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.

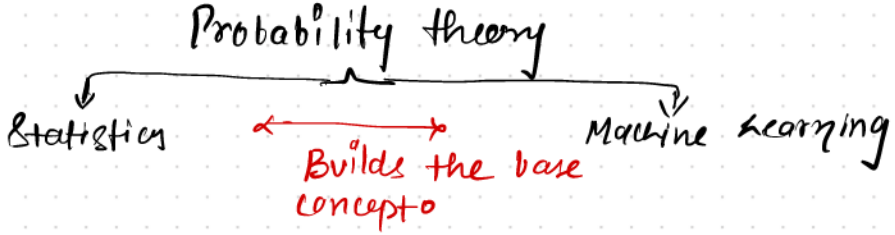
- Whether that person have cancer or not.

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 enable 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. ↳ symbols

* Events :- An event is an outcome or defined collections 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.

1 → absolute certainity that the patient has flu.
0 → absolute certainity that the patient don't have flu.

Probability Axioms

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

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

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

$$\mathbf{P}(A_1 \cup A_2 \cup \cdots) = \mathbf{P}(A_1) + \mathbf{P}(A_2) + \cdots$$

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

P(A) + P(B) = P(A∪B)

Some example :-

P(A Nim or B win) = P(A∪B) = P(A) + P(B) = 0.2 + 0.4 = 0.6

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

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 $A_1, A_2, \cdots, A_n$:

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

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(getting a head) = $\frac{1}{2}$ or 0.5 or 50%

Say that we've calculated that there's 30% chance that we'll get heads on every toss in a small five-toss experiment.

The large of large nos. states that the more expt's 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 :-

$X = \begin{cases} 0 \text{ (heads)} \\ 1 \text{ (Tails)} \end{cases}$ → Random events

∵ 0, 1 are the possible values of an expt.

→ A RV has whose set of values & can take any of those values randomly.

R.V set of values = Ω

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

$P(X=1) = \frac{1}{6}$; $P(X=2) = \frac{1}{6}$ ...

Probability of RV 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 state (finite, non-finite)
  - → Could be category (head, tail)
  - → 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 = Amt of rainfall on a given day, where $X \in [0, \infty]$ this helps us in taking out $P(X \geq 2\text{cm})$ or such probs.

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

Probability Mass function :-

→ 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.

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

* $\sum_x P_X(x) = 1$ ↳ Sum of probabilities sum up to 1.

$P_X(x) \geq 0$ → It is non-negative.

Eg:- PMF of rolling a dice :-

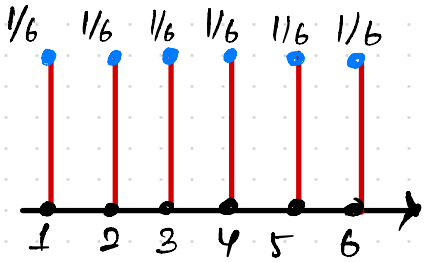


fig: Graph of PMF!!

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

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

⋮

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

Note:- Probabilities can changes as well according to the example.

Probability Density Functions :-

→ It gives the probablity for continuous random variable.

→ Tells out the probability coming within a distinct range of values, as opposed to only one value.

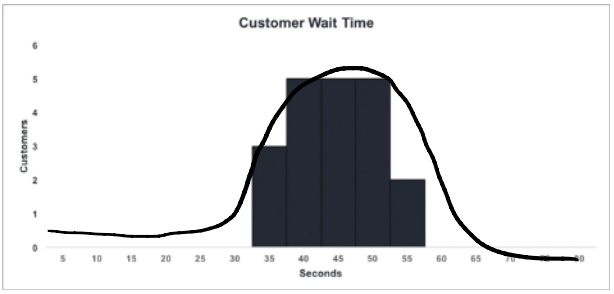
Histogram :-

* Used to summarize data both (continuous & discrete) by showing the no. of pt.s 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.

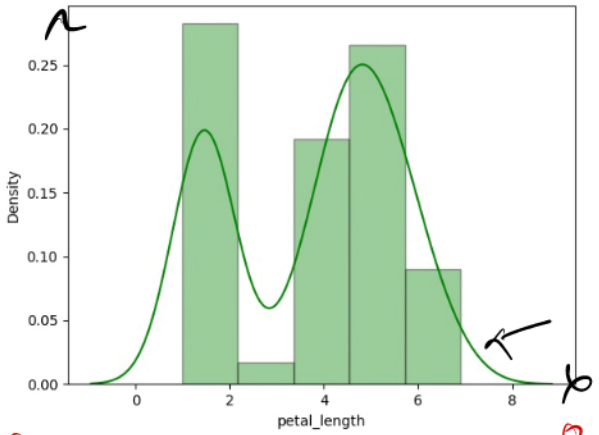
5 customer waiting b/w 1 & 40 & so on!

Density Plots :-

* It can be thought of as plots of smoothed histogram.

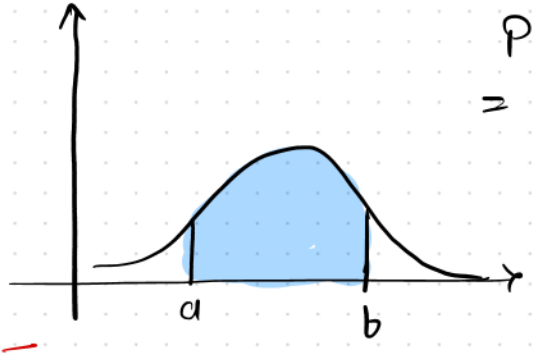
Wdym??

Examples are :-



Note :- We will take a detail look & analysis on histogram & density plots later.

Coming back ...



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

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

Properties :-

* $\int p(x)\,dx = 1$ * Every $p(x) \ge 0$

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

$$\mathbb{E} = \sum_x x P(x)$$

$$\mathbb{E} = 0 \times 0.18 + 1 \times 0.34 + 2 \times 0.35 + 3 \times 0.11 + 4 * 0.02 \simeq \boxed{1.45}$$

* Expected no. of goals a team will score

Calculating Mean :-

$$\bar{x} = \frac{\sum_i^n x_i}{n}$$

↳ mean of the sample

Law of large nos. implies that as we increase the no. of exptns, our $\bar{x}$ will converge to $\mathbb{E}$.

Take a look at the practical example of this.

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

$$\mathbb{E}(X) = 1\times\left(\frac{1}{6}\right) + 2\cdot\left(\frac{1}{6}\right) + 3\cdot\left(\frac{1}{6}\right) + 4\cdot\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\cdot\left(\frac{1}{6}\right)$$

$$= 3.5$$

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

$$\bar{x} = \frac{\sum_{p}^{n} x_p}{n} = 3.0$$

So, avg. value is 3.0, with 0.5 of distance from the $\mathbb{E}$ 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.

$\mu = \mathbb{E}(X)$ ← 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 / Majorly used for continuous data.
* Sample mean = $\bar{x} = \frac{\sum_{i}^{n} x_i}{n}$
* Population mean = $\mu = \frac{\sum_{i}^{n} x_i}{n}$

* 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

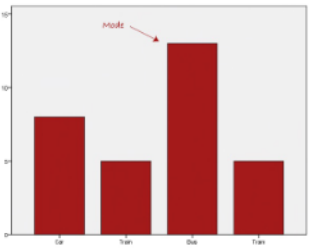
middle value

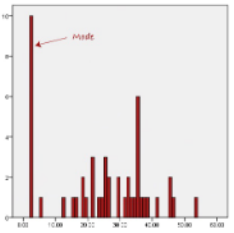
find out how to take media of an even no. of scores ?

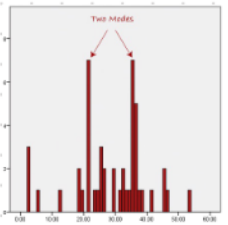
Mode :-

* Most frequent score in our dataset.

Problems this 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

Quantiles 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 (lower Quartile) Q_2 (Middle Quartile) Q_3 (upper Quartile)

↳ also called median!!

Interquartile Range

* The interquartile range is from Q_1 to Q_3:

25%	25%	25%	25%

Q_1 Q_2 Q_3

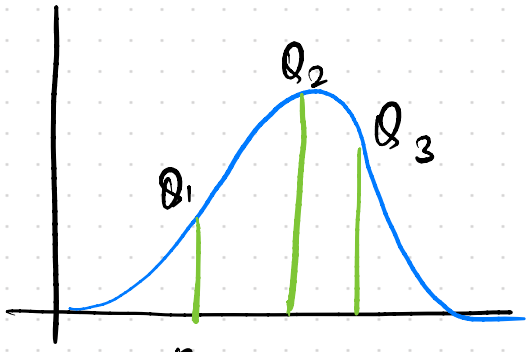
$= Q_3 - Q_1$

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 :- → Q_1

$n * (2/4)$

Eg :- 59, 60, 65, 65, **68**, 69, 70, 72, 75, **75**, 76, 77, 81, 82, 84, 87, 90, 95, 98

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 avail data.

→ 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 score 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 spred 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 = $100 - 2 = 98$

so bad !!

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

Percentile :- certain % of scores fall below that no.
if you know that your score is in the 90th percentile, that means you scored better than 90% of people who took the test.

Say for an :- the 70th percentile on the 2013 LSAT 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 divide 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

↓ 10% 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 spred out from their avg. value.

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

* Measure of spredd from data.

Formula :-

variance = $(\sigma)^2$ → standard deviation

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$$

→ mean value of all observation

→ the value of one observation

↑ the total no. of observation

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

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

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

Eg:-

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

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

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

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

$= 0$ $= 0$

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

Note:- Both R.Vs have same mean value; their distribution is completely different.

$\mu = E[Y]$ ↳ it's mean = 0 $X = \underbrace{100, \text{ or } -100}$ 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; whereas Var(X) is large.

Variance of X :-

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

↓

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 & it's mean, the result is 0.

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

Variance of X

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

$Var(Y) = (0-0)^2(1) = 0.$

You can see the $Var(X)$ is way larger than $Var(Y)$

let's say we have a dataset which state the size of watermelons in meters; but $Var(X)$ has different units. Say X is in meters, then $Var(X)$ will be in $(meters)^2$

So, unit of $X \neq Var(X)$ unit

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

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

* this introduces another measure called Standard deviation.

So, $\sigma_X = \sqrt{10{,}000} = 100$ & $\sigma_Y = \sqrt{0} = 0$

Computational formula for variance :-

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

Hint :- $Var(X) = E[(X-\mu_X)^2]$ → expand it using $(a-b)^2$ identity.

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

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

$$[E(X)]^2 \rightarrow \mu_X^2 \rightarrow \sum_{\text{all } x} x P_X(x) = \mu = (\mu)^2$$

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

$R_X = \{1,2,3,4,5,6\}$ and $P_X(k) = \frac{1}{6}$ for $k = 1, 2 \cdots 6$

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

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

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

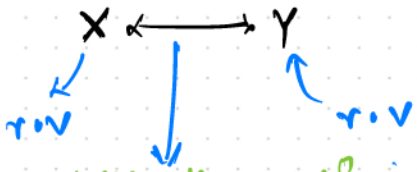
$$= \frac{91}{6}$$

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

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

Note :- Theorems are given in reading materials.

Covariance :-



tells the relationship b/w two R.V

A finance example :-

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

S&P 500 ⟷ ABC corp.

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

Data :-

	S&P 500	ABC corp.
2013	1692	68
2014	1978	102
2015	1884	110
2016	2151	112
2017	2519	154
	↑ X (R.V)	↑ Y (R.V)

X ⟷ 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 increases & other decreases then we say it has negative covariance.

Covariance formula :-

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

$$Cov(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n}$$

$$Cov(X,Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$ → Sample covariance

Where :-

→ x_i - the values of the X-var
→ y_i - the values of the Y-var
→ $\bar{x}$ - the mean of the X-var
→ $\bar{y}$ - the values of the Y var.
→ n - the no. of data pts

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

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

1692 − 2044	68 − 109.20	−352.80 × −41.20
1978 − 2044	102 − 109.20	−66.80 × −7.20
1884 − 2044	110 − 109.20	−160.80 × 0.80
2151 − 2044	112 − 109.20	106.20 × 2.80
2519 − 2044	154 − 109.20	474.20 × 44.80

Multiply

14,535.36
480.96
−128.64
+ 297.36
21,244.16
36,429.20

Cov(S&P500, ABC Corp.)

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

$= 9,107.30$

positive covariance

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 vars 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

$$\rho_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

- ρ_{xy} → pearson product moment correlation coef.
- $cov(x,y)$ ← covariance of var. x & y.
- σ_x → standard deviation of x.
- σ_y → standard deviation of y.

Joint Probability Distribution

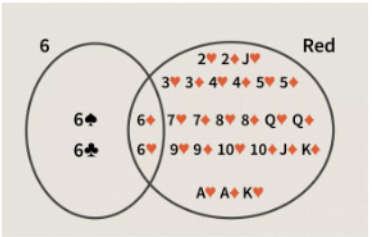
* Probability distributions can represent probability of multiple r.v.s simultaneously.
* Probability of both $x = x$ & $y = y$ is: $P(x = x, y = y)$

eg:-

→ $P(6 \cap 52) = 2/52 = 1/26$

$P(6) \times P(red) = 4/52 \times 26/52 = 1/26$

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



Marginal Probability

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

Sum rule :-

$$\forall x \in X, \; P(X=x) = \sum_{y} P(X=x, Y=y)$$

(X: r.v)

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.

→ People who prefer cats = 7/22 = 0.32
→ People who prefer Fish = 7/22 = 0.32
→ People who prefer Dogs = 8/22 = 0.36

} Margine

Conditional Probability

* The likelihood of an event occuring based on the occurence of a previous event or outcome.
* Two events are independent if one event does not affect the another event, however if one event occuring not affect the probability of other event, then it is dependent.
* if events are independent, then it does not relates to conditional probability.

P(A|B)
↳ Probability of A given B.

Conditional Probability formula :-

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \quad \text{or} \quad = \frac{P(A \cap B)}{P(A)}$$

} 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?

$P(R|C)$ — R: rain, C: cloudy

Formula:-

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

Some special case:-

- 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

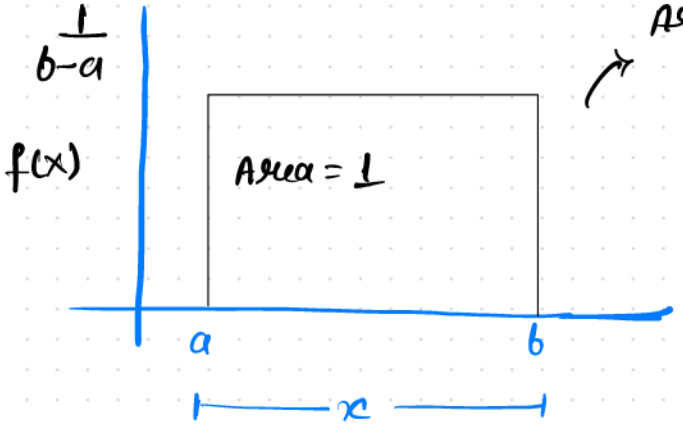
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}$$

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

Distributions

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

Continuous uniform distribution :-



Area = b × l

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

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

So, the PDF of the uniform distribution

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

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

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

↳ you don't need to know how it came!

★ AUC of the uniform distribution is just a rect.

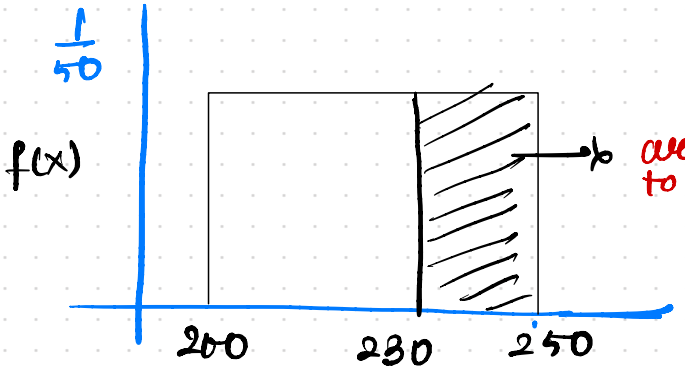
Example (Statistic refo.) :-

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

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

$$= 0 \quad \text{otherwise}$$

What is $P(X > 230)$?



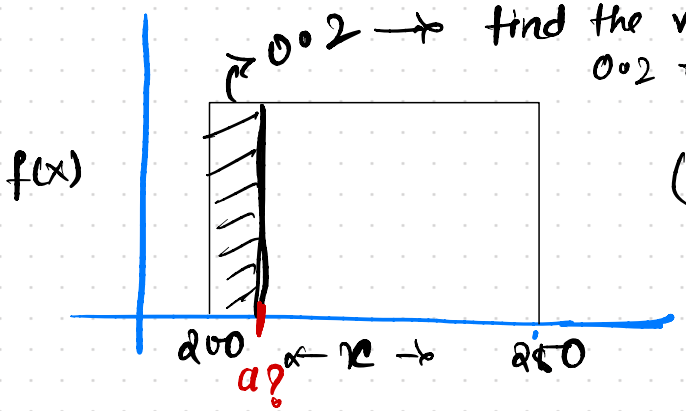
area under the curve to the right of 230.

↓

this is simply a rect.

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

What is the 20th percentile ?



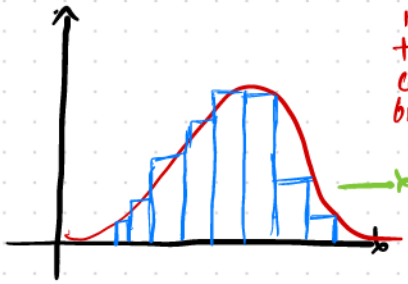
find the value a which yields 0.2 to the left ?

$$(a - 200) * \frac{1}{50} = 0.2$$

$$= 210$$

Normal Distribution :-

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



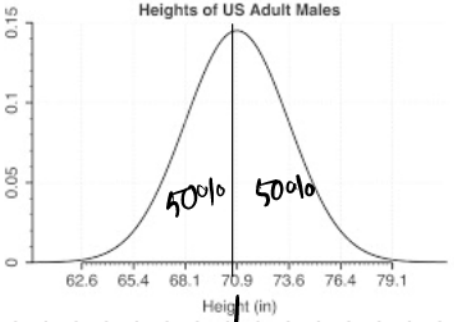
→ There are many cases in which data tends to be around a central value with no bias left or right.

→ Normal distribution ('Bell curve')

Many such examples follow N.D :-

→ Height of people

When we say that data is Normally distributed :-



→ mean = median = mode

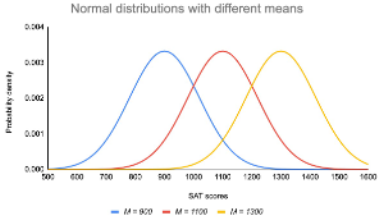
→ Symmetric about the centre

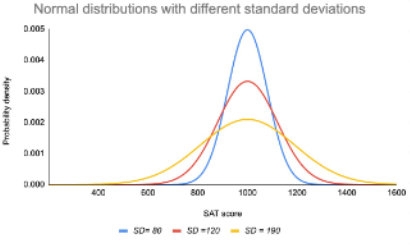
→ 50% of values are less than mean & 50% of values are > than mean.

↳ mean = median = mode

All N.D can be described by 2 pars :- 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^2)

* $N(\mu = 0, \sigma = 1)$ → Normal distribution has mean μ and standard deviation σ.

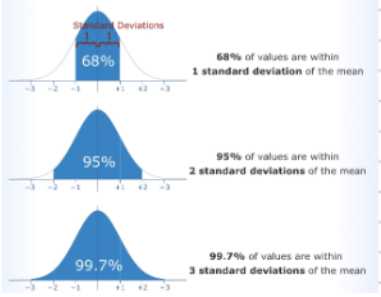
if mean = 0 and standard deviation is 1, then it's called standard Normal distribution.

Standarizing with z-scores :-

$$z\text{-score} = \frac{x - \mu}{\sigma}$$

Empirical Rule :-

→ The empirical rule gives us how much of the data lies within one, two or three std. deviation 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**:

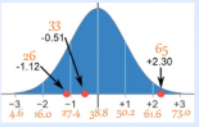
- first subtract the mean: 26 − 38.8 = −12.8,
- then divide by the Standard Deviation: −12.8/11.4 = −1.12

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

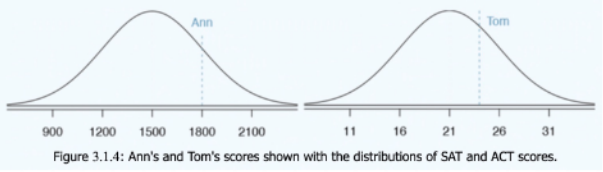


Figure 3.1.4: Ann's and Tom's scores shown with the distributions of SAT and ACT scores.

→ Ann is 1 Std. deviation above from average.
→ Tom's is 0.6 Std. deviation above from average.

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

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

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

* 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

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 nth trial is given by

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

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

$p = 0.4$

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

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

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

here, μ & expected value are same. On avg. it takes $1/p$ to get success under the geometric distribution.

→ if the P(success) is ↑ then you don't have to wait for long for a success.
eg : P(success) = 0.8 ; $\frac{1}{0.8}$ = 1.25 trials on avg.

→ if the P(success) is ↓ then you do have to wait for long for a success.
eg : P(success) = 0.1 ; $\frac{1}{0.1}$ = 10 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 examples of null hypotheses that are generally accepted as being true are :-

1. Taking Viorx 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.