & (\text{Tails}) \end{cases} \quad \because 0,1 are the possible values
of an expt. \rightarrow \text{Random events}$$

→ A R.V has whole set of values & can take any of those
values randomly.

$$\text{R.V set of values} = S$$

→ Eg: Throwing a dice :- $X = \{1, 2, 3, 4, 5, 6\}$ ✓

$$\underbrace{P(X=1)}_{6} = \frac{1}{6}; P(X=2) = \frac{1}{6} \dots$$

Probability of R.V taking the
value of $X=1$

Random Variable

Discrete random variable

Continuous random variable

* **Discrete random variable :-** It's a r.v. that can take finite or countably finite values.

→ Countable no. of states (fin. non-fin.)

→ Could be category (heads, tails)

→ Could be integer (result of rolling die)

* **Continuous random variable :-** It's a r.v. that can take infinite no. of possible values.

Eg:- $X = \text{Amt. of rainfall on a given day}$, where $X \in [0, \infty]$
this helps us in figuring out $P(X \geq 2 \text{ cm})$ or such prob.

Eg:- $X = \text{height of students}$, where we can take out the probability such as $P(X \geq 180 \text{ cm})$

↳ Probability that a student have height greater or equals to 180 cm.

Eg:- $X = \text{Time spent on a given day}$ such as $P(X \geq 10 \text{ min}) = 80\%$,
where $X \in [0, \infty]$

Probability Mass functions :-

→ It gives the probabilities for the given discrete random variables.

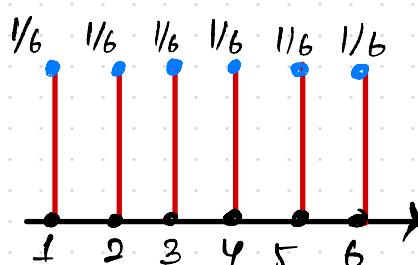
→ It's the function i.e. the probability distribution of a discrete r.v. and provides values & their probabilities.

$$P_X(x) = P(X=x)$$

* $\sum_{x \in S} P_X(x) = 1$ ↳ Sum of Probabilities sum up to 1.

$P_X(x) \geq 0 \rightarrow$ It is non-negative.

Eg:- PMF of rolling a dice :-



$$X = \{1, 2, 3, 4, 5, 6\}$$

$$P_X(1) = \frac{1}{6}$$

⋮

$$P_X(6) = \frac{1}{6}$$

fig: Bar graph of PMF !!

Note:- Probability can change & as well according to the example.

Probability Density Functions :-

- It gives the probability for continuous random variable.
- Panels out the probability coming within a distinct range of values, as opposed to giving only one value.

Histogram :-

* Used to summarize data both (continuous & discrete) by showing the no. of pts that fall in a specified range.

Example :-

Customer Wait Time in Seconds (n=20)	
43.1	42.2
35.6	45.5
37.6	30.3
36.5	31.4
45.3	35.6
43.5	45.2
40.3	54.1
50.2	45.6
47.3	36.5
31.2	43.1

Bin Ranges	
5	10
15	20
25	30
35	40
45	50
55	60
65	70
75	80



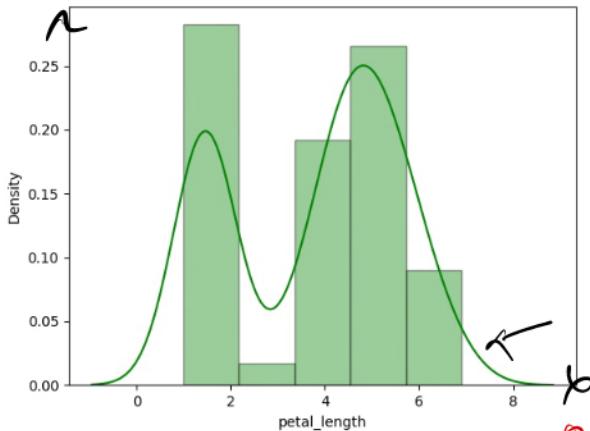
3 customer waiting b/w 1 & 35.

5 customer waiting b/w 1 & 40.
8 & so on!

Density Plot :-

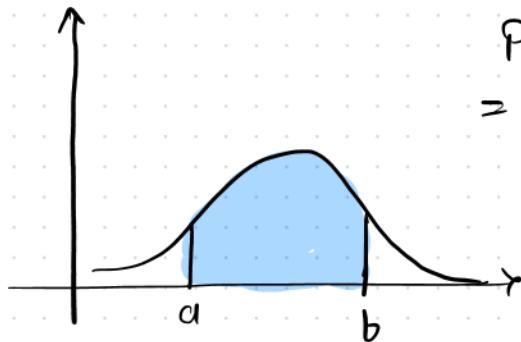
* It can be thought of as plot of smoothed histogram.
Why??

Examples are :-



Note :- We will take a detail look & analysis in histograms density plots later.

Coming back ::::



$$P(a \leq x \leq b) ?$$

$$= \int_{[a,b]} p(x) dx$$

Properties :-

$$* \int p(x) dx = 1 * \text{Every } p(x) \geq 0$$

Expected value :-

- Mean of a probability distribution. It represents avg. value we expect before collecting any data.
- Whereas Mean is typically used as when we calculate the mean of a sample. here we collect data.

Calculating Expected value :-

Eg:-

Goals (X)	Probability P(X)
0	0.18
1	0.34
2	0.35
3	0.11
4	0.02

$$E = \sum x P(x)$$



$$E = 0 \times 0.18 + 1 \times 0.34 + 2 \times 0.35 + 3 \times 0.11 + 4 \times 0.02 = 1.64$$

Calculating Mean :-

$$\bar{x} = \frac{\sum_0^n x p}{n}$$

↳ mean of the sample

* Expected no. of goals a team will score

Law of large no.s implies that as we increase the no. of expts., our \bar{x} will converge to E .

Take a look at the practical example of this.

Example:- $X = [1, 2, 3, 4, 5, 6]$ unbiased 8x-sided die.

$$\begin{aligned}E(X) &= 1 \times \left(\frac{1}{6}\right) + 2 \times \left(\frac{1}{6}\right) + 3 \times \left(\frac{1}{6}\right) + 4 \times \left(\frac{1}{6}\right) + 5 \times \left(\frac{1}{6}\right) + 6 \times \left(\frac{1}{6}\right) \\&= 3.5\end{aligned}$$

Now, say that we threw die 10 times, outcomes are
5, 2, 6, 2, 2, 1, 2, 3, 6, 1

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 3.0$$

so, avg. value is 3.0, with
0.5 of distance from the E
value 3.5.

Now, when you roll the die N times, & N become larger
then you will notice that avg. value will converge to the
expected value.

$\mu = E(X)$ ← this is what going to
happen.

Why?? ↴

Because of Law of large numbers. So as you increase
N, each possible value of $P_x(x)$ will occur with equal prob.
of $1/6$, i.e terms avg. to the expected value!

Measures of Central tendency :-

- * A measure of C.T is a single value that attempts to
describe a set of data by identifying the central position
within that set of data.

Measures

Mean Median Mode

Mean :-

- * Commonly / Majority used for continuous data.
- * Sample mean = $\bar{x} = \frac{\sum_i^n x_i}{n}$
- * Population mean = $\mu = \frac{\sum_i^n x_i p_i}{n}$
- * Mean tells us the value that is most common.
- * Mean isn't the actual value present in data.

When Mean doesn't work?

→ if your data have outliers, then mean can get influenced and may give you wrong interpretation of data. So we have median which doesn't gets influence by outliers.

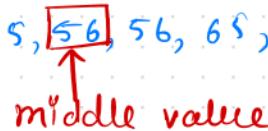
Median :-

- * Middle score of the dataset that has been arranged in order of magnitude.
- * It's less affected by outliers.

Eg:-

14, 35, 45, 55, 55, 56, 56, 65, 87, 89, 92

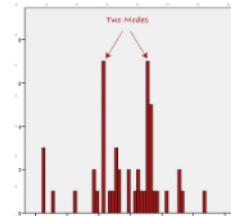
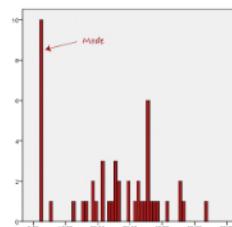
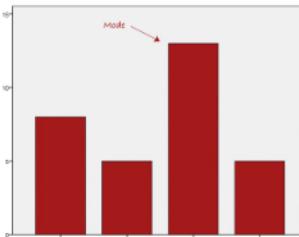
find out how to take media of an even no. of scores?


middle value

Mode :-

- * Most frequent score in our dataset.

Problems that have :-



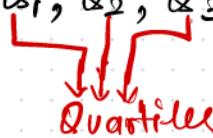
Quantiles: Percentiles, Quartiles, and Deciles

- * A Quantile is where a Sample is divided into equal-sized, sub-groups. It can also refer to dividing a probability distribution into areas of equal probability.
- * Quartiles:- It divides the distribution into four equal parts.
- * Percentiles:- It divides the distribution into 100 equal parts.
- * Deciles:- It divides the distribution into 10 equal parts.

Quartiles :-

- Divides the entire set into 4 equal parts.
- There are 3 Quartiles, Q_1 , Q_2 , Q_3

Quartiles formula:-



- * Put the data in order.
- * Cut the data / distribution in 4 equal parts.
- * The Quartiles are at the cuts.

2, 4, 4, 5, 6, 7, 8
 Q_1 Q_2 Q_3
(lower quartile) (middle quartile) (upper quartile)
↳ also called median!!

Interquartile Range

- * The Interquartile range is from Q_1 to Q_3 :

$$\begin{array}{|c|c|c|c|} \hline 25\% & 25\% & 25\% & 25\% \\ \hline Q_1 & Q_2 & Q_3 & \\ \hline \end{array}$$
$$= Q_3 - Q_1$$

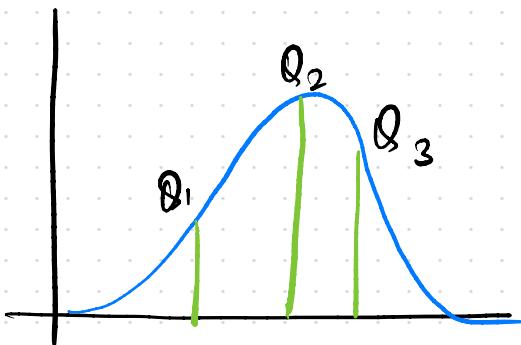
Interquartile range :-

2, 4, 4, 5, 6, 7, 8

$$Q_3 - Q_1 = 7 - 4 = \boxed{3}$$

Interquartile range

Quartiles divide the probability distribution in 4 equal parts :-



Interpreting Quartiles :- $\rightarrow Q_1$

Eg :- 59, 60, 65, 65, $\boxed{68}$, 69, 70, 72, 75, $\boxed{75}$, 76, 77, 81, 82,
84, 87, 90, 95, 98

$\downarrow n * (1/4)$

$\uparrow Q_2$ (Median)

* Q_1 is the central point b/w smallest and the median.

$$* Q_1 = n * (1/4) = 19 * \frac{1}{4} = \frac{19}{4} = 4.75 \approx 5$$

Q_1 is at 5th index / pos.

* Q_3 is the middle value b/w Q_2 & the highest score.

$$Q_3 = (82 + 84)/2 = 83$$

$Q_3 = 84$ (as it is the median b/w Q_2 & max)

Score of Q_1 represent 1st Quartile & is the 25th percentile.

It represents the median of the lower half of the score set in the overall data.

$\rightarrow Q_1$ tells 25% of the scores are less than 68 & 75% of the scores are greater.

→ Q_2 is the median & 50th percentile & shows that 50% of the scores are less than 75% & vice-versa.

→ Q_3 is the 75th percentile, reveals that 25% of the scores greater than 84 & 75% less than 84%.

→ IQR tells how far apart the 1st & 3rd Quartile are, indicating how spread out 50% of our dataset is!

IQR is resistant to outliers:-

* helps in measuring the spread of a dataset.

Eg :- 2, 3, 3, 4, 5, 6, 6, 7, 8, 8, 8, 9

$$Q_1 = 3.5, Q_2 = 6, Q_3 = 8$$

* IQR = $8 - 4 = 4$

* Range = $9 - 2 = 7$

What if in data we introduce 100 as a pt in data?

range = $100 - 2 = 98$

Note:- We will see the use of IQR in identification of outlier soon.

Percentiles :- certain % of scores fall below that %.
if you know that your score is in the 90th percentile, that means you scored better than 60% of people who took the test.

Say for ex :- the 70th percentile on the 2013 IITJEE was 156.
If you scored 156 on the exam, your score was better than 70% of test takers.

→ The 25th percentile is called the first quartile.

→ The 50th percentile is called the median.

→ The 75th percentile is called the 3rd Quartile.

→ The difference b/w third & first Quartile.

Percentile Rank :- In above example, 70th percentile, 70 is a percentile rank.

Decile :- It divides the distribution into ten equal parts.

→

15	22	24	27	32	36	40	41	50	90
↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
1	2	3	4	5	6	7	8	9	10

1st decile → 15 & 6th decile = 36

↓
100% of students scored below 15.

↓
60% of students scored below 36.

Variance :- Variance is a measure of how data pts. differ from the mean. It is how far set of data are spread out from their avg. value.

↑ More variance ↑ More data is scattered or spread out from the mean

* Measure of spread from data

Formula :-

$$\text{Variance} = (\sigma)^2 \quad \text{standard deviation}$$

→ mean value of all observation

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad \begin{array}{l} \text{→ the value of one observation} \\ \text{the total no. of observations} \end{array}$$

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

Why for sample variance we have $n-1$?
(We will talk about it)

If we talk about probability distribution, or variance of R.V.
then concept of Expectation takes place.

Eg:-

$$X = [-100, 100] ; P_X(x) = \begin{cases} 0.5 & \text{for } x = -100 \\ 0.5 & \text{for } x = 100 \\ 0 & \text{otherwise} \end{cases}$$

$$Y = [0] ; P_Y(y) = \begin{cases} 1 & \text{for } y = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \sum_{\text{all } x} g(x) p(x)$$

$$= -100 \cdot 0.5 + 100 \cdot 0.5 = 1 \times 0 \\ = 0$$

$$E[Y] = \sum_{\text{all } x} g(x) p(x)$$

So, $E[X] = E[Y] = 0$ (to be noted)

Note:- Both R.V.s have same mean value, their distribution is completely different.

$$y = E[Y]$$

↳ it's mean = 0

$$x = \underbrace{100, \text{ or } -100}_{\text{far away from mean}}$$

far away from mean

* Variance is measure of how spread out the distribution of a r.v. is from mean.

* Here, the variance of Y is quite small since its distribution is concentrated at a single value, while $\text{Var}(X)$ is large.

Variance of X :-

$$E[(X - \mu)^2] = \sum_{\text{all } x} (x - \mu)^2 p(x)$$

$$\downarrow$$

$$\text{Var}(x)$$

Why we are squaring?

$$E[X - \mu_X] = E[X] - E[\mu_X] = \mu_X - \mu_X = 0$$

If we don't sq. the difference b/w X & μ 's mean, the result is 0.

$$\text{Var}(X) = E[(X - \mu_X)^2] = \sum_{i=1}^n (x_i - \mu_X)^2 P_X(x_i)$$

Variance of X

$$\text{Var}(X) = (-100 - 0)^2(0.5) + (100 - 0)^2(0.5) = 10,000$$

$$\text{Var}(Y) = (0 - 0)^2(1) = 0.$$

You can see the $\text{Var}(X)$ is way larger than $\text{Var}(Y)$

let's say we have a dataset which states the size of Watermelons in meters; but $\text{Var}(X)$ has different units. Say X is in meters, then $\text{Var}(X)$ will be

80, unit of $X \neq \text{Var}(X)$ unit \rightarrow in (meters)²

It's an issue!! like \rightarrow we can take the square root of variance.

$$\sqrt{\text{Var}(X)} \leftarrow \sigma_X$$

* this introduces another measure called Standard deviation.

$$\text{But, } \sigma_X = \sqrt{10,000} = 100 \quad \& \quad \sigma_Y = \sqrt{0} = 0$$

Computational formula for Variance :-

$$E[(X - \mu)^2] = E(X^2) - [E(X)]^2 \quad (\text{How to prove it?})$$

$$\text{Hind :- } \text{Var}(X) = E[(X - \mu_X)^2]$$

\rightarrow expand it using $(a-b)^2$ identity.

$$\text{Var}(X) = E[X^2] - [E(X)]^2$$

↓ ↗ $\sum_{\text{all } x} x^2 p_x(x) = \mu = (\mu)^2$

$$E[X^2] = \sum_{\text{all } x} x^2 p_x(x)$$

Example :- Roll a fair die & let X be the resulting no. Find $E[X]$; $\text{Var}(X)$; & σ_X

$$R_X = \{1, 2, 3, 4, 5, 6\} \text{ and } P_X(x) = \frac{1}{6} \text{ for } x = 1, 2, \dots, 6$$

$$E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= \frac{7}{2}$$

$$E[X^2] = 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6}$$

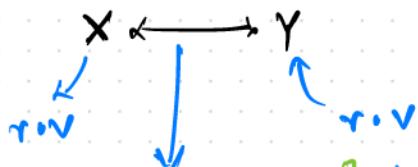
$$= \frac{91}{6}$$

$$\text{So; } \text{Var}(X) = E[X^2] - [E(X)]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2$$

$$\sigma_X = \sqrt{\text{Var}(X)} \approx \sqrt{2.92} \approx 1.71 = \frac{91}{6} - \frac{49}{4} \approx 2.92$$

Note :- Theorems are given in reading materials.

Covariance :-



A finance example :-

Say you're a investor, his portfolio track performance of the S&P 500 and you want to add the stock of ABC Corp.

Data :-

S&P 500 ABC Corp.

S&P 500 \longleftrightarrow ABC Corp.

→ Wants to assess the directional relationship b/w the stock of ABC Corp. & S&P 500

2013 1692 68

2014 1978 102

2015 1884 110

2016 2151 112

2017 2519 154

\uparrow
 $X(R.o.V)$

$X \longleftrightarrow Y$ Identify

→ Wants to identify the directional relationship.

* If both of the stocks tend to increase together, then we say it has positive covariance.

* If both of the stocks where one increase & other decrease then we say it has negative covariance.

Covariance formula :-

Covariance b/w two R.o.V X & Y for random variable :-

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}, \text{ Sample Covariance}$$

Where :-

- X_i - the values of the X-var → \bar{Y} - the values of the Y-var
- Y_i - the values of the Y-var → n - the no. of data pts.
- \bar{X} - the mean of the X-var

$$\text{Mean (S&P 500)} = \frac{1692 + 1978 + 1884 + 2151 + 2519}{5} = 2,044.80$$

$$\text{Mean (ABC Corp.)} = \frac{68 + 102 + 110 + 112 + 154}{5} = 109.20$$

$$1692 - 2044$$

$$68 - 109.20$$

$$-352.80 \times -41.20$$

$$1978 - 2044$$

$$102 - 109.20$$

$$-66.80 \times -7.20$$

$$1884 - 2044$$

$$110 - 109.20$$

$$-16.80 \times 0.80$$

$$2151 - 2044$$

$$112 - 109.20$$

$$10.80 \times 2.80$$

$$2519 - 2044$$

$$154 - 109.20$$

$$47.40 \times 44.80$$

$\text{Cov(S&P 500, ABC Corp.)}$

$$= \frac{36,429.20}{5-1}$$

$$= 9,107.00$$

positive covariance ↑

$$14,585.36$$

$$480.96$$

$$-128.64$$

$$297.36$$

$$21,244.16$$

$$\underline{\underline{36,429.20}}$$

Note:- We will take a look at covariance matrix when looking in Colab.

Correlation :-

→ Another measure to examine the relationship b/w two variables. While covariance b/w two var measures the directional relationship.

But correlation measures the strength of the relationship.

Correlation Coefficient ranges from -1 to 1.

- * if correlation coef. of -1 describes a perfect negative or inverse correlation, in which one rises other decreases.
- * if correlation coef. of 1 describes a perfect pos. correlation in which one increases & other as well.
- * if relation coef. is 0, then there is no relationship.

→ The most common is Pearson coefficient or Pearson's which measures the strength and direction of linear relationship b/w two variables.

Correlation Coefficient Equation

$$r_{xy} = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

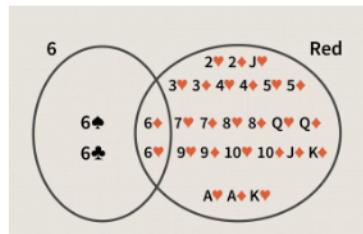
Pearson product moment correlation coef. $\sigma_x \rightarrow$ Standard deviation of x .
 $\sigma_y \rightarrow$ Standard deviation of y .

Point Probability Distribution

- * Probability distributions can represent probability of multiple r.v.s simultaneously.
- * Probability of both $x=x$ & $y=y$ is: $P(x=x, y=y)$

Eg:- $P(6 \cap 52) = 2/52 = 1/26$
 $P(6) \times P(\text{red}) = 4/52 \times 26/52 = 1/26$

" \cap " is referred as intersection which means it will happen at the same time.



Marginal Probability

→ It's the probability of a single event occurring, independent of other events

Sum rule :-

$$\forall x \in X, P(x=x) = \sum_y P(x=x, y=y)$$

Example: Calculate the Marginal Probability of Pet Preference among men and women.

	Cats	Fish	Dogs	
Men	2	4	6	12
Women	5	3	2	10
	7	7	8	22

More examples with different scenarios is in reading materials.

Total no. of people = 22.

$$\begin{aligned} \rightarrow \text{People who prefer Cats} &= 7/22 = 0.32 \\ \rightarrow \text{People who prefer Fish} &= 7/22 = 0.32 \\ \rightarrow \text{People who prefer Dogs} &= 8/22 = 0.36 \end{aligned} \quad \left. \right\} \text{Marginal}$$

Conditional Probability

- * The likelihood of an event occurring based on the occurrence of a previous event or outcome.
- * Two events are independent if one event does not affect the another event, however if one event occurring / not affect the probability of other event, then it is dependent.
- * If events are independent, then it does not relate to conditional probability.

$$P(A|B)$$

↳ Probability of A given B.

Conditional Probability formula

$$\begin{aligned} P(B|A) &= \frac{P(A \text{ and } B)}{P(A)} \\ &= \frac{P(A \cap B)}{P(A)} \end{aligned} \quad \left. \right\} \text{Revisiting this}$$

Eg:- We say that there is 23% of the days are rainy.

$P(R) = 0.23$, where R is the event that it will rain on random day.

Let's assume we pick a day & say that it's a cloudy day,
extra information

What's the prob. that it rains on a given day

given that the day is cloudy?

$$P(R|C)$$

$\xrightarrow{\text{rain}} \xrightarrow{\text{cloudy}}$

formula:-

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, P(B) \neq 0$$

Some Special cases :-

- When A & B are disjoint: $A \cap B = \emptyset$, so,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\emptyset)}{P(B)} = 0$$

Since, A & B are disjoint, they can't occur at the same time.

- When B is a subset of A: then whenever B happens, A also happens.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

When A is a subset of B, then

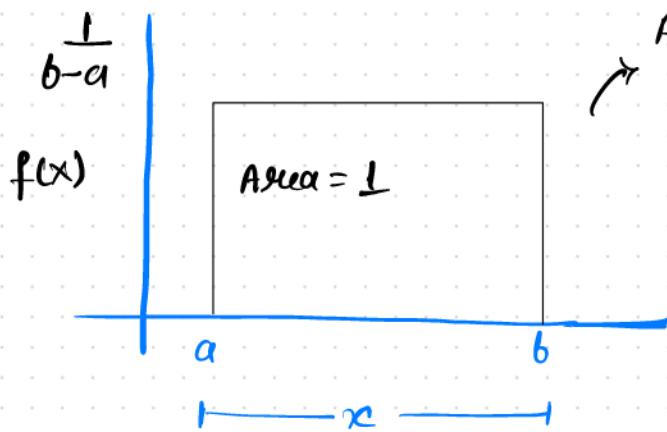
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}$$

Note:- The solved problems of conditional probability will be released as soon as possible.

Distributions

1.) Uniform distribution :- Probability is constant over all the possible values of x .

Continuous uniform distribution :-



$$\text{Area} = b \times l$$

$$\Rightarrow (b-a) f(x) = 1$$

$$\Rightarrow f(x) = \frac{1}{b-a}$$

So, the PDF of the uniform distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } c \leq x \leq d \\ 0 & \text{otherwise} \end{cases}$$

$$\text{median} = \frac{a+b}{2}; \text{ mean} = \frac{c+d}{2}$$

$$\sigma^2 = \frac{1}{12} (d-c)^2$$

you don't need to know how it came!

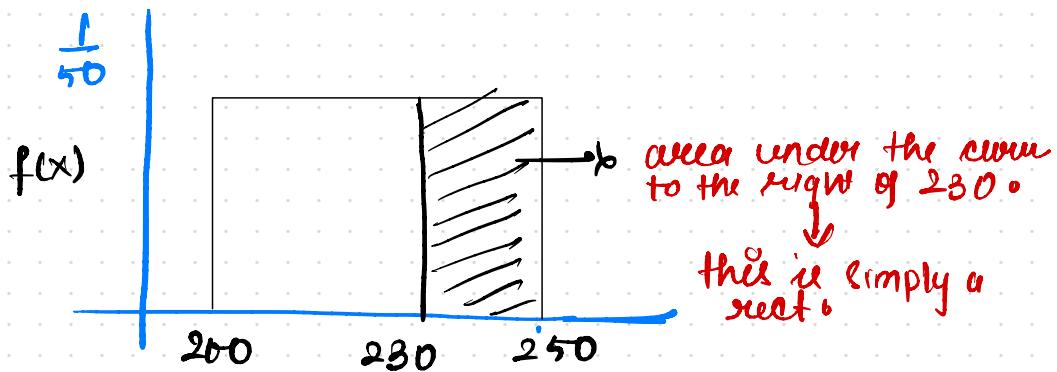
* AUC of the uniform distribution is just a rectangle.

Example (Statistic ref.) :-

- Say X is a r.v. that has a uniform distribution with $a=200$ and $b=250$. What is $f(x)$?

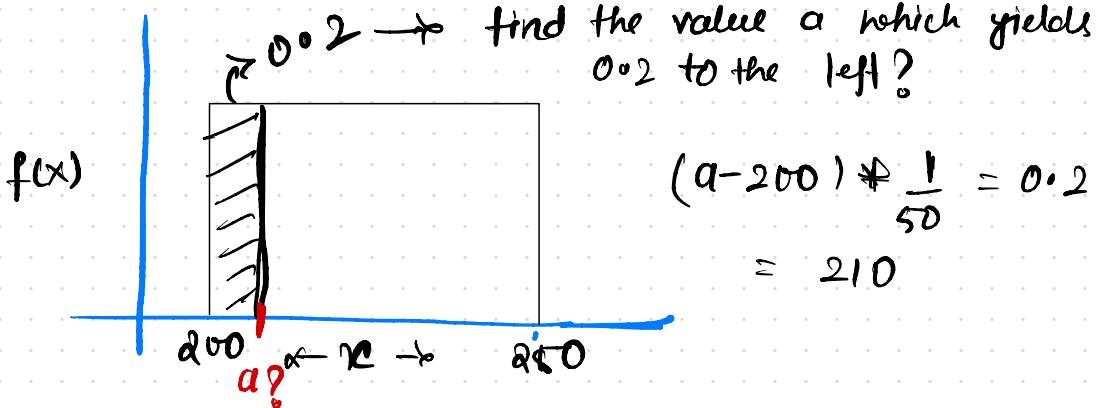
$$f(x) = \frac{1}{b-a} = \frac{1}{250-200} = \frac{1}{50} \quad \text{for } 200 \leq x \leq 250 \\ = 0 \quad \text{otherwise}$$

What is $P(X > 230)$?



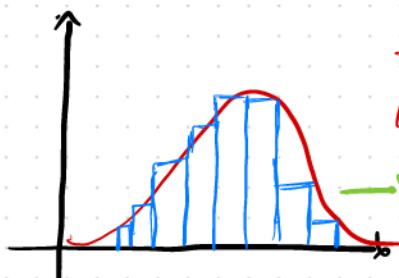
$$P(X > 230) = (250 - 230) * \frac{1}{50} = 0.4$$

What is the 20th percentile?



Normal Distribution :-

- * It is symmetric about the mean, showing that data near the mean are more frequent to occur than data far from mean.



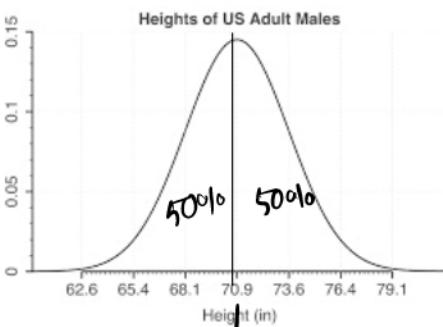
→ There are many cases in which data tends to be around a central value with no bias left or right.

→ Normal distribution (Bell curve)

Many such examples follow N.D -

→ Height of people

When we say that data is Normally distributed

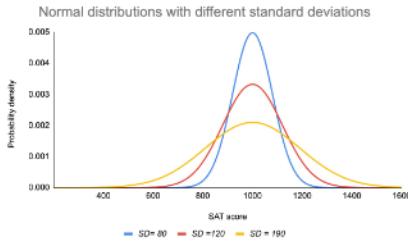
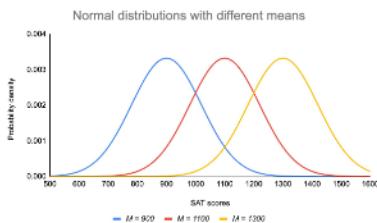


- mean = median = mode
- Symmetric about the centre
- 50% of values are less than mean & 50% of values are $>$ than mean.

→ mean = median = mode

All N.D can be described by 2 pairs to mean & standard deviation

→ Mean is the location parameter and standard deviation is the scaling parameter.



Empirical Rule :-

It tells where most of our values lie in a Normal distribution.

Around 68% of the values are within 1 S.D from the mean.

Around 95% of the values are within 2 S.D from the mean.

Around 99.7% of the values are within 3 S.D.

formula :- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ (You only need μ & σ)

* $N(\mu = 0, \sigma = 1)$

Normal distribution has mean μ and standard deviation σ .

If mean = 0 and standard deviation is 1, then it's called Standard Normal distribution.

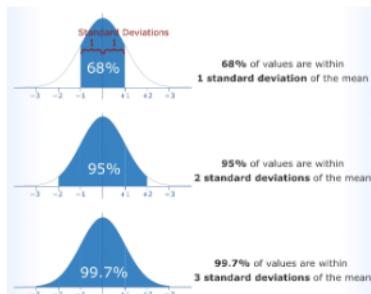
Standardizing With Z-score :-



$$Z\text{-Score} = \frac{x - \mu}{\sigma}$$

Empirical Rule :-

→ The empirical rule gives us how much of the data lies within one, two or three std. deviations from the mean.



Example: Travel Time

A survey of daily travel time had these results (in minutes):

26, 33, 65, 28, 34, 55, 25, 44, 50, 36, 26, 37, 43, 62, 35, 38, 45, 32, 28, 34

The Mean is 38.8 minutes, and the Standard Deviation is 11.4 minutes (you can copy and paste the values into the [Standard Deviation Calculator](#) if you want).

Convert the values to z-scores ("standard scores").

To convert 26:

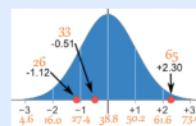
- first subtract the mean: $26 - 38.8 = -12.8$,
- then divide by the Standard Deviation: $-12.8/11.4 = -1.12$.

So 26 is **-1.12 Standard Deviations** from the Mean

Here are the first three conversions

Original Value	Calculation	Standard Score (z-score)
26	$(26-38.8) / 11.4 =$	-1.12
33	$(33-38.8) / 11.4 =$	-0.51
65	$(65-38.8) / 11.4 =$	+2.30
...

And here they are graphically:



You can calculate the rest of the z-scores yourself!

Standardizing with Z-score :-

Mean and Standard deviations for the SAT and ACT

	SAT	ACT
Mean	1500	21
SD	300	5

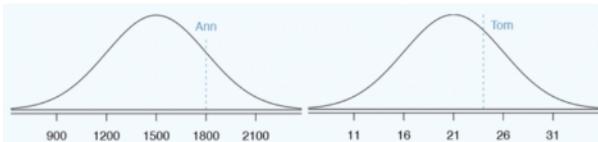


Figure 3.1.4: Ann's and Tom's scores shown with the distributions of SAT and ACT scores.

- Ann is 1 std. deviation above from average.
- Tom's is 0.6 std. deviation above from average.

$$\mu_{\text{SAT}} = 1500, \sigma_{\text{SAT}} = 300, \text{ and } x_{\text{Ann}} = 1800$$

$$z_{\text{Ann}} = \frac{x_{\text{Ann}} - \mu_{\text{SAT}}}{\sigma_{\text{SAT}}} = \frac{1800 - 1500}{300} = 1$$

$$z_{\text{Tom}} = \frac{x_{\text{Tom}} - \mu_{\text{ACT}}}{\sigma_{\text{ACT}}} = \frac{24 - 21}{5} = 0.6$$

- * The Z-score of an observation is the no. of standard deviations it falls above or below the mean. we compute Z-score for an observation x that follows a distribution with mean μ and standard deviation σ using :-

$$Z = \frac{x - \mu}{\sigma}$$

[https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_OpenIntro_Statistics_\(Diez_et_al\)/03%3A_Distributions_of_Random_Variables/3.01%3A_Normal_Distribution](https://stats.libretexts.org/Bookshelves/Introductory_Statistics/Book%3A_OpenIntro_Statistics_(Diez_et_al)/03%3A_Distributions_of_Random_Variables/3.01%3A_Normal_Distribution)

Geometric distribution :-

- How long should we expect to flip a coin until it turns up heads?
- How many times should we expect to roll a die until we get 1?

The Questions are answered using geometric distribution!!

Geometric distribution :- Used to describe how many trials it takes to success.

If the probability of a success in one trial is p and the probability of a failure is $1-p$, then the probability of finding the first success in the n th trial is given by

$$(1-p)^{n-1} p$$

Eg:- Suppose you're playing game of darts. The probability of success is 0.4. What is the probability you will hit the bullseye on the 3rd try?

$$P=0.4$$

$$P(X=n) = (1-p)^{n-1} p; P(X=3) = (1-0.4)^{3-1} (0.4); P(X=3) = 0.144$$

The mean (expected value), variance and std. deviation of this wait time are given by :-

$$H = \frac{1}{p}; \quad \sigma^2 = \frac{1-p}{p^2}; \quad \sigma = \sqrt{\frac{1-p}{p^2}}$$

here, H & expected value are same. On avg. it takes $\frac{1}{p}$ to get success under the geometric distribution.

→ If the $P(\text{Success})$ is ↑ then you don't have to wait for long for a success.

$$\text{eg: } P(\text{Success}) = 0.8; \quad \frac{1}{0.8} = 1.25 \text{ trials on avg.}$$

→ If the $P(\text{Success})$ is ↓ then you do have to wait for long for a success.

$$\text{eg: } P(\text{Success}) = 0.1; \quad \frac{1}{0.1} = 10 \text{ trials on avg.}$$

Bernoulli Distribution

→ This distribution is quite simple left for students to explore.

Binomial distribution :-

→ Used to describe the no. of successes in a fixed no. of trials. Geometric distribution describes which is the no. of trials we must wait before we observe a success.

Note :- It is not used too much, though in assignment you're asked to explore about it!

Hypothesis test :-

* Say you run an expt. and find that a certain drug is effective at treating headaches.

What is Hypothesis?

* It's an educated guess about something.

a guess based on knowledge
and experience.

* A claim to test, it can be an expt. or an observation.

Eg :-

→ Drug A works better than Drug B.

→ A way of teaching you might think be better.

→ it can be really anything at all as long as you can put it to the test.

What is a Hypothesis Statement?

* If you're going to propose a hypothesis; it's customary to write a statement.

"If I ... (do this) ... then (this will happen)"

Eg :- • if I give patients counseling in addition to medication)
then (their overall depression scale will decrease).

Null hypothesis - H_0 - The null hypothesis is always accepted as fact. Simple example of null hypotheses that are generally accepted as being true are:-

1. Smoking烟 can increase your risk of heart problems.

Alternate hypothesis - H_a - The alternate hypothesis is also called research hypothesis. It involves the claim to be tested.

What is Hypothesis Testing?

* It's a way for you to test the results of a survey or expt. to see if you have meaningful results.

