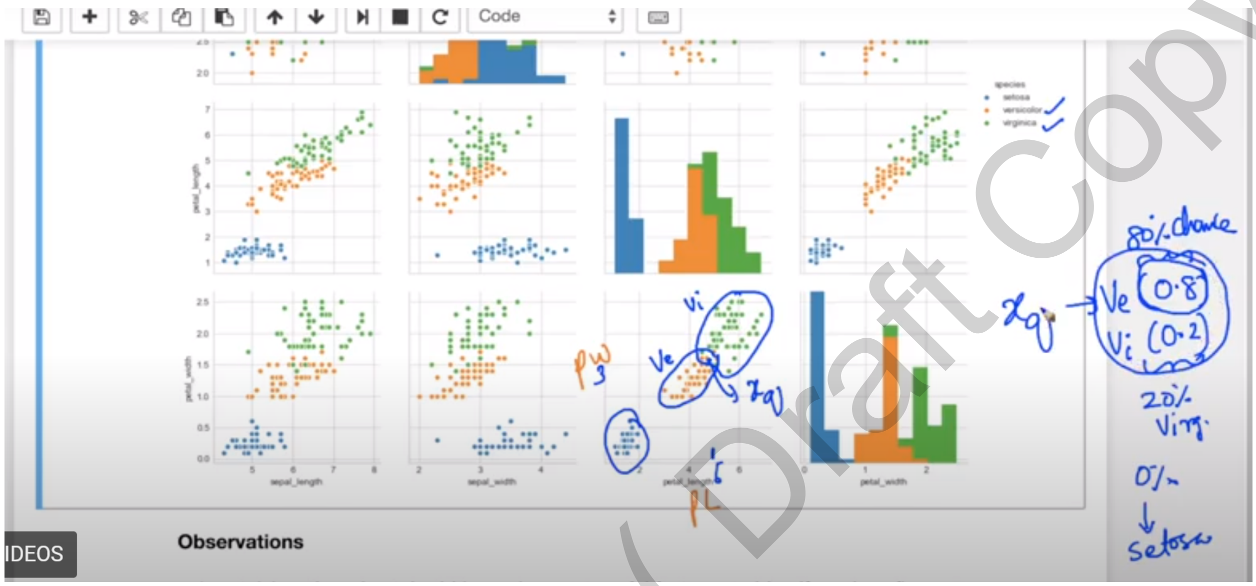
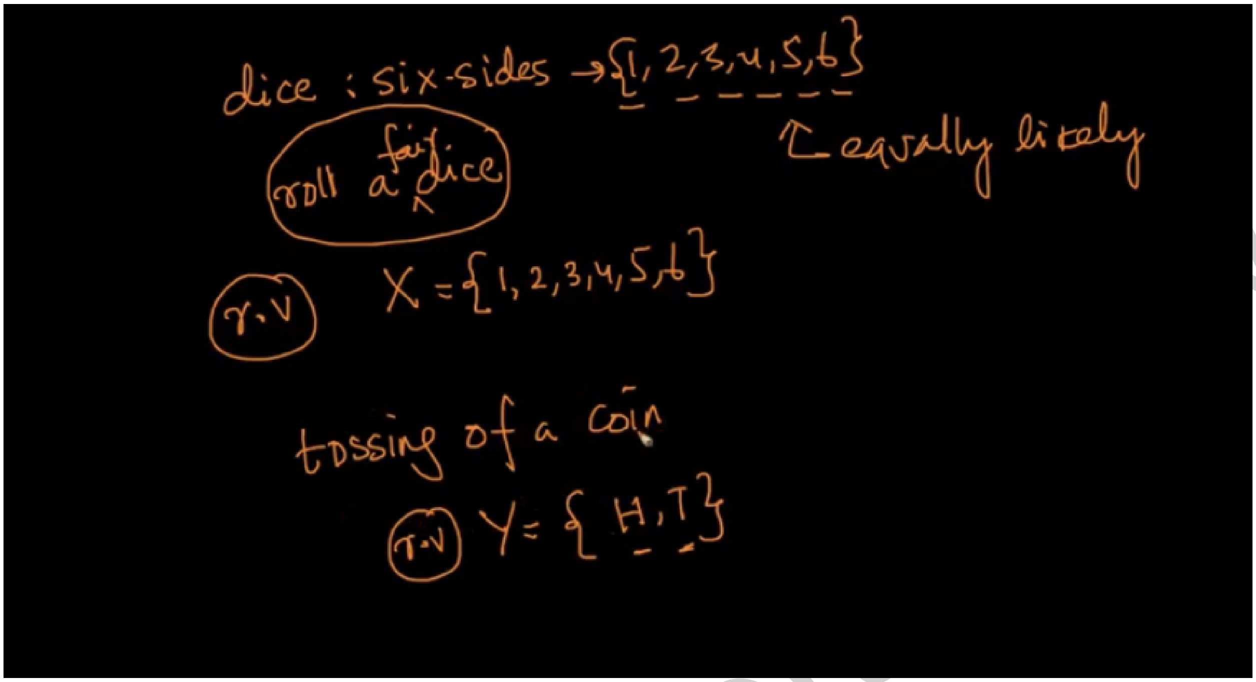
21.2 Introduction to Probability and Statistics

Timestamp 3.41

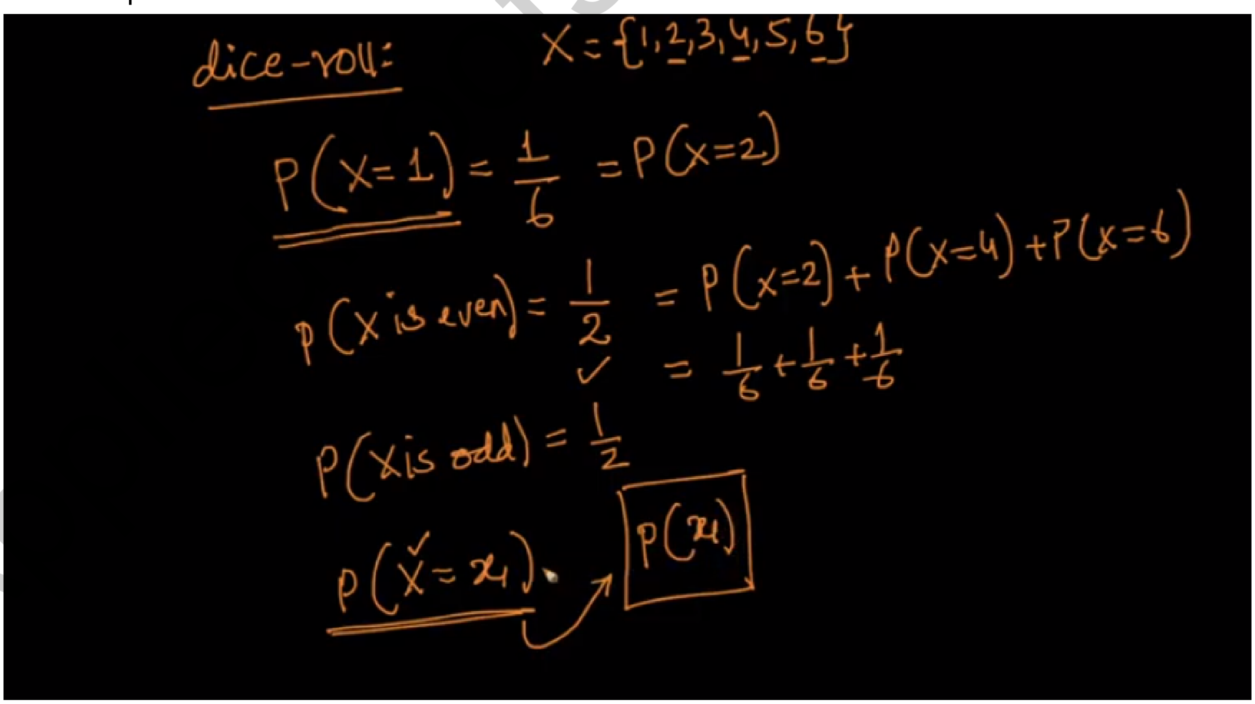


1.As shown above if we consider a query datapoint Xq ,using probability we will be able to describe there is 0.8 probability for it to be versicolor and 0.2 probability for it to be virginica and 0 probability for setosa .Lets understand the probability and statistics with respect to Machine Learning



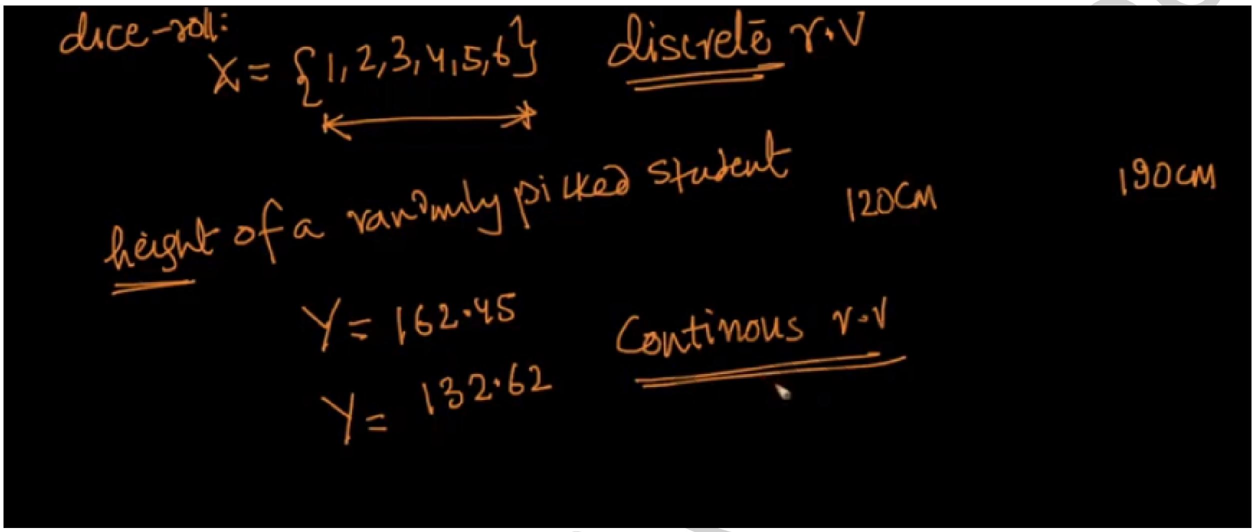
1. In an experiment of rolling a fair dice, random variable X can take any of the values from set {1,2,3,4,5,6} and is equally likely.
2. Similarly tossing a fair coin random variable y can take any of values from set {H,T} and is equally likely.

Timestamp 11.12



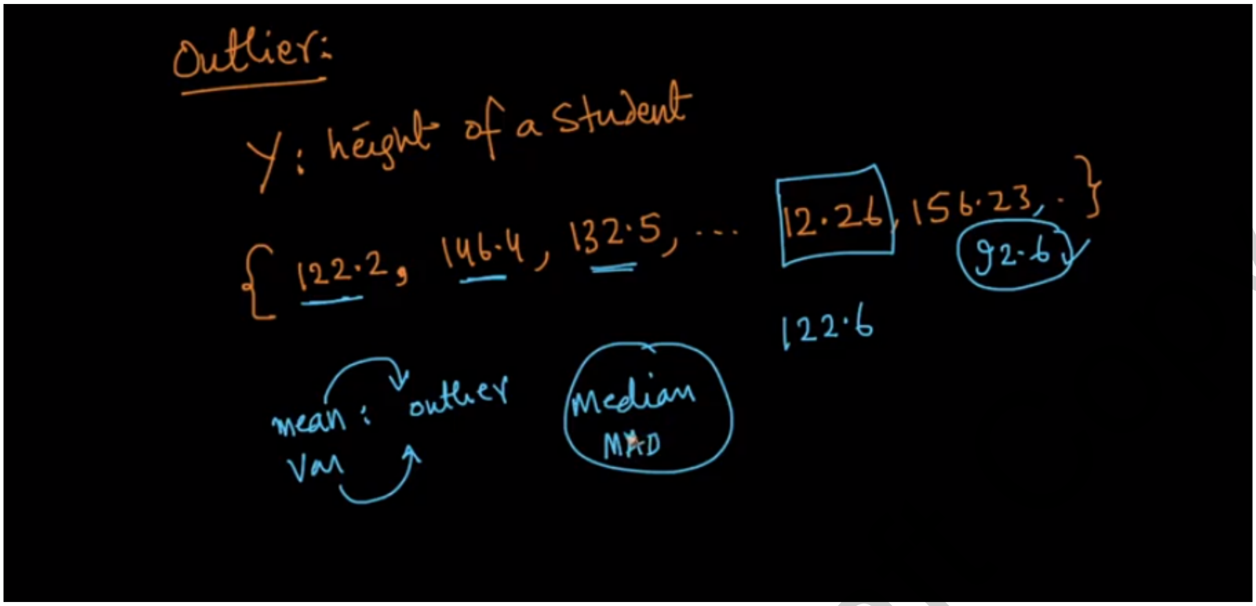
1. Probability of random variable x taking a value 1 when a dice is rolled is ⅙ and is same as probability of x taking value 2
2. When a fair dice is rolled the probability of random variable x taking an even number is ½ as shown above which is the same as x taking an odd number.

Timestamp 12.11



1. There are two types of random variables,consider an experiment where a dice is rolled a random variable x can take any value in set{1,2,3,4,5,6}.Another experiment where the height of a randomly picked student lies between 120-160, random variable y can take a real value in the range.
2. A random variable x which can take one value in a finite set of outcomes(x can take a value from a discrete set) is a discrete random variable.
3. A random variable y which can take real values is called a continuous random variable

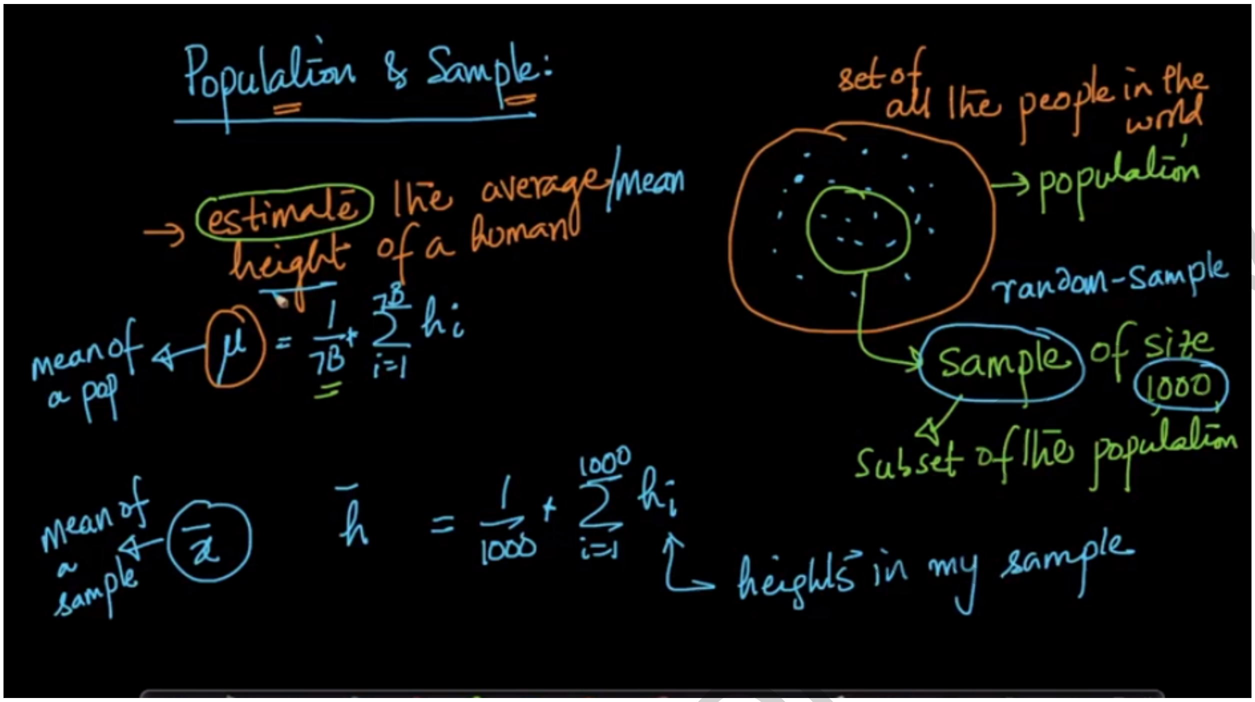
Timestamp 14.4

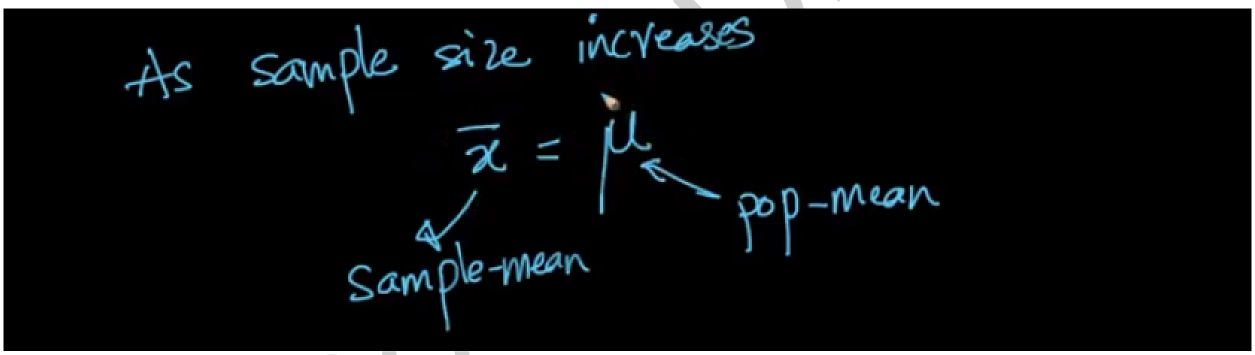


1. Let's say we have collected the observations of random variable y(heights of students)as shown above.
2. If there is an observation which is too small or too large out of all other observations such as observation is called an outlier.Outlier can happen because of amy reason it may be a human error,observation error or it could be genuine outlier,but outliers can corrupt our analysis as mean and variance are very much impacted by outliers.

21.3 Population and Sample

Timestamp 4.21

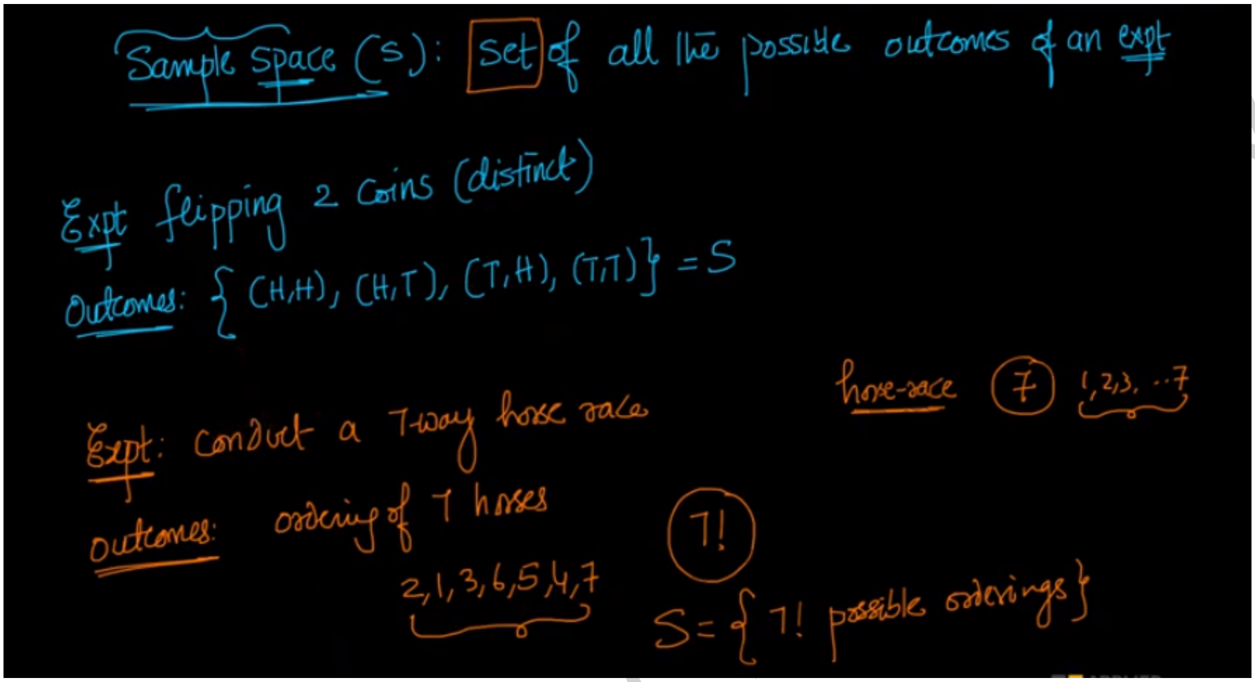


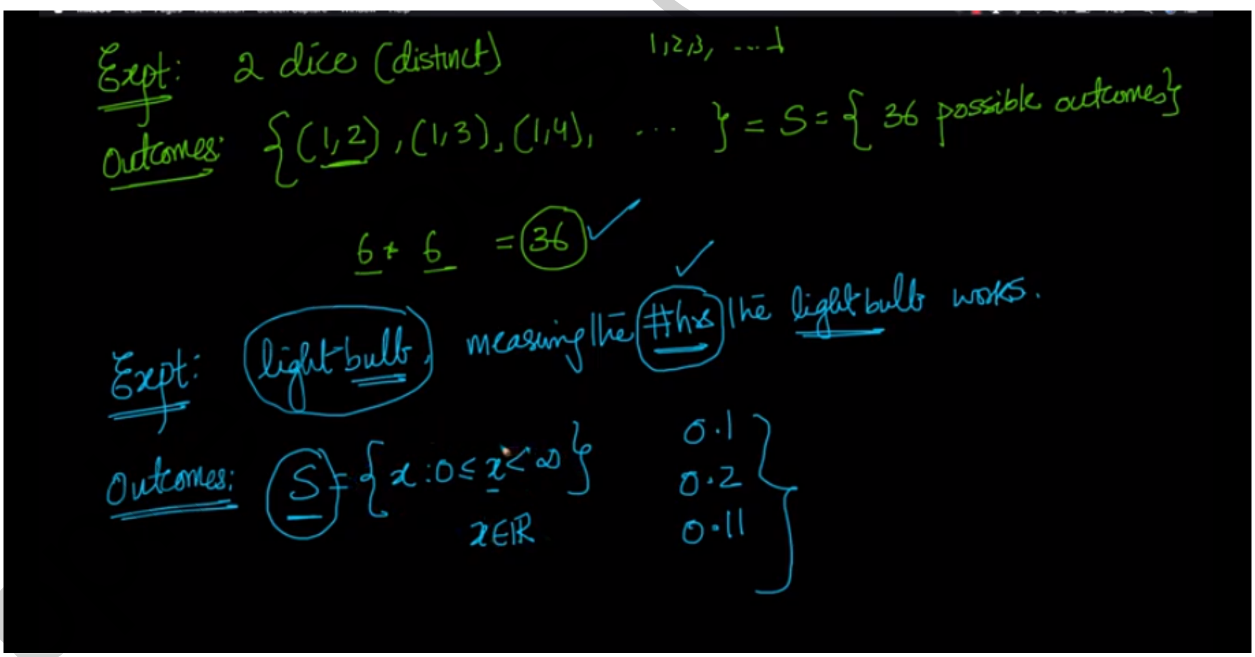


1. Let's assume set of all people in the world and it's the population,now we want to estimate the average /mean height of the humans as shown above. We cannot find the heights of all the people in the population inorder to find the mean height of the population (μ) so we are going to estimate the mean height using a sample(say a sample of size 1000).
2. Sample is a subset of population(if sample is collected randomly such a sample is called random sample),now we can find the mean height of the sample as shown.
3. As the size of sample increases, sample will be equal to population mean.
4. Population is set of all events or observations whereas a sample is a subset of population which we use to estimate some statistic or some property of population .Here height is the property of population we have estimated.

21.4 Probability from a set theoretic view

Timestamp 9,45



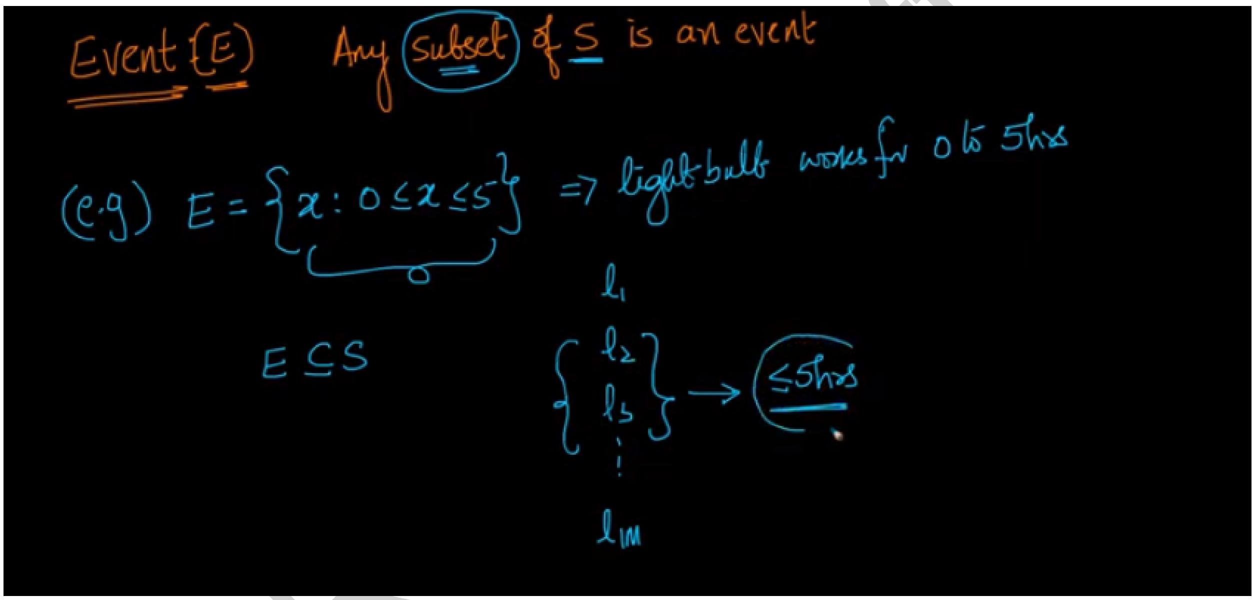


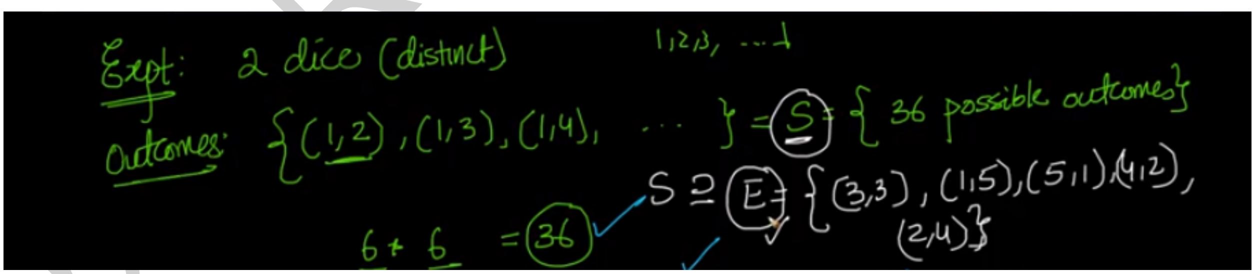
1. Let's study probability from set theoretic view
2. Let's say an experiment of flipping 2 coins the possible outcomes are as shown above.Set of all possible outcomes if an experiment is called the sample space(S) .

3. Imagine another experiment where we conduct a seven horse race ,the set of all possible outcomes are the ordering of horses after the race is sample space i.e) 7! possible orderings is the sample space
4. In experiment if throwing two distinct dice the set of all possible outcomes are 36(6*6)which is sample space.
5. Consider another experiment where we have a light bulb and we will be measuring the number of hours the light bulb works .In this case our possible outcomes lie in range **0 ≤ X <**∞ which is sample space.

The possible outcomes need not be discrete finite sets; it can also be an infinite set of outcomes. Sample space is the space of all the possible outcomes that we care about as experiment is concerned.

Timestamp 16.57



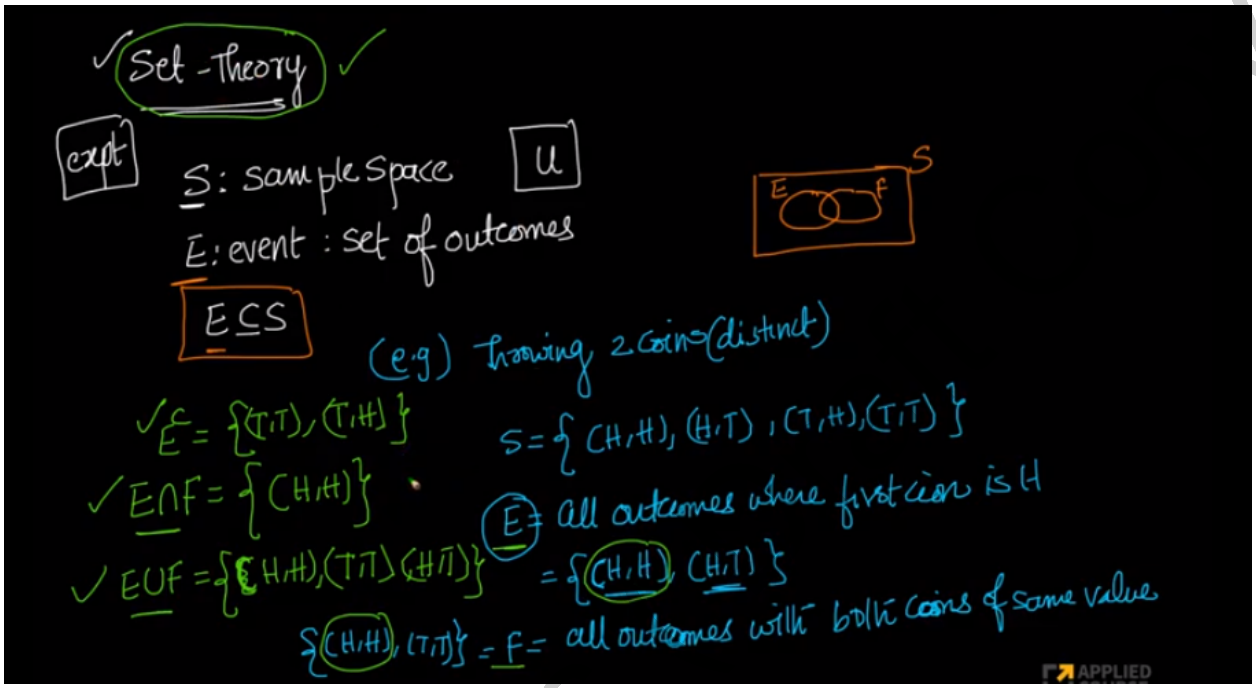


1. An event is any subset of S(Sample space)including null set.
2. Consider the experiment of the light bulb discussed above and we can have an event E such that the light bulb works for 5 hours as shown . E: **0 ≤ X ≤ 5**
3. In the experiment of throwing two dice lets define an event E such that the sum of the dice is 6   E:{(3,3),(2,4),(4,2),(1,5),(5,1)}
4. Above are all possible ways in which the event(getting sum 6 when two dice are rolled)can occur.

Similarly we can define events on experiments where E ⊆ S .

21.5 Axioms of Probability, Properties and Examples-

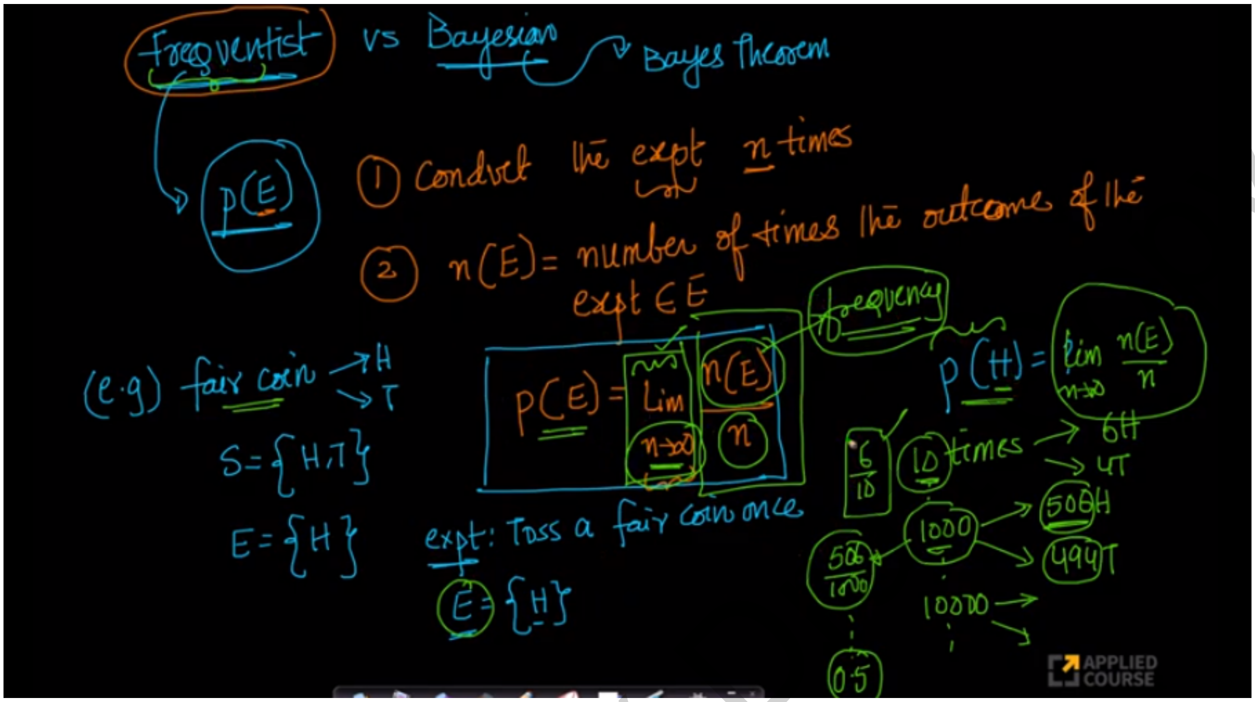
Timestamp 4.37



Lets study probability in set theoretic perspective

1. If we consider an experiment we define a sample space (S) which is a set of all outcomes.An event E which is a set of outcomes (subset of S).
2. Let's say we are flipping two coins and we have our sample space as s={HH,HF,FH,HH}
3. Consider an event E of getting outcomes where first coin is H (E={HH,HT})
4. Also consider another event F of getting outcomes where both coins are of same value
5. Now we can think of these events E and F as sets and can perform all operations which we can perform on sets. i.e)set union ,intersection,complement etc as shown above.

Timestamp 12.48



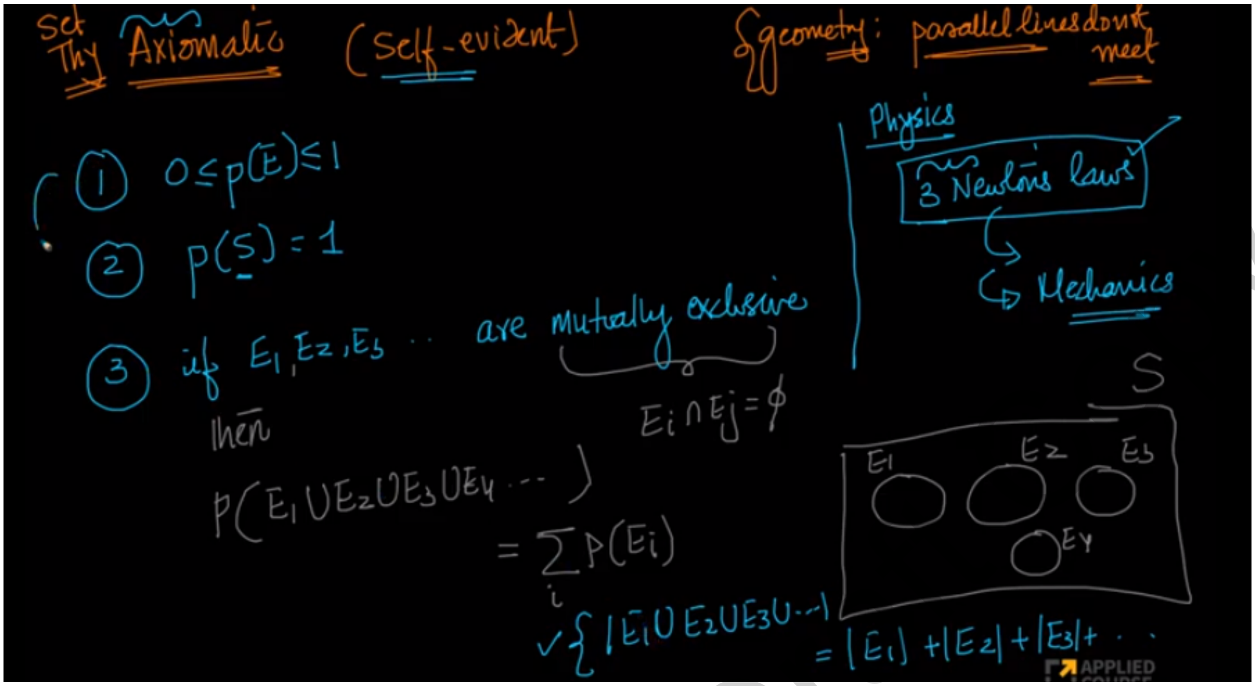
Let's understand the Frequentist and Bayesian approach of probability

From a frequentist approach probability of an event p(E) can be defined as

1. let's say we are conducting an experiment n times (assume the experiment of tossing two fair coins)
2. n(E)be the number of times the outcome of the experiment belongs to event E

Then p(E) = $\lim_{n \to \infty}$ n(E)/n

n(E) is the frequency of occurrence of the event and n is total number of times we are conducting the experiment as n tends to infinity(we conduct the experiment infinite times)

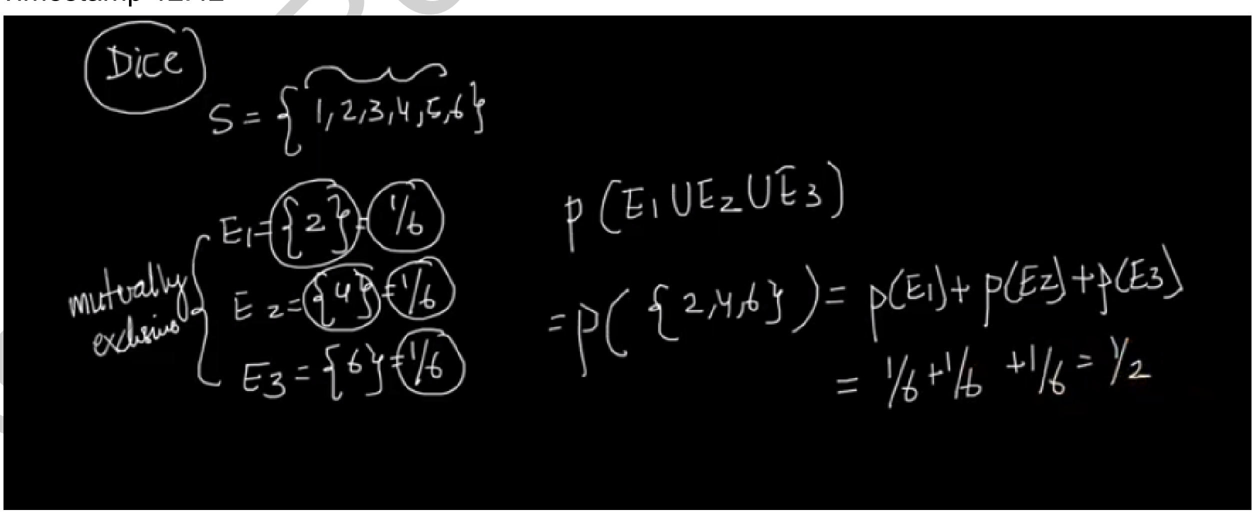


Axiomatic(self evident) approach of probability says

1. The probability of any event lies between 0 and 1 i.e) 0≤p(E)≤1
2. The probability of sample space is p(S)=1
3. If there are events E1,E2,E3…….etc which are mutually exclusive(E1∩E2 =Φ) then $p(E1UE2UE3UE1....) = \sum_{i=1} p(Ei)$

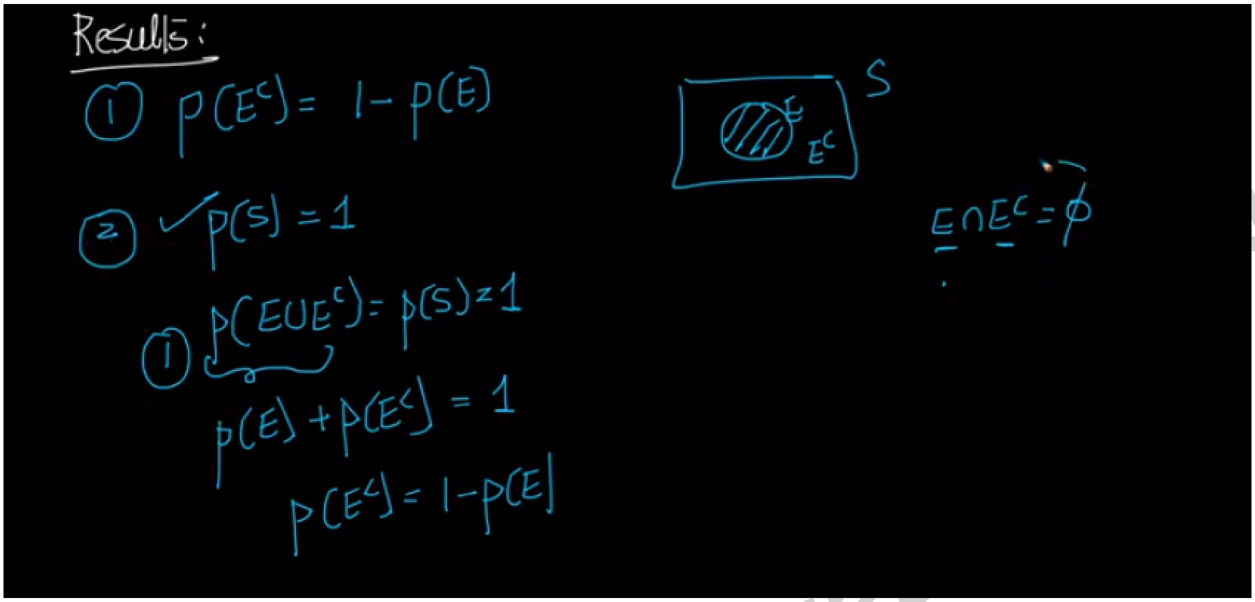
Using the above axioms and properties in set theory we can actually prove all theorems in probability

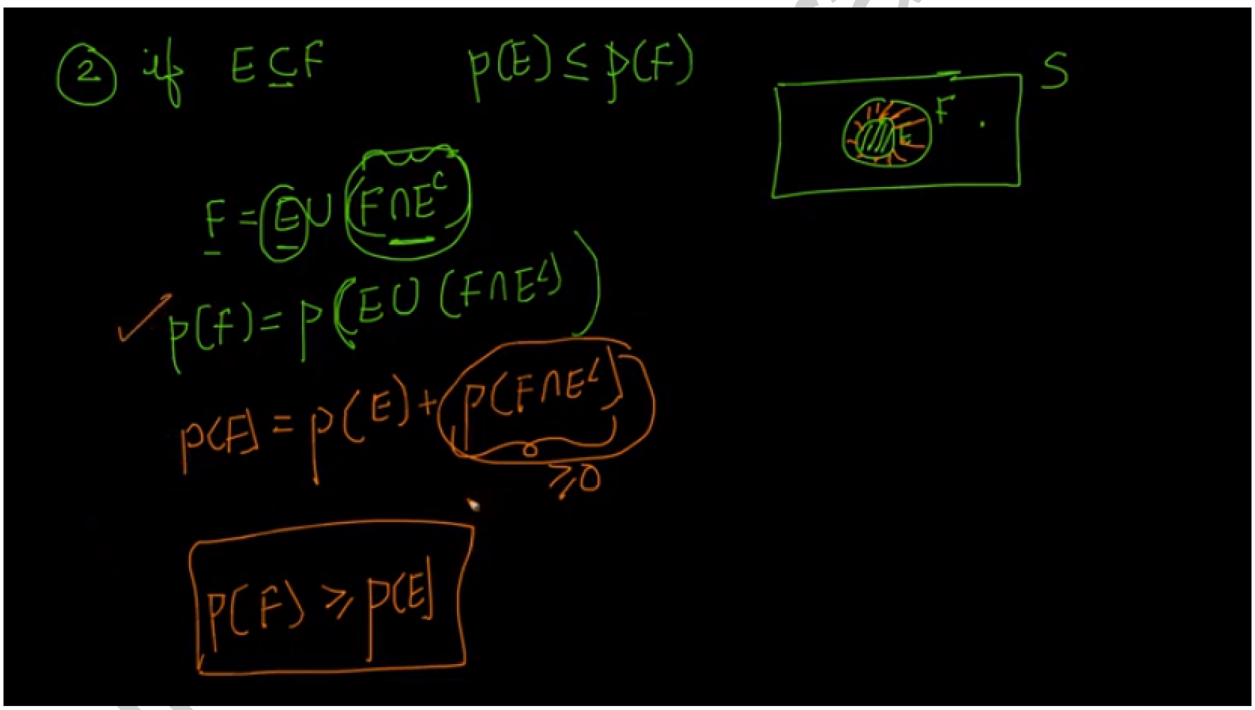
Timestamp 12.42

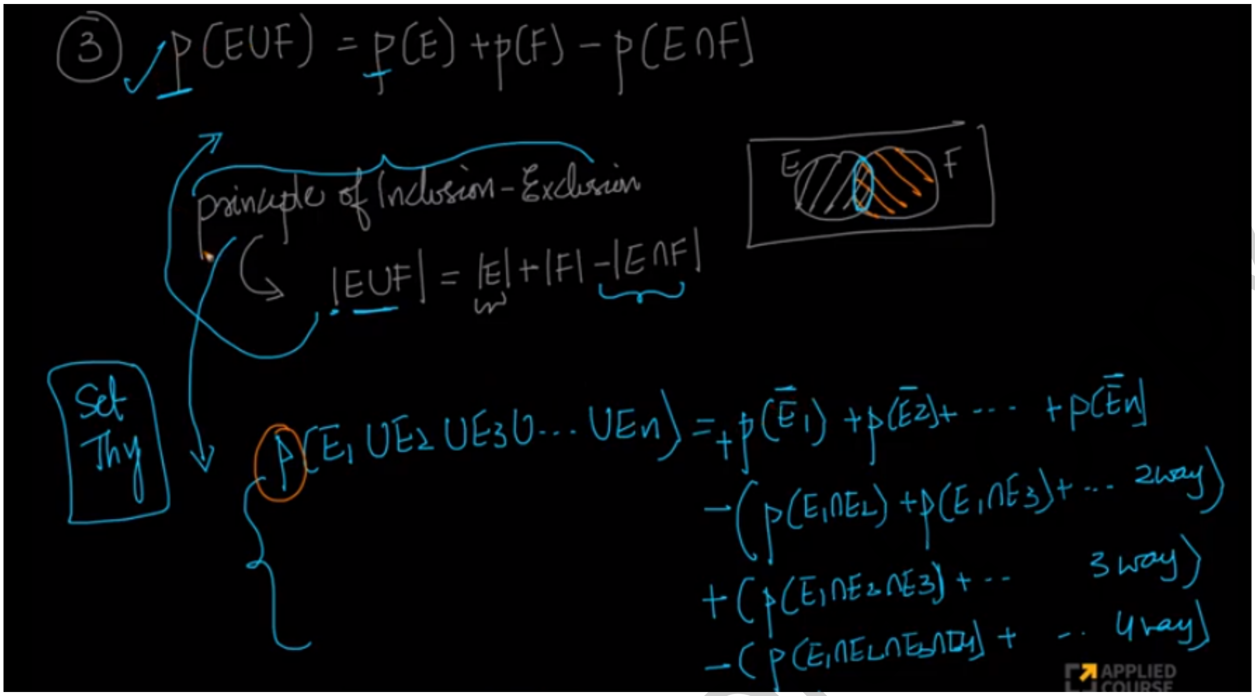


Above is an example where we have three mutually exclusive events and their union is equal to the sum of the events .

Timestamp 23.35

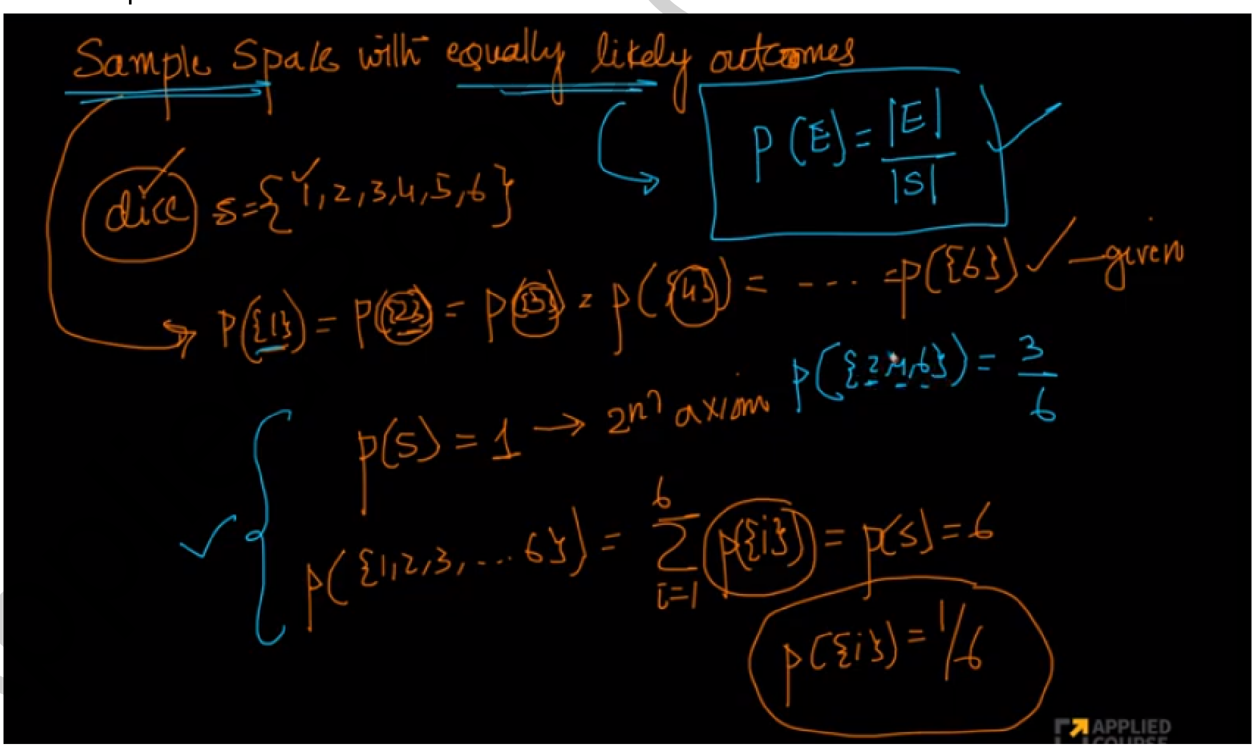






- Above are few examples of proof that we can solve all theorems in probability using these axioms and properties in set theory

Timestamp 34.43



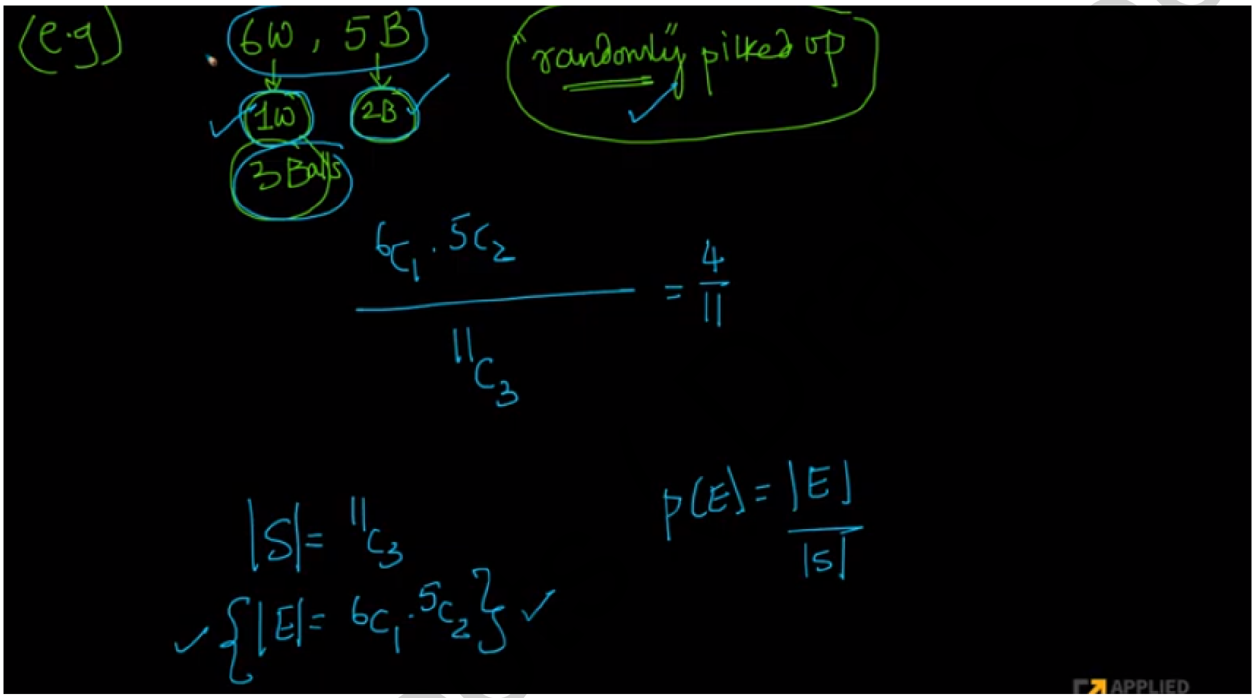
- In an experiment of rolling a dice,we have a sample space with equally likely outcomes as shown above.

S={1,2,3,4,5,6}

p({1})=p({2})=p({3})=p({4})=p({5})=p({6})=⅙ and We know that probability of sample space is p(S)=1

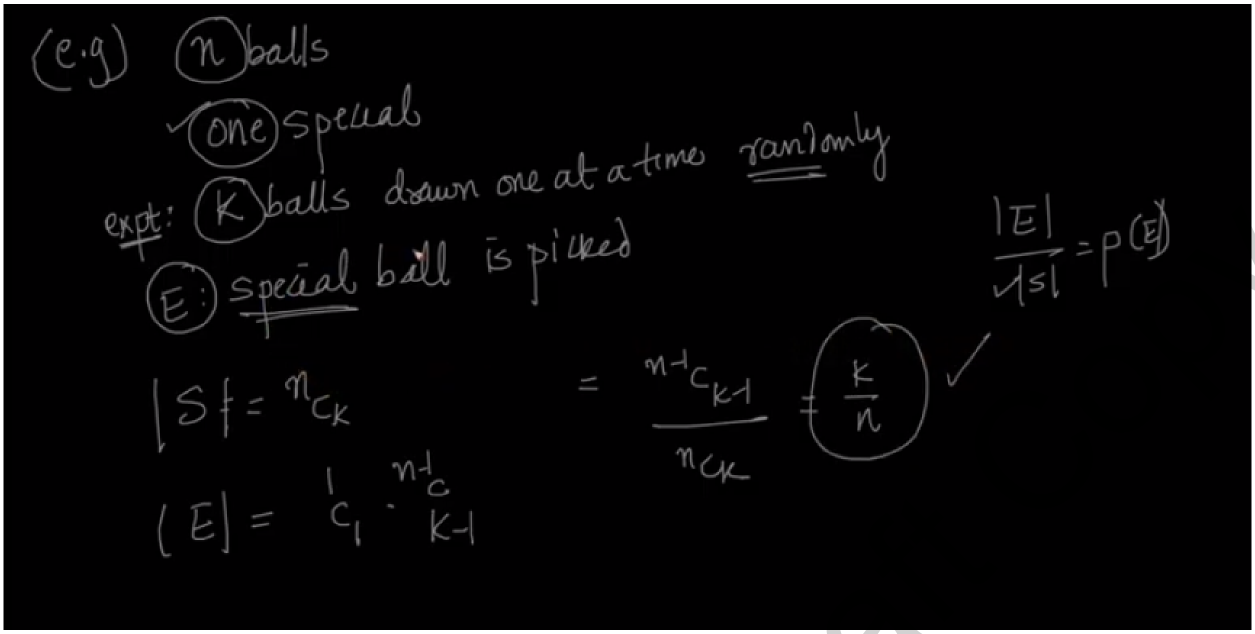
- Hence,p(E)=|E|/|S| this is true for sample space with equally likely events

Timestamp

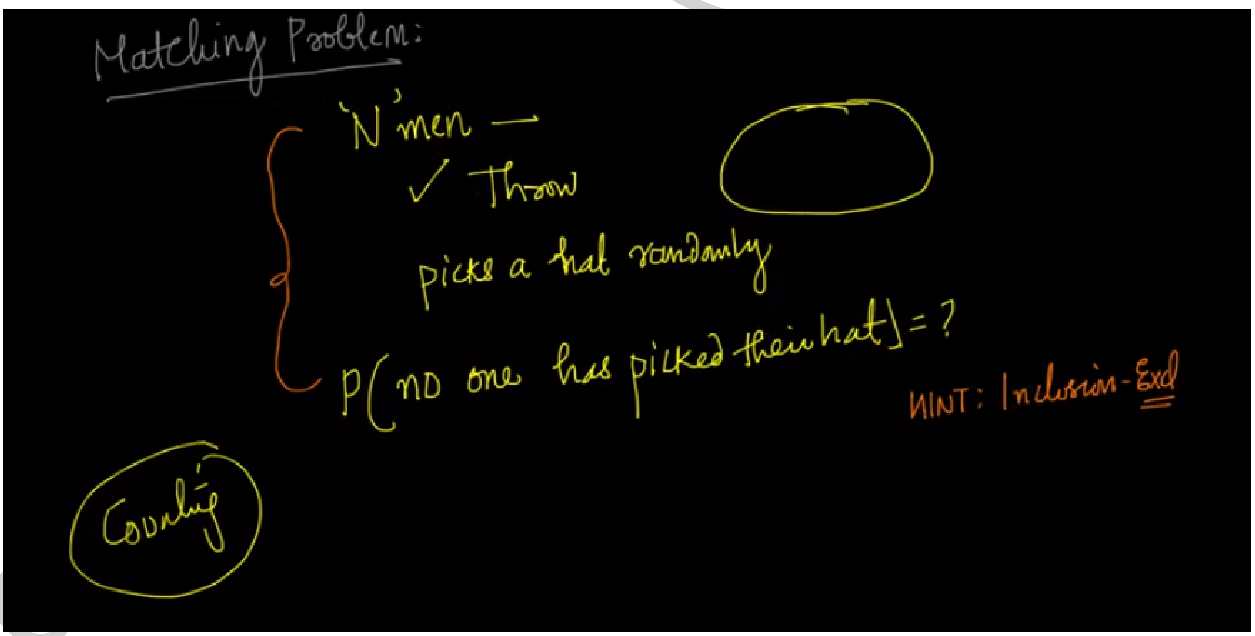


- In the above example we have to find probability that we have to choose 3 balls out of 11 balls such that we need to choose exactly 1 white and 2 black.Out of 11 balls we have 6 are white and 5 are black.

- As shown above using combinatorics we can find the probability is 4/11



1. Above example we have n balls out of which one ball is special.lets say in an experiment of drawing k balls out of n balls randomly.
2. Let say event E where special ball is picked
3. p(E) can be calculated using combinatorics as shown above



- In a party where N men throw their hats into the bag and then each of them picks a hat randomly.We have to find probability that No one picks their own hat.It can be solved using principle of inclusion and exclusion as shown below

$E_i = i^{th}$ person has picked the correct hat

$P(E_1 \cup E_2 \cup \dots E_n)$ = prob. of atleast one person picking their hat

$1 - P(E_1 \cup E_2 \dots \cup E_n)$ = prob of no person picking their hat

$$P(E_1) + P(E_2) + \dots P(E_n) \rightarrow \frac{1}{n} + \frac{1}{n} + \dots \frac{1}{n} = \frac{n}{n} = 1/1!$$

$$-\left(P(E_1 \cap E_2) + P(E_1 \cap E_3) + \dots\right) \rightarrow \frac{1}{2!}$$

$$\left(\frac{1}{n} \cdot \frac{1}{n-1}\right) {}^nC_2 = \frac{\cancel{n \times n-1}}{2!} \times \frac{1}{\cancel{n}} \cdot \frac{1}{\cancel{n-1}}$$

$$+\left(P(E_1 \cap E_2 \cap E_3) + \dots\right) \rightarrow \frac{1}{\cancel{n}} \cdot \frac{1}{\cancel{n-1}} \cdot \frac{1}{\cancel{n-2}} \cdot \frac{\cancel{n}\,\cancel{n-1}\,\cancel{n-2}}{3!} = \frac{1}{3!}$$

P 3

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots$$

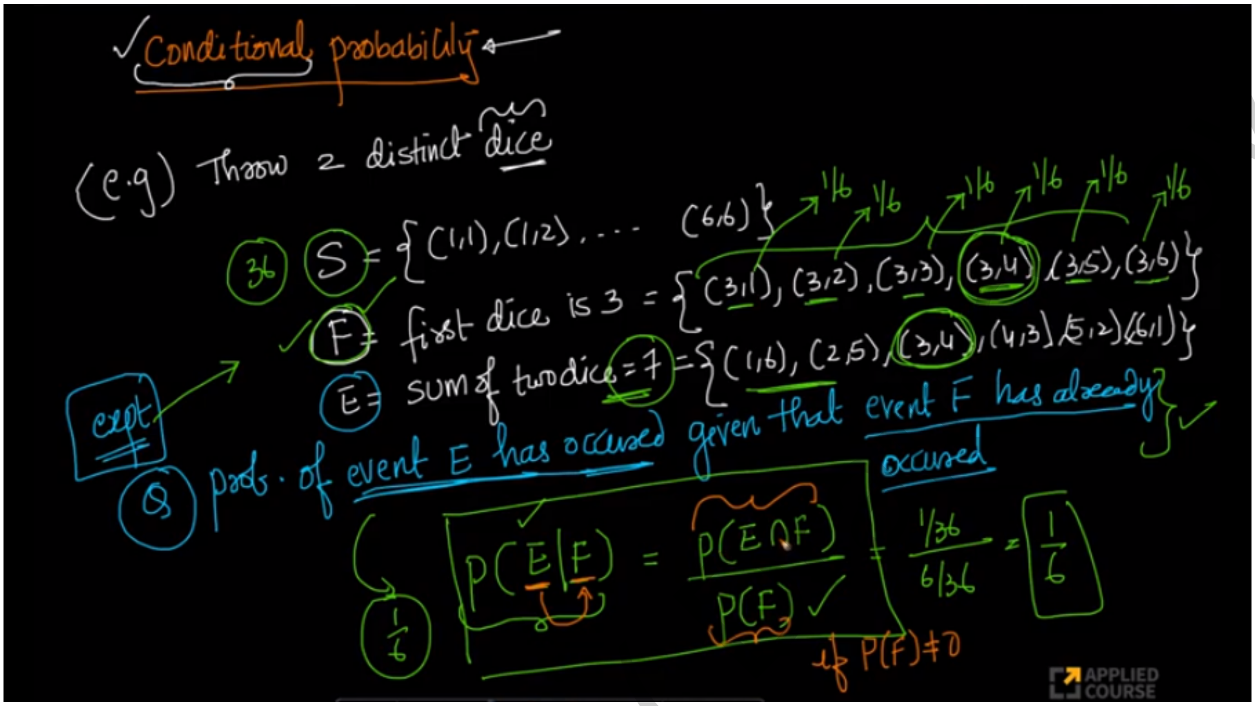
$$1 - P\left(\bigcup_{i=1}^{n} E_i\right) = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} + \dots \pm \frac{1}{n!}$$

∞ - series

if $n = \infty$ $\quad e^{-1} = 0.36788$

21.6 Conditional Probability & Examples

Timestamp



Let's try to understand conditional probability using an example

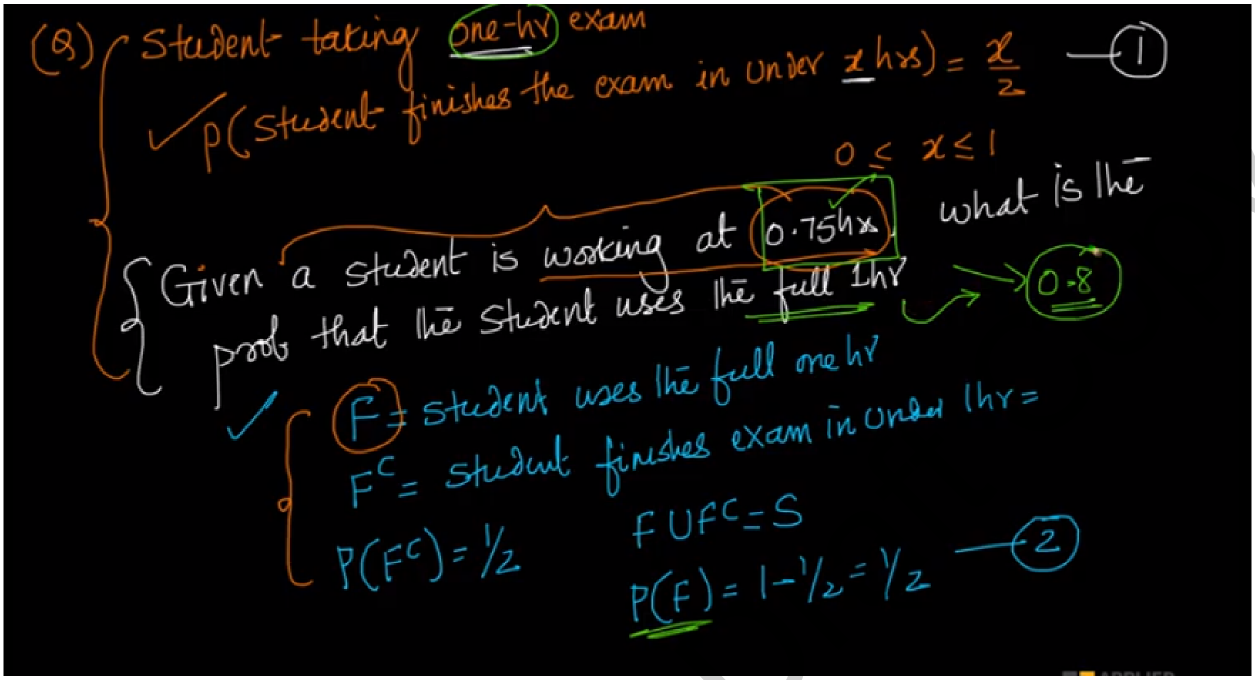
- Consider an experiment of throwing 2 dice,we have Sample space S={(1,1),(1,2)........(6,6)}
- Let's say
  Event F where we get 3 on the first dice -{(3,1),(3,2),(3,4),(3,3),(3,5),(3,6)}
  Event E where sum of two dice is 7   -{(1,6)(2,5),(3,4),(4,3),(5,2),(6,1)}

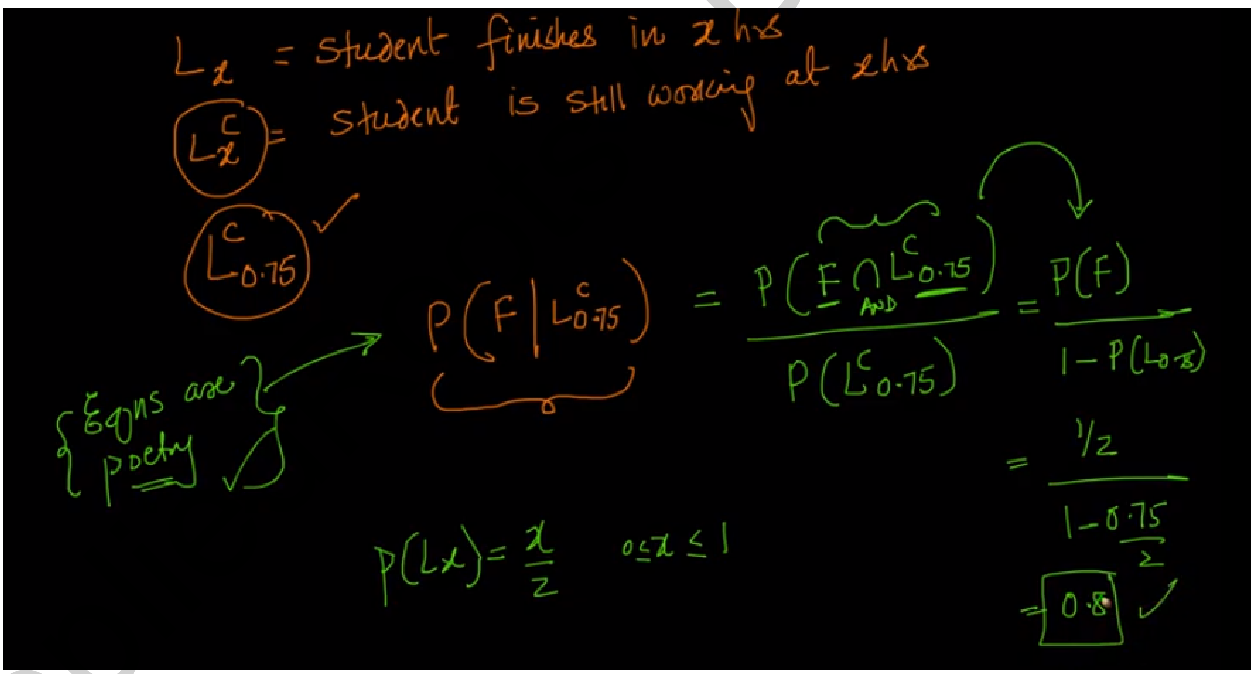
Now we can find probability of event E given that event F has already occurred using conditional probability

P(E|F) = P(E ∩ f)/P(F) as shown above

Conditioned on the fact that F has already occurred the probability of event E occurring will be given by ***P(E ∩ f)/P(F)*** and probability holds only when ***P(F)*** $\neq 0$.

Timestamp 13.30





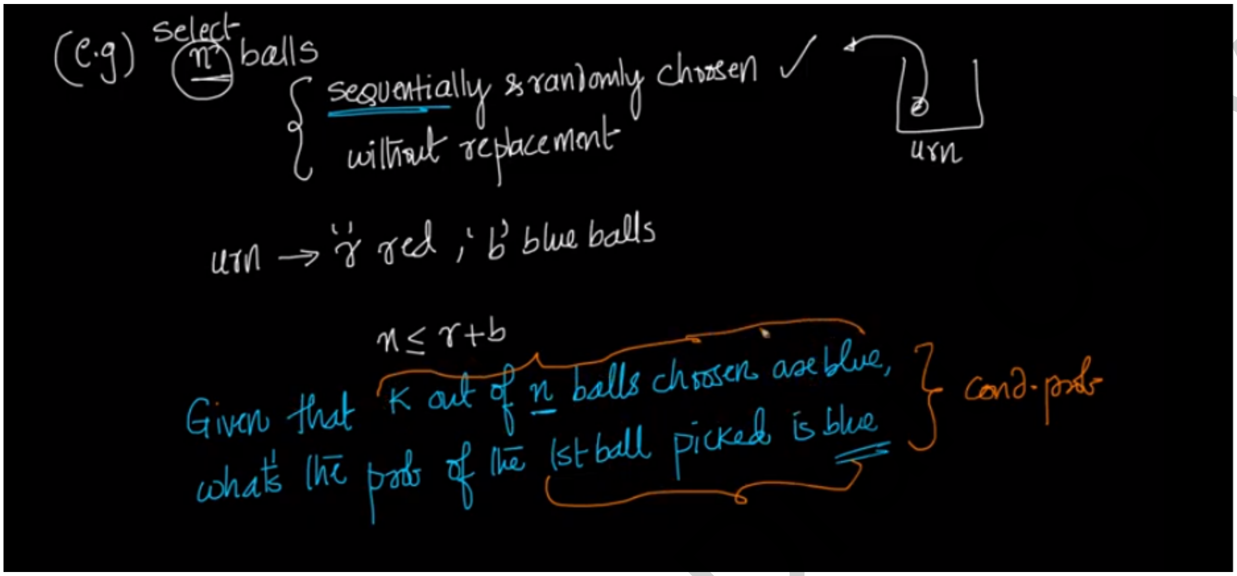
Consider the above example where a student is talking one hour exam.

- Given probability that student finishes the exam in under xhrs=x/2
- Given a student is working at 0.75 hrs We need to find the probability that the student uses full hour.

We can solve the problem as shown above using conditional probability.

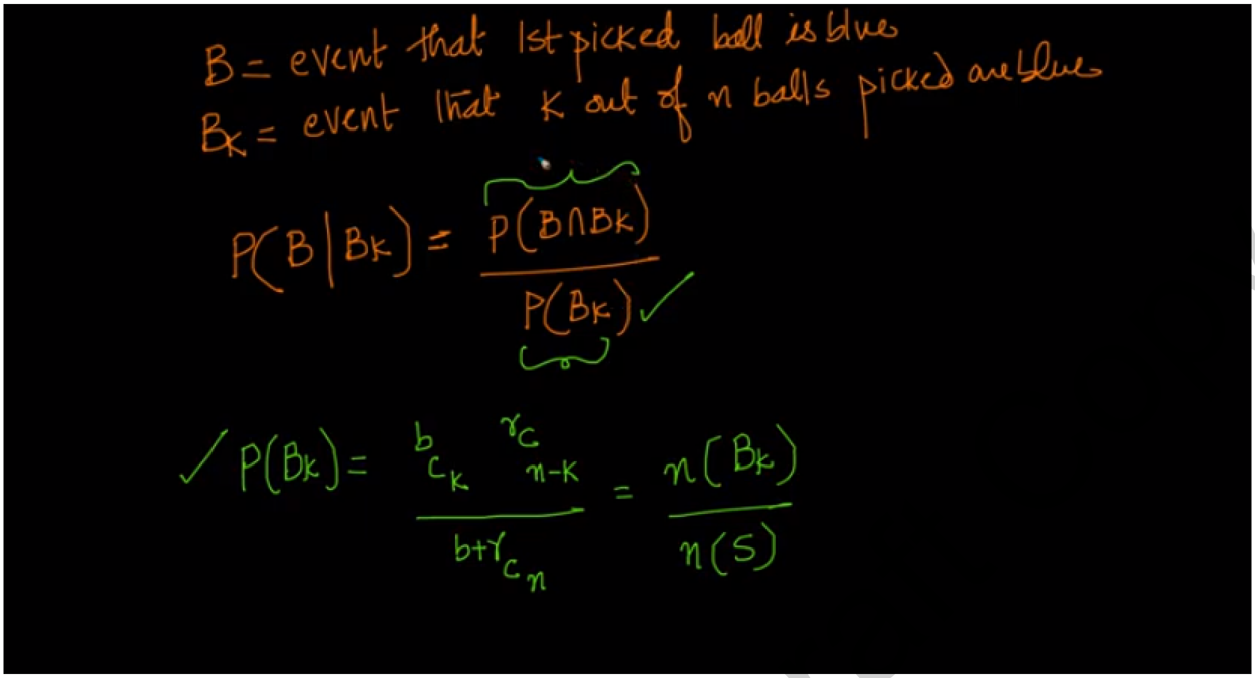
21.7 Multiplication theorem

Timestamp



- For understanding conditional probability let's consider the above example
- There are r red balls and b blue balls in an urn ,we have to select n balls sequentially and randomly chosen without replacement.

- Given that out of n balls chosen K balls are blue.We need to find the probability of the 1 st ball being picked is blue.



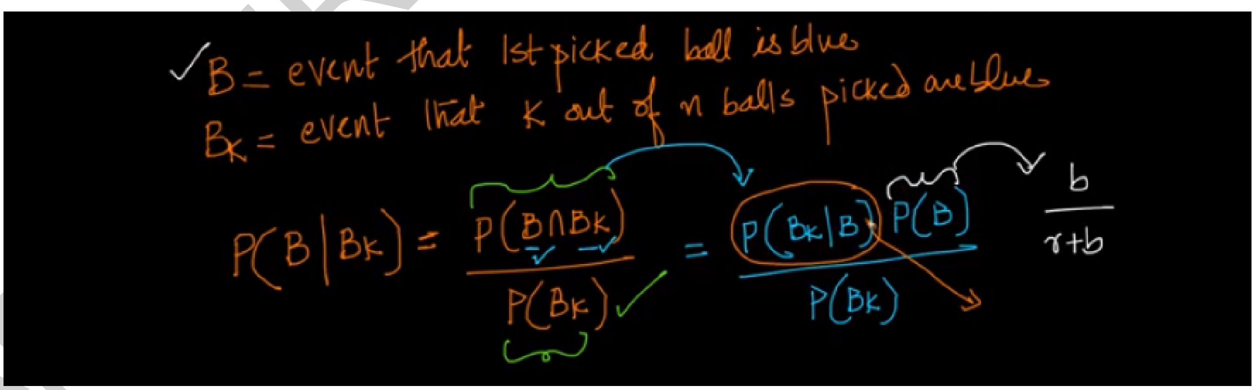
We define events as shown above and the we need to find

P(B|BK) = P(B ∩ BK)/P(BK)

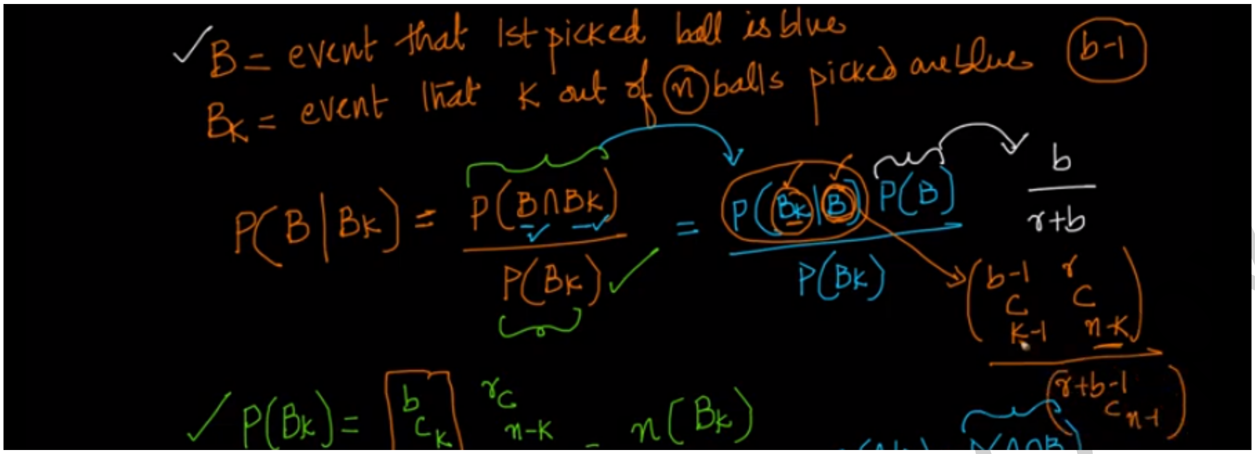
We calculate the probability of event BK (k out of n balls picked are blue)as shown above .

- We are choosing n balls out of b+r balls ,this is our sample space.
- We select k balls out of b blue balls and n-k balls out of r red balls

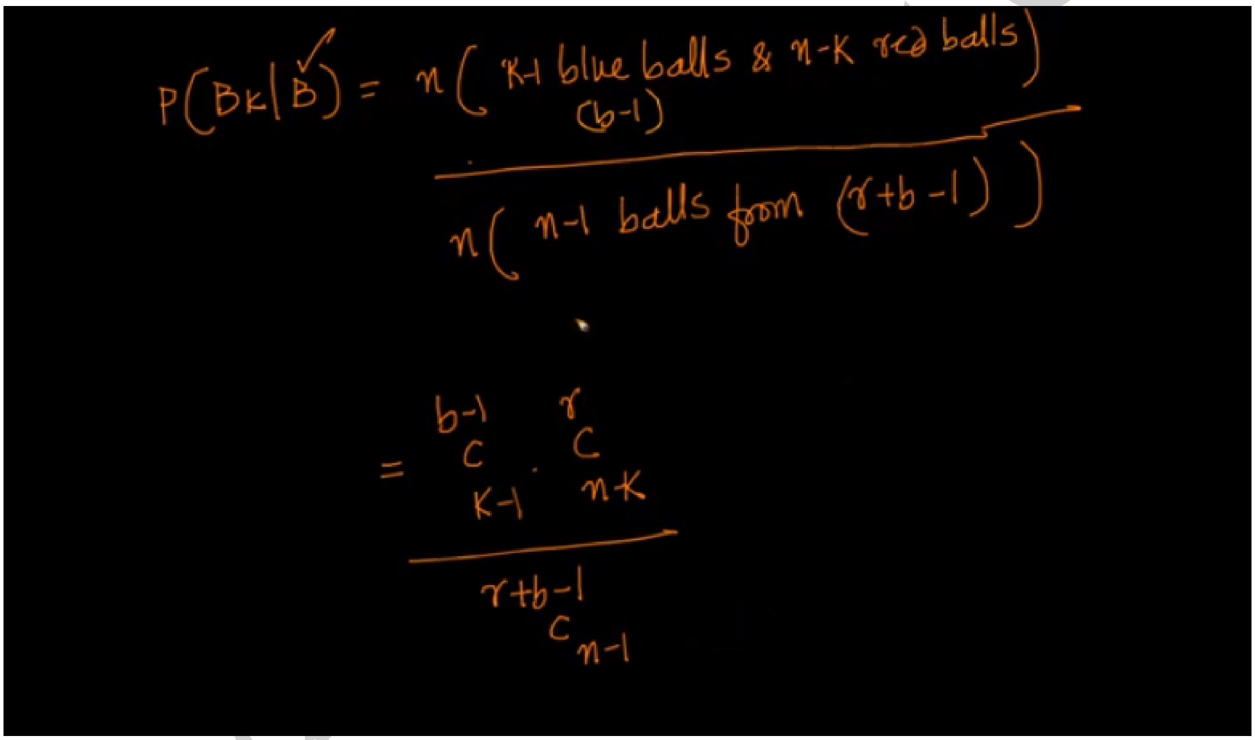
Timestamp 10.19



Using conditional probability we can rewrite ***P(B ∩ BK)*** .Now we find P(B), the probability that first picked ball is blue i.e)b/r+b as shown above.



Timestamp 16.17



- Given that we have already picked first ball blue,now we have to find the probability that we pick k blue balls out of n balls ***i.e) P(Bk / B)***
- Since we have already picked first ball blue now we choose k-1 balls out of b-1 blue balls and n-k balls out of r red balls as shown above

Now we have our numerator and denominator,we just substitute the values and we get
P(B|BK) = P(B ∩ BK)/P(BK) = k/n

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad \text{if } P(E_2) \neq 0$$

Above is the definition of conditional probability,We can think of multiplication rule as a generalization of conditional probability

Timestamp 22.44

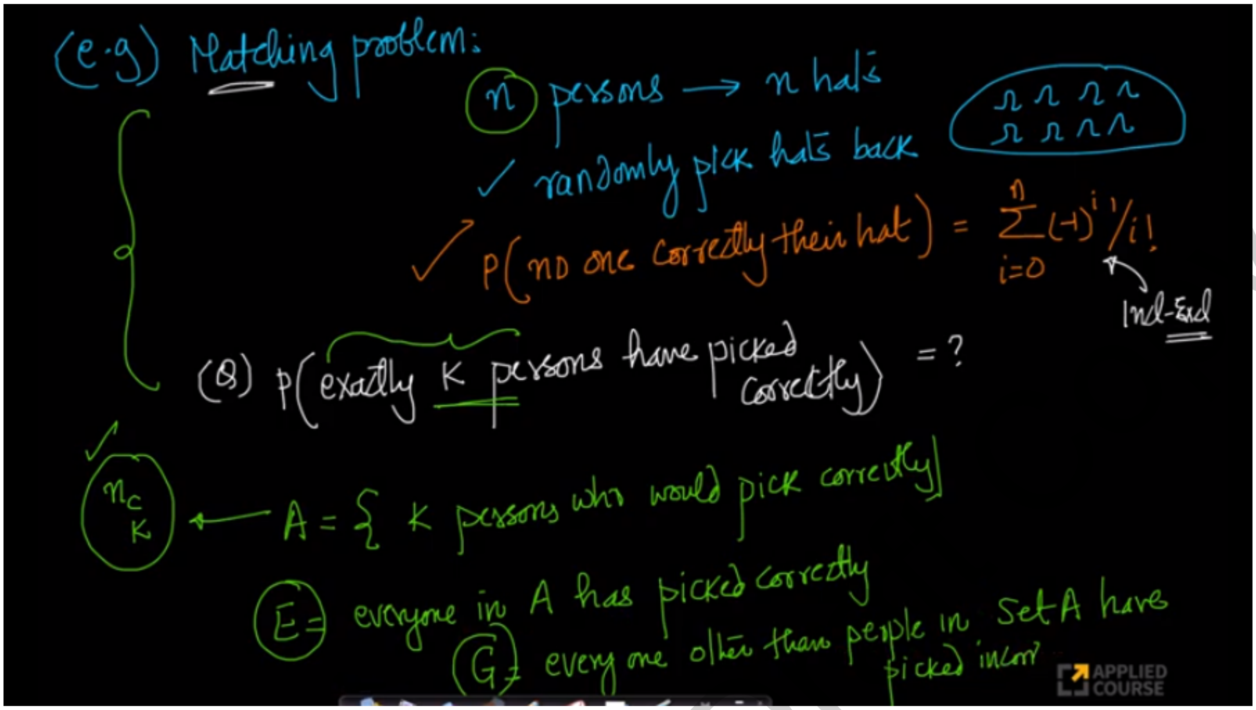
$$\Rightarrow P(E_1 \cap E_2) = P(E_1|E_2)\,P(E_2)$$

$$P(E_1 E_2) = P(E_1|E_2)\,P(E_2)$$

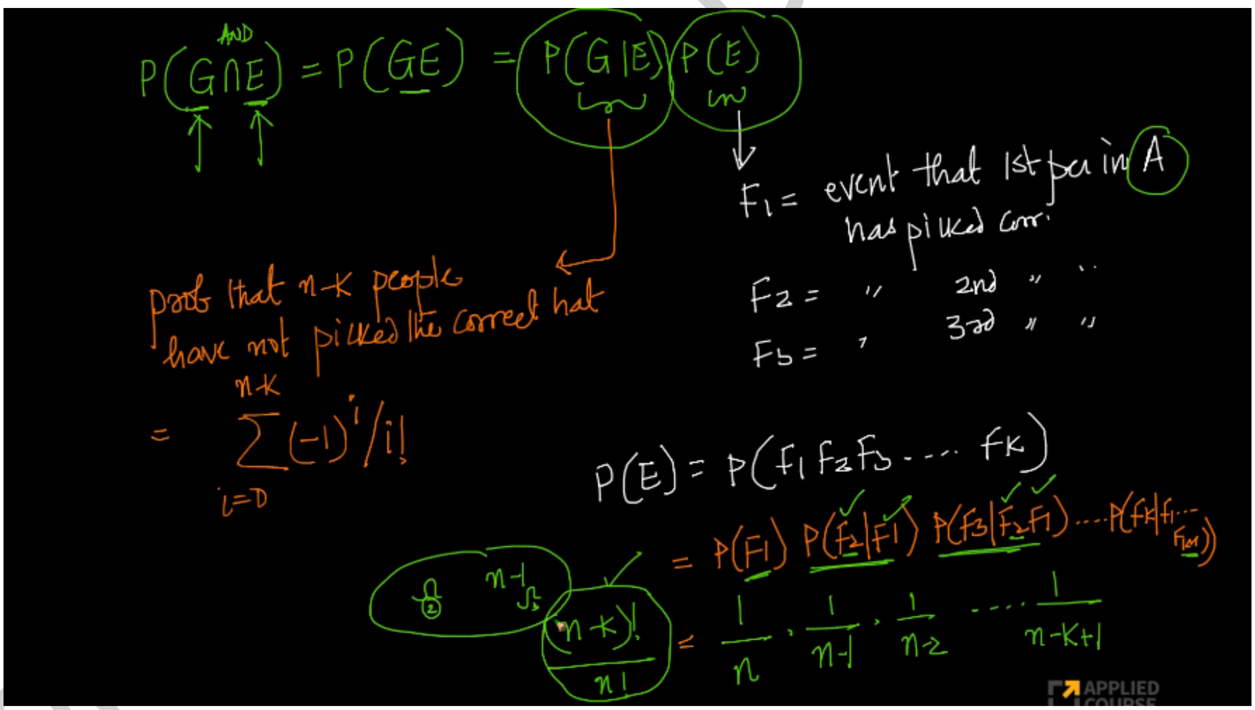
Gen:

$$P(E_1 E_2 E_3 \cdots E_n) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_2 E_1) \cdots P(E_n|E_{n-1} E_{n-2} \cdots E_1)$$

By repeatedly applying conditional probability we obtain the multiplication rule as shown above.



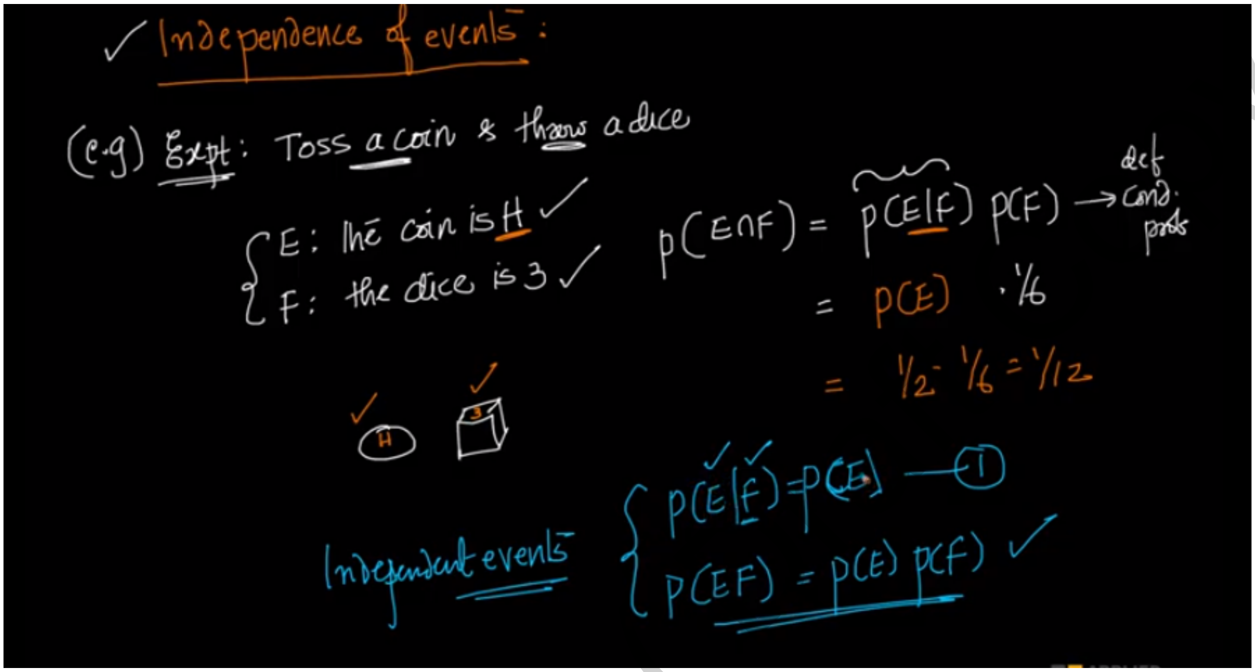
Timestamp 34.42



- Above is a problem where n persons randomly pick their hats will be solved as shown above

21.8 Independent events

Timestamp 4.54



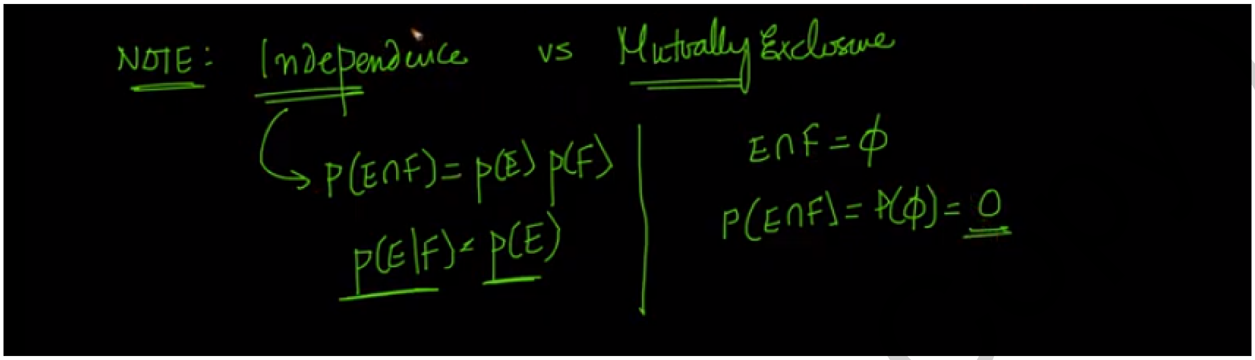
1. Consider an experiment of tossing a coin and throwing a dice.As shown above define two events E and F.
   E= probability of getting a head
   F= probability of getting 3

2. The probability of both events happening E∩F can be obtained using conditional probability.
   p(E∩F) = p(E/F) . p(F)

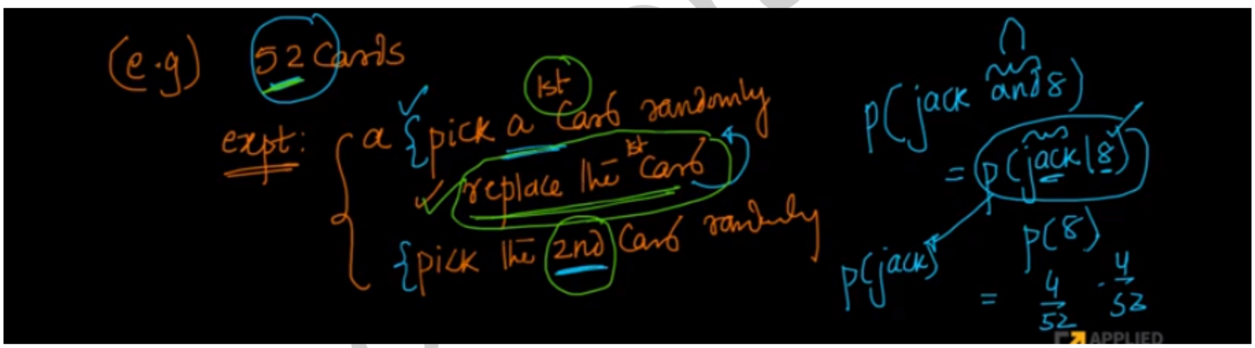
3. The fact that F has already happened has no impact on E happening.Event E happening doesn't depend on event F (E is independent if F)so p(E/F) = p(E).
   Then

   p(E∩F) = p(E) . p(F)
4. Two events are said to be independent events when their intersection is equal to the product of the product of events.

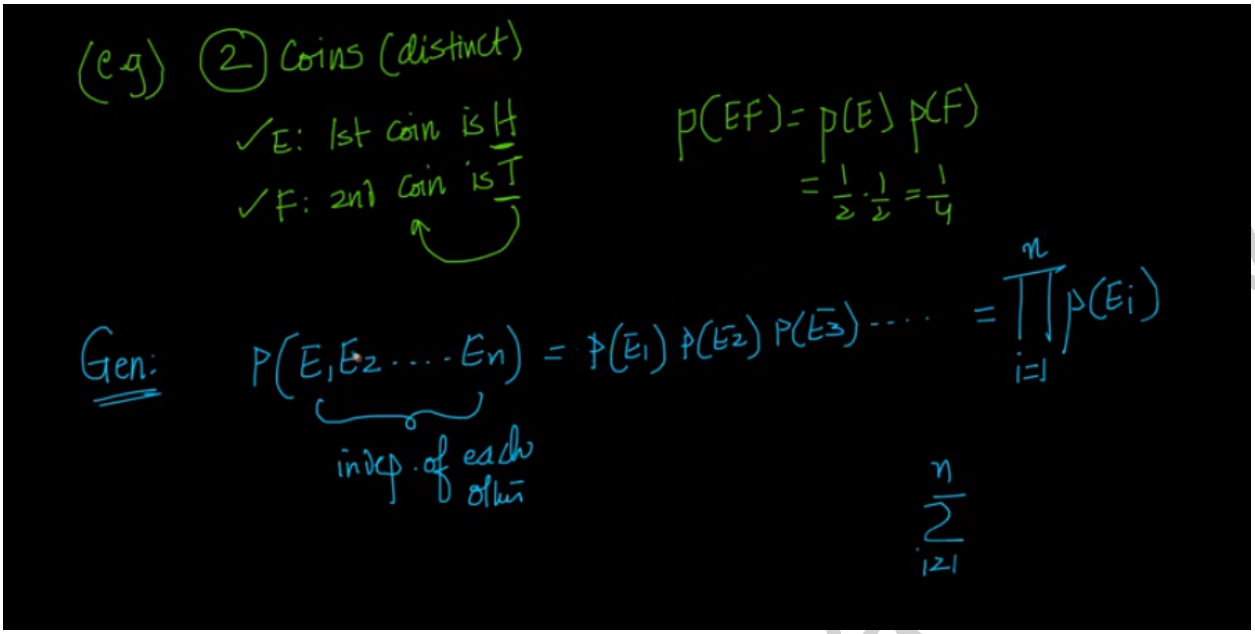
Timestamp 11.53



- For mutually exclusive events the intersection of events is null , independent events when their intersection is equal to the product of the product of events.

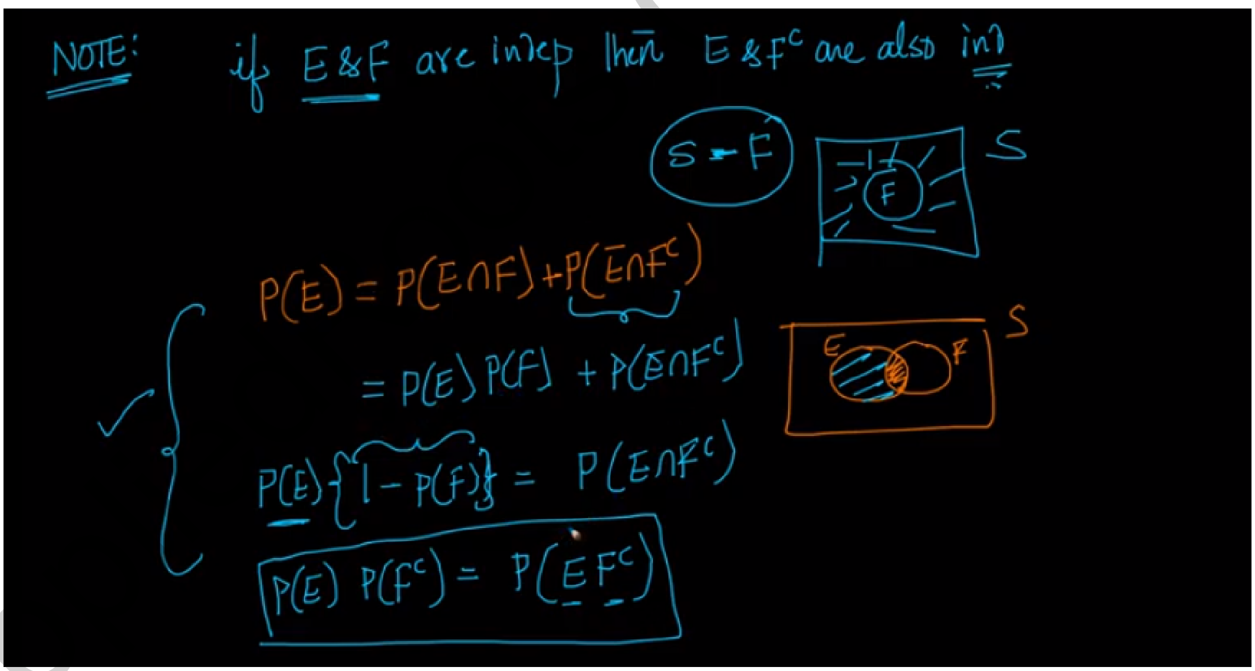


- In the above example probability of picking a second card doesn’t depend on the event of picking the first card since we are replacing the first card.They both are independent events .



- We can extend the concept of independent events and generalize it for n events as shown above .Its nothing but the product of all the events which are independent.

Timestamp 17.9



- As shown above if two events E and F are independent then E and F^ are also independent.

Timestamp 24.30

(e.g) An infinite seq. of trails | expt (indep) is performed; a success prob = p

trail: throwing a dice
success: outcome is 1 ($p = 1/6$)
failure: outcome is not 1 ($1-p$)

(a) atleast one success in (n) indep trails

$= 1 - P(\text{no success in } n \text{ trails})$

$= 1 - P(F_1 F_2 F_3 F_4 \cdots F_n)$

$= 1 - P(F_1)P(F_2) \cdots P(F_n) = 1 - (1-p)^n$

(b) exactly K successes in n trails

$= {}^{n}C_{K}\, p^{K} (1-p)^{n-K}$

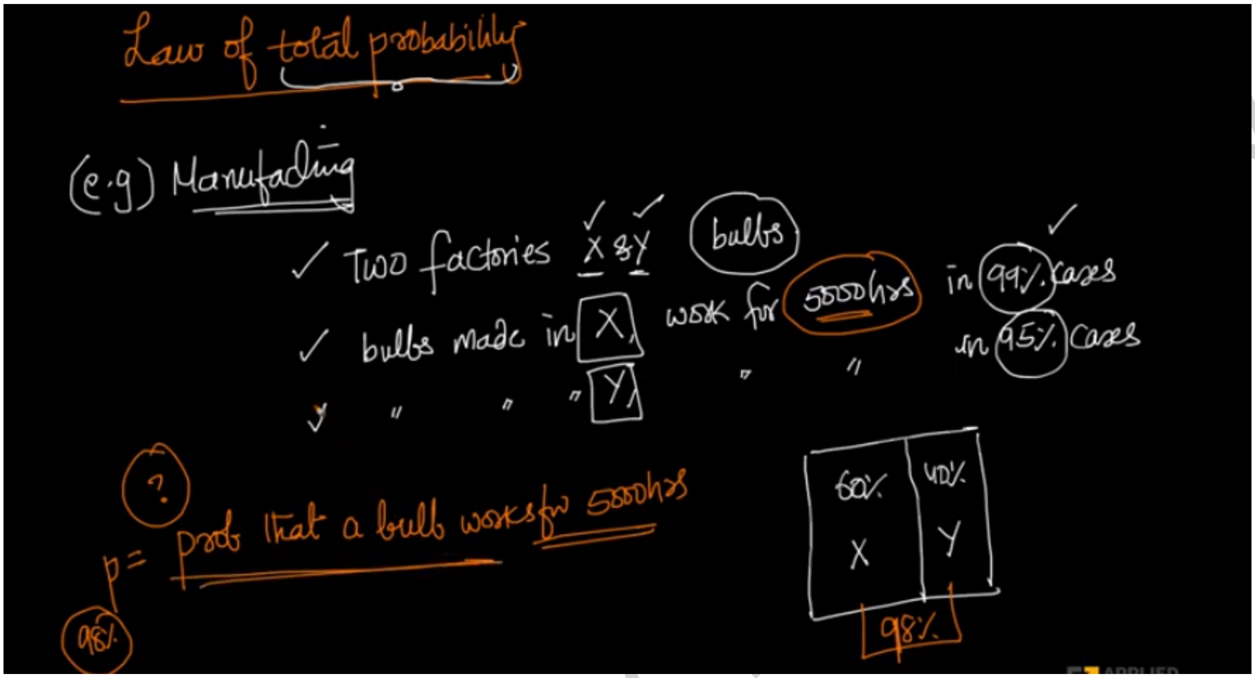
(c) all n trails are succ.

$= p \cdot p \cdots n\text{ times} = p^n$

- Above is an example of using independent events to solve problems.

21.9 Law of total Probability

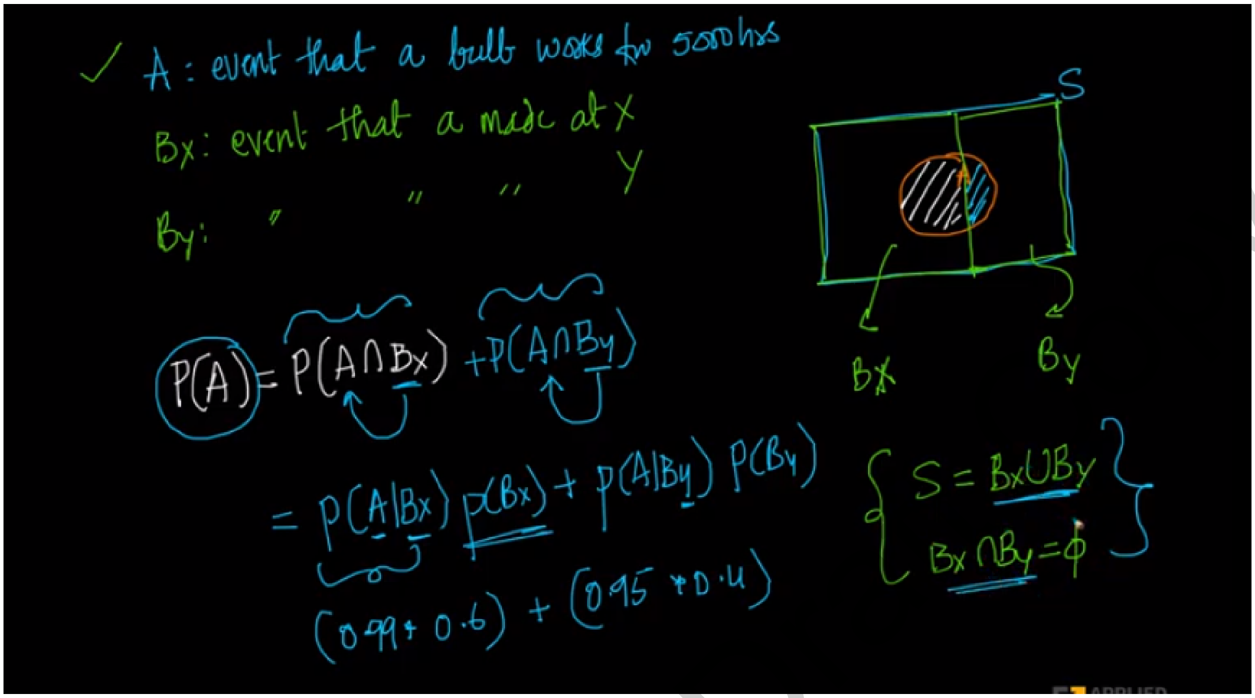
Timestamp 3.00



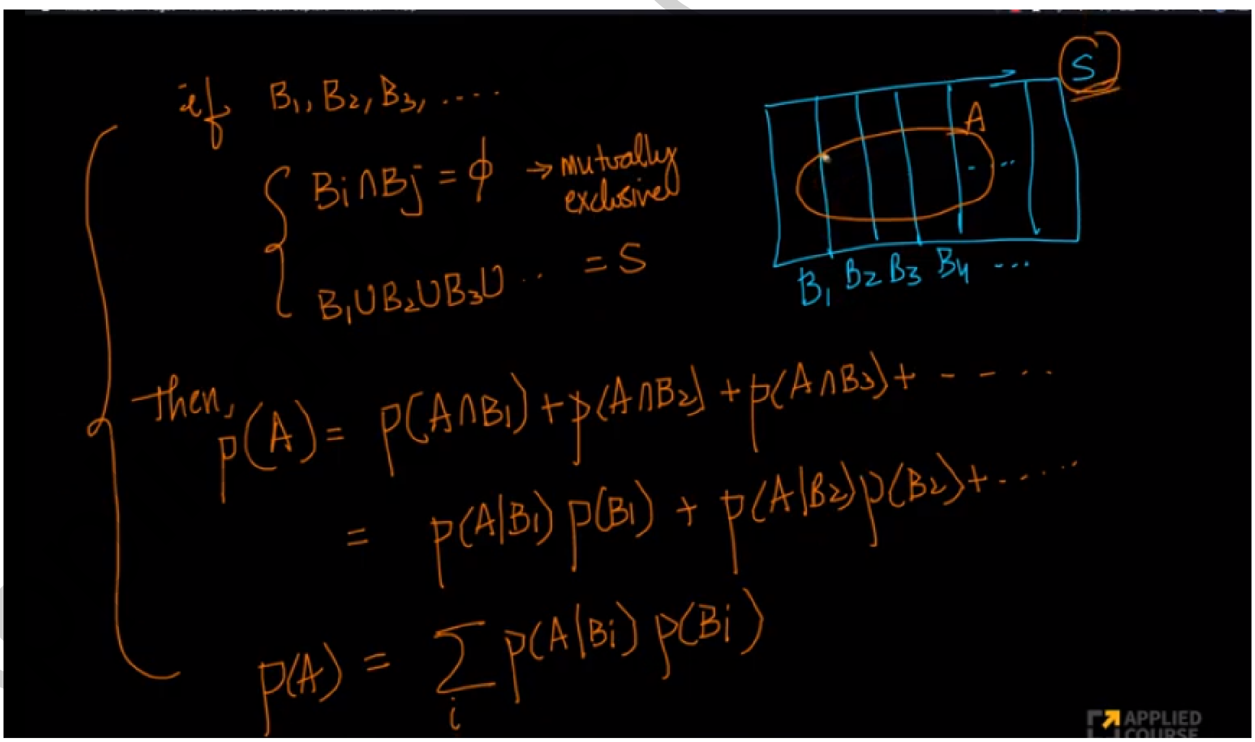
1. Above is an example of use case where we use Law of total probability

2. We have two manufacturing factories X and Y.As shown above if we have a package where we have 60 % of bulbs from factory X and rest 40 % from factory Y.Now we need to find the probability that the bulb works for 5000 hours.

Timestamp 10.6



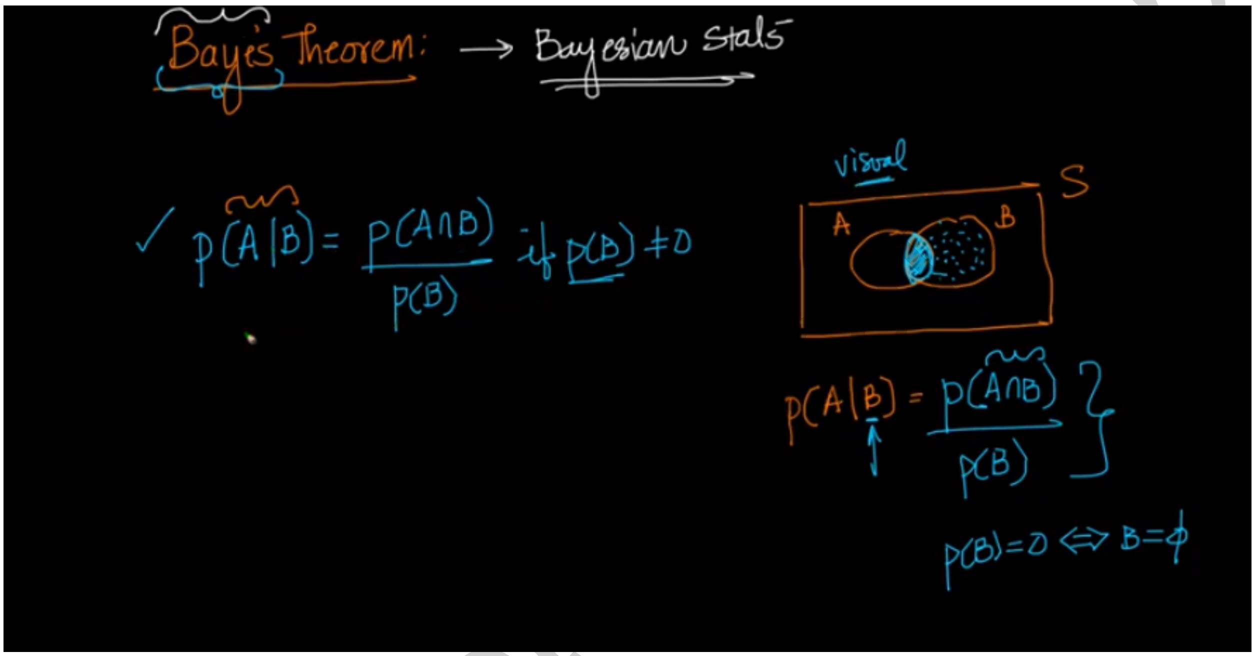
- We can solve the problem as shown above using total probability.



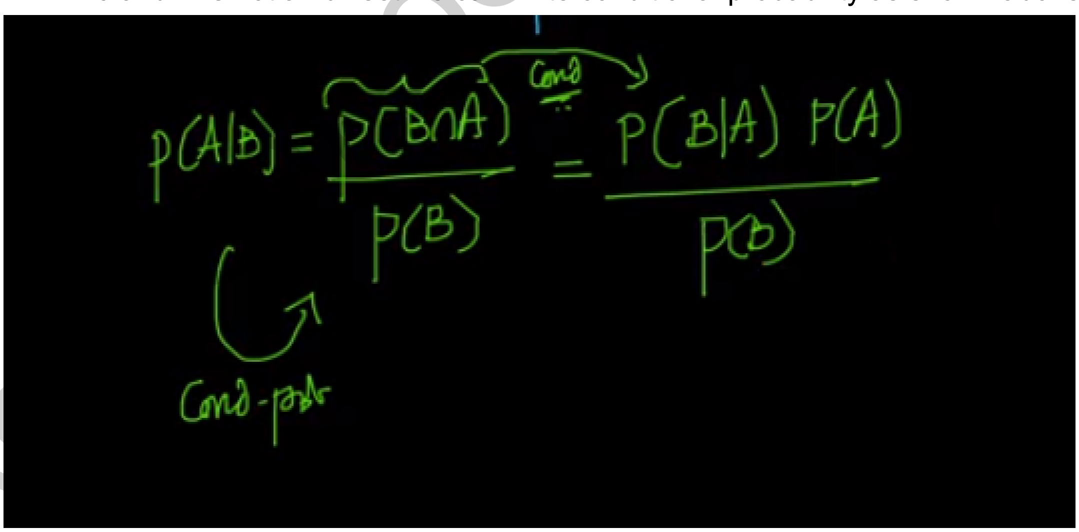
- If you have several events B1,B2,B3,B4…….etc. such that all are mutually exclusive events and union of all the events is sample space S(mutually exhaustive events).Then probability of any event A can be written as shown above using law of total probability.

21.10 Bayes Theorem

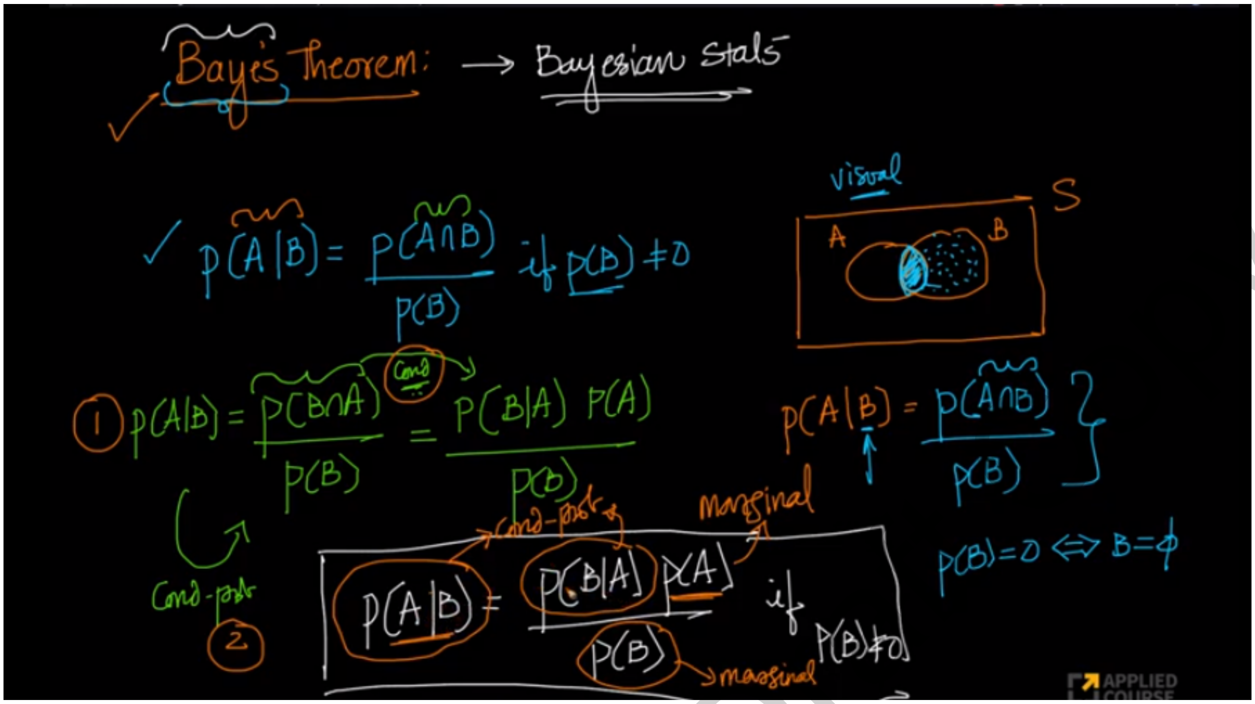
Timestamp 4.35



- We can obtain Bayes theorem using conditional probability.By assuming that p(B) is not 0 and B is not a null set we can write conditional probability as shown above.

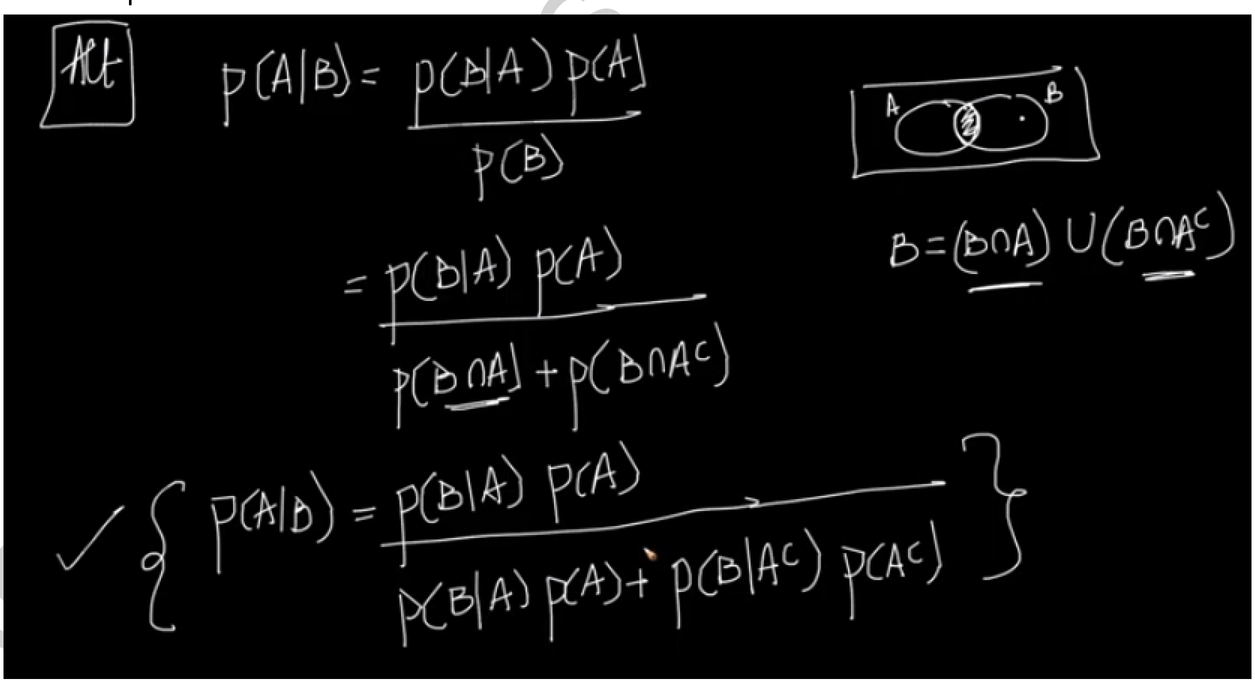


- We apply conditional probability again as shown above to get Bayes' theorem.

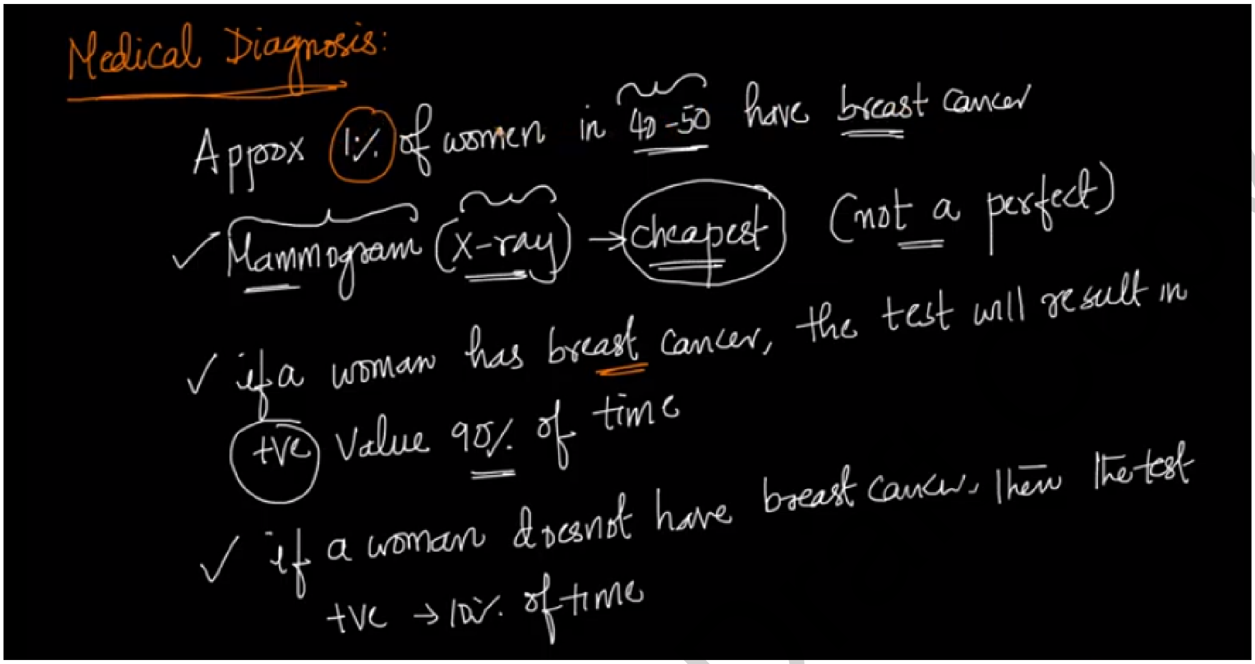


- The probability of events A and B called marginal probabilities.
- Bayes theorem can also be thought of as an equation that connects the conditional probabilities two events and their marginal probabilities

Timestamp 10.58

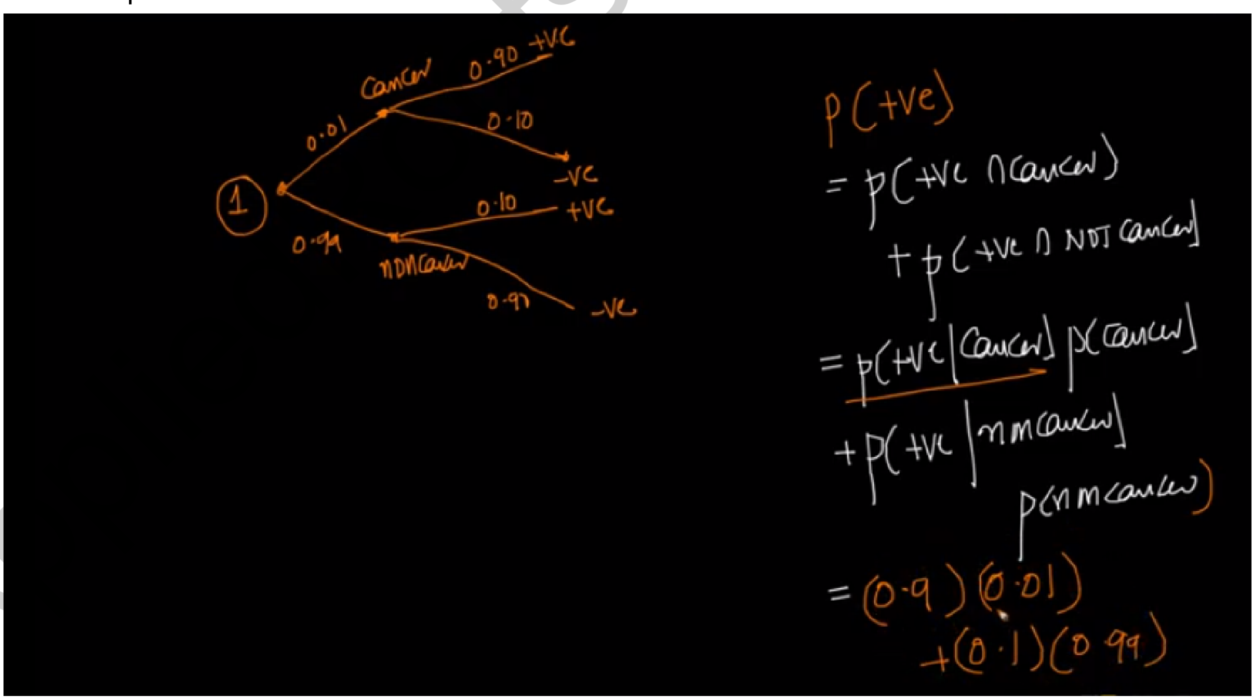


- Alternative definition for Bayes theorem can be derived as shown above. Which can be easily derived using set theory.

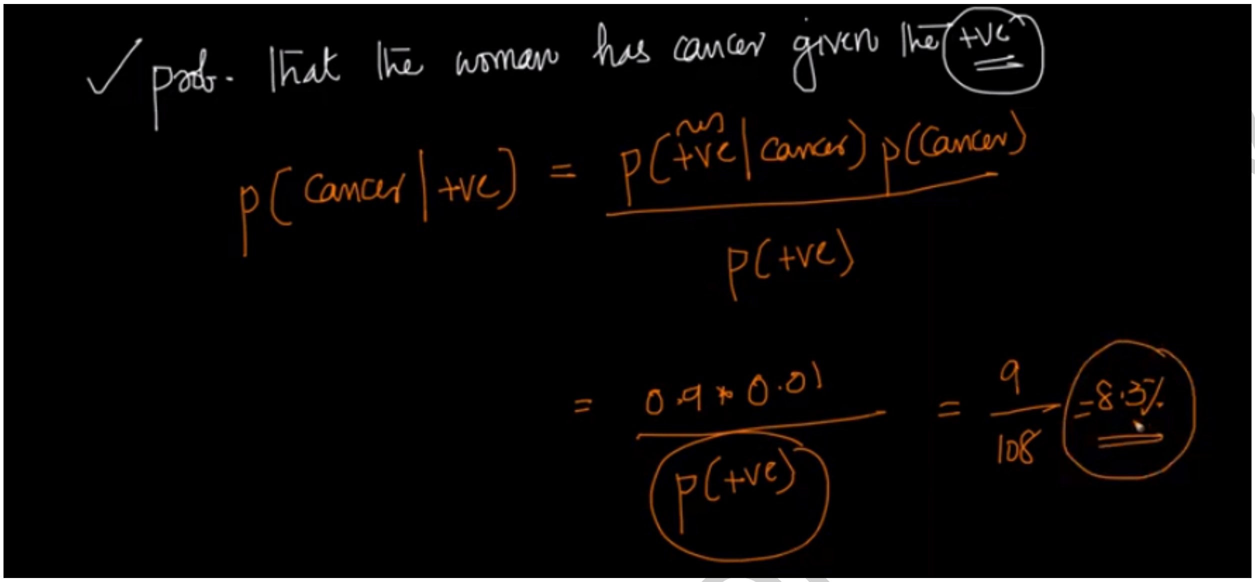


- Above is an example solved using Bayes theorem.
- We have to solve the probability that a woman actually has cancer given the test is positive.

Timestamp 21.17

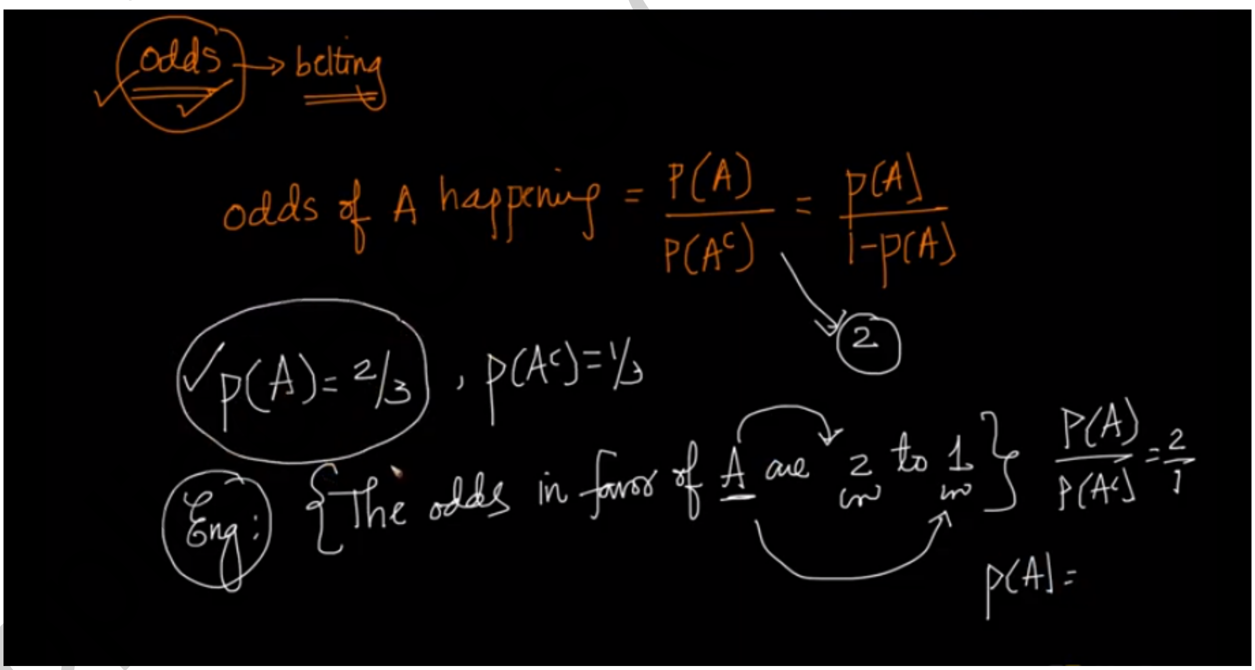


Timestamp 23.17



The problem can be solved as shown above

Timestamp 29.21

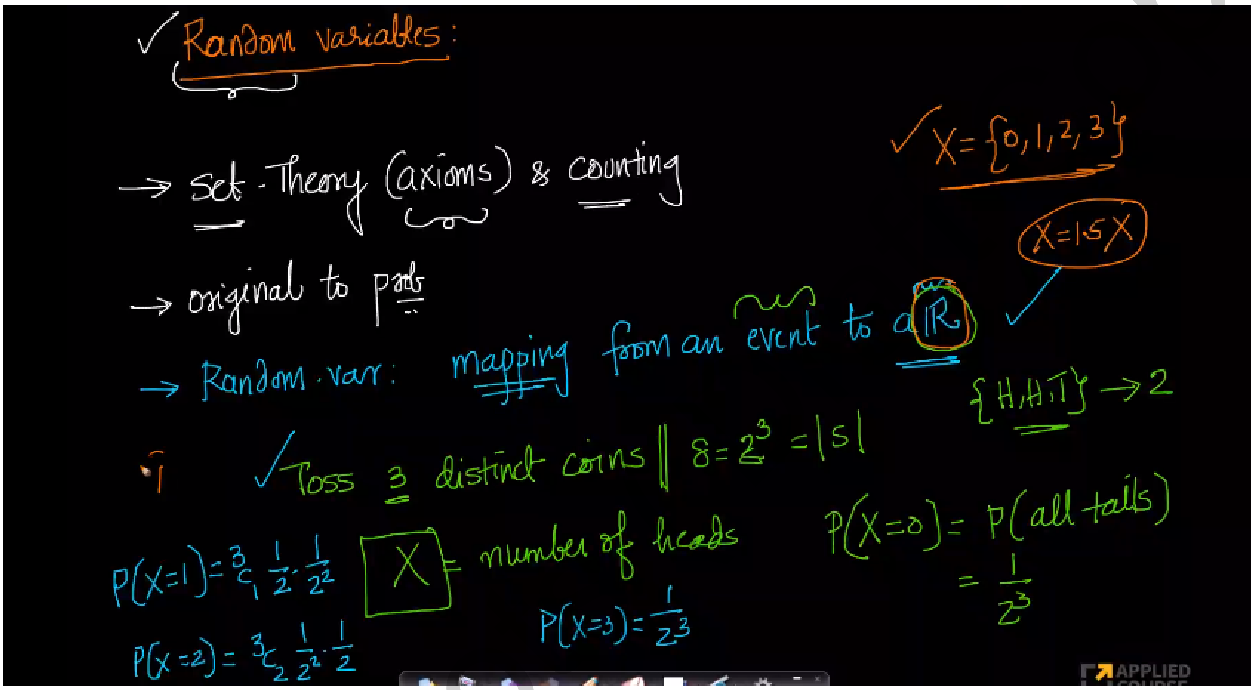


- There is a concept called odds .The Probability of Event A happening is ⅔ and not happening is ⅓ then we can say that the odds in favour if A are 2 to 1 which means there's ⅔ probability that A ill happen and ⅓ probability that it won't happen.

21.12 Random variables with examples

Timestamp 5.25

A random variable can be thought of as a mapping or function from an event to a real number. It is also called as a stochastic variable



- In the above example of tossing 3 distinct coins let X be a random variable where X=number of heads .Here X is a discrete random variable as it can only take only real values 0,1,2,3.

(2) ✓ X = amount of rainfall on a given day

✓ $p(X \geq 2cm) = 0.95$ ✓ farmer

✓ $p(X \leq 1cm) = 0.99$

✓ $p(X \geq 20cm) = 0.999$

✓ $X \in [0, \infty)$ 2.1cm, 1.067cm

benches/chairs ✓

(3) X = height of students

$p(X \geq 180cm) = 1\%$

(4) X = time spent on a website

let $p(X \geq 10min) = 80\%$ ✓

let $p(X \leq 1min) = 90\%$ (wrong) → ✓

$X = [0, \infty)$

Continuous

2.1 min
1.067 min

(5) X = # visitors to a website on a given day → discrete

✓ $P(X \geq 1000) = 0.1\%$

$\frac{1}{1000} = \frac{1}{3yrs}$

S/W & H/W
servers

$X = \{0, 1, 2, 3, \ldots\}$

1.5X
2.1X

(6) X = # children in a family

discrete

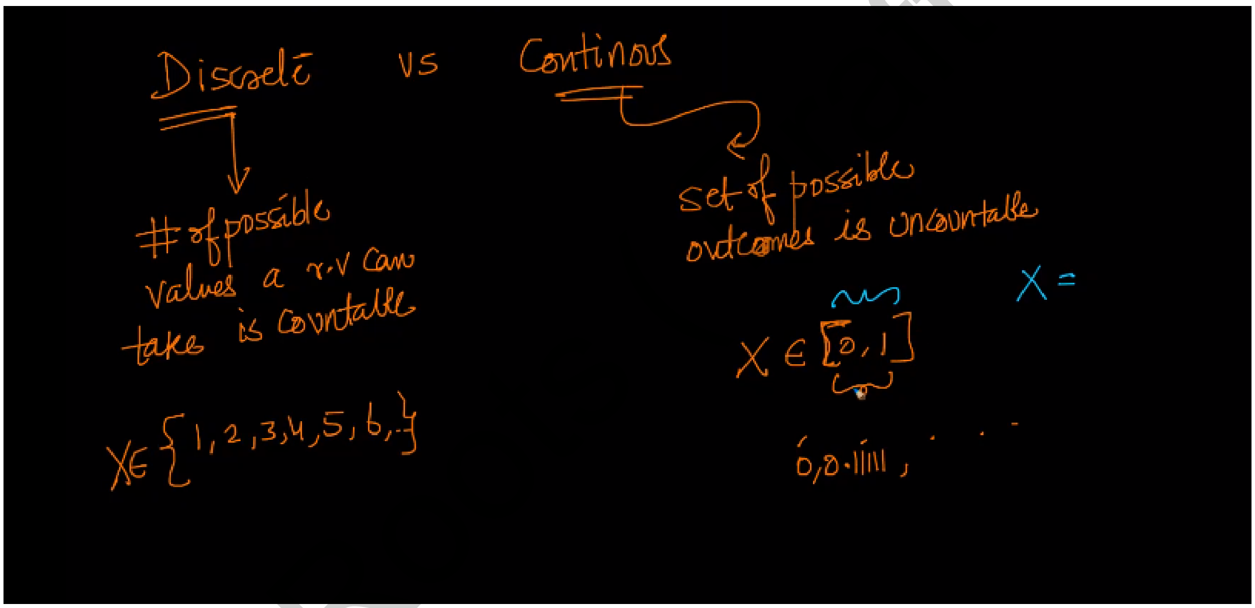
✓ $p(X \leq 2) = 95\%$ ✓ → India

$X \in \{0, 1, 2, 3 \ldots\}$

✓ $p(X = 0) = 90\%$ → Japan ✓

1. The random variable describing the amount of rainfall on a given day is continuous random variable
2. The random variable describing the height of students is continuous random variable
3. The random variable describing the time spent on a website is a continuous random variable as it can take any value between 0 to infinity.
4. The random variable describing number of visitors to a website on a given day is discrete random variable
5. The random variable describing number of children in a family is discrete random variable

Timestamp 22.20

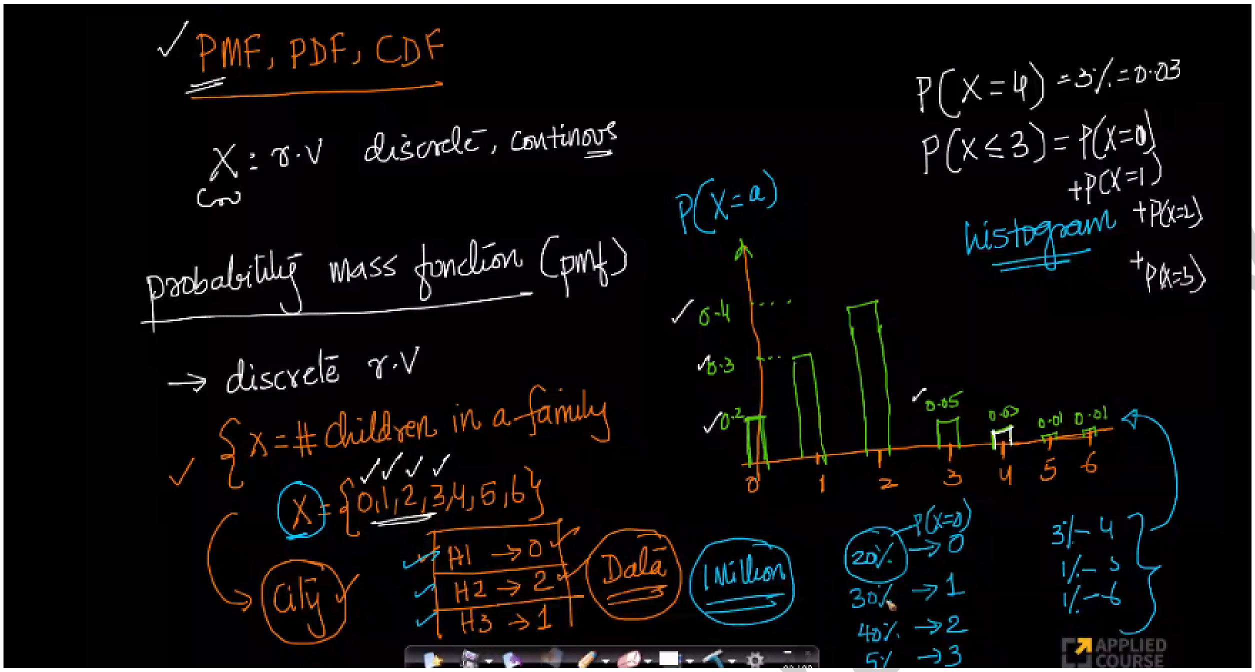


Discrete and continuous random variables are defined as shown above

21.13 PMF, CDF and PDF of random variables

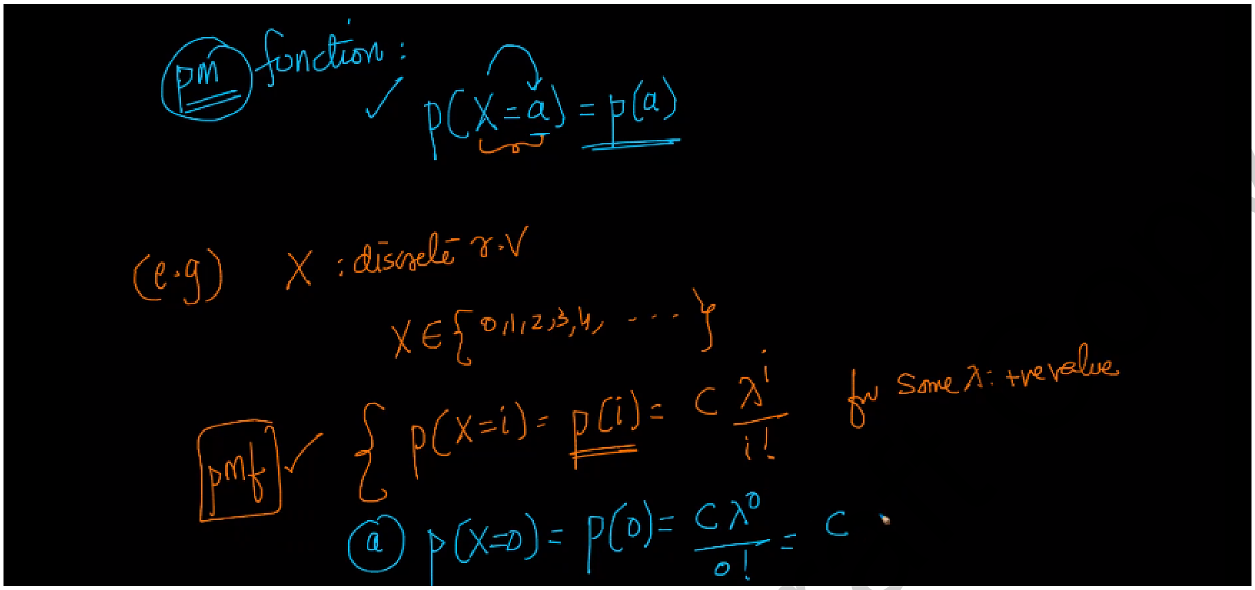
There are three important functions that are defined on random variables PDF,CDF,PMF,lets discuss

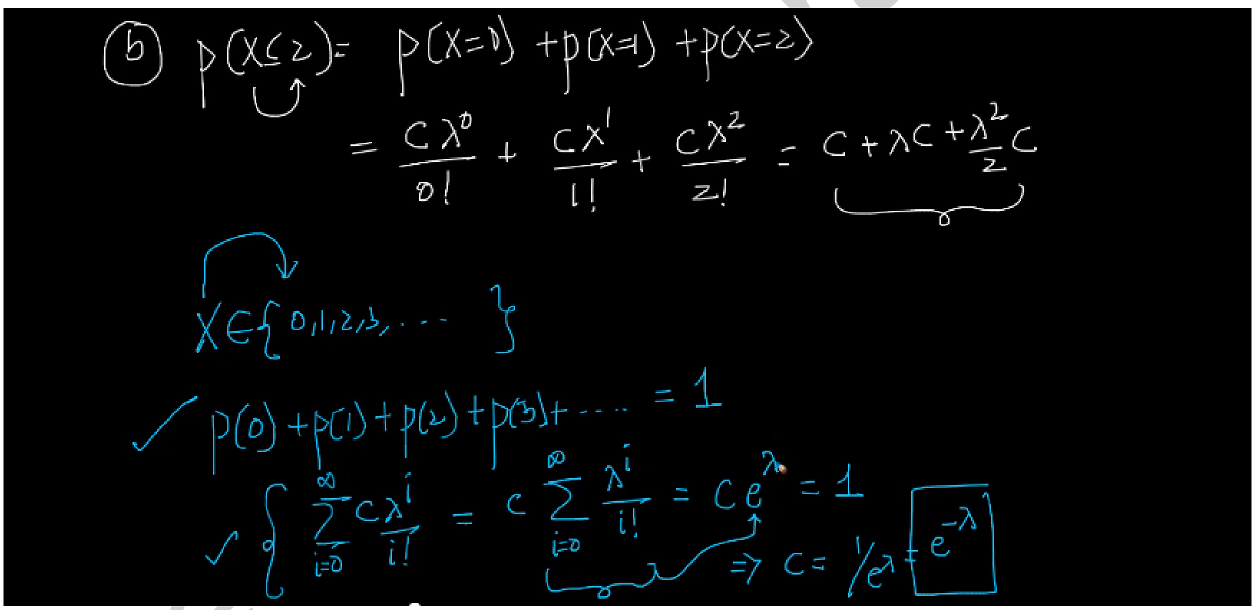
Timestamp 8.20



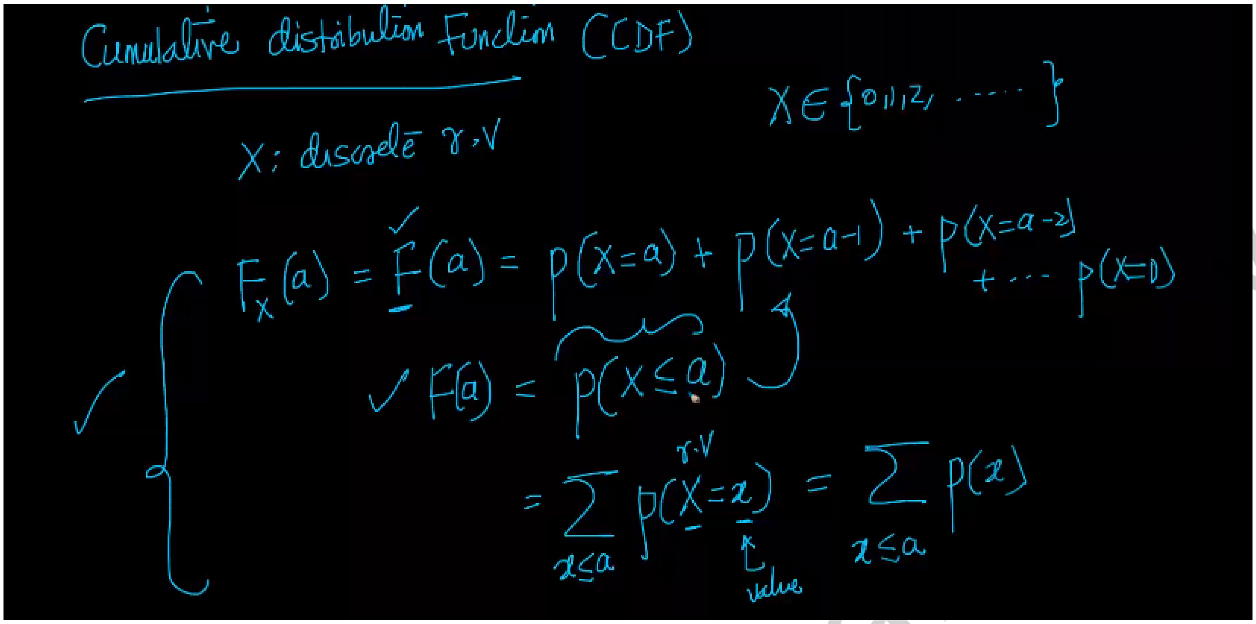
- PMF is described for discrete random variables.Let x be a random variable describing the number of children in a family which is discrete.
- We are plotting P(x=a) on y axis and values of x in x-axis and we draw the histogram as shown above .Using the histogram we can answer questions as shown.

Timestamp 12.25

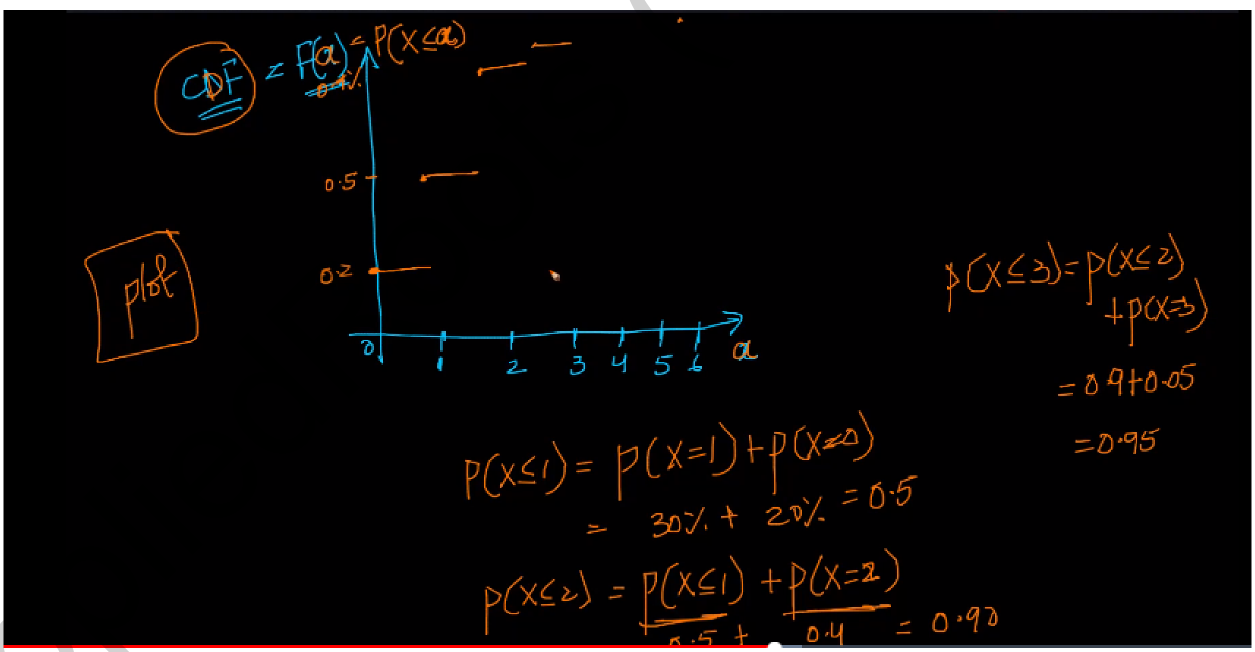




- PMF is a function which gives the probability that random variable X takes value a i.e) P(X=a) or P(a)
- As shown above if we have the PMF function we can easily find the probability of random variable X taking value a .
- The sum of all the probabilities that variable x can take will be equal to 1.

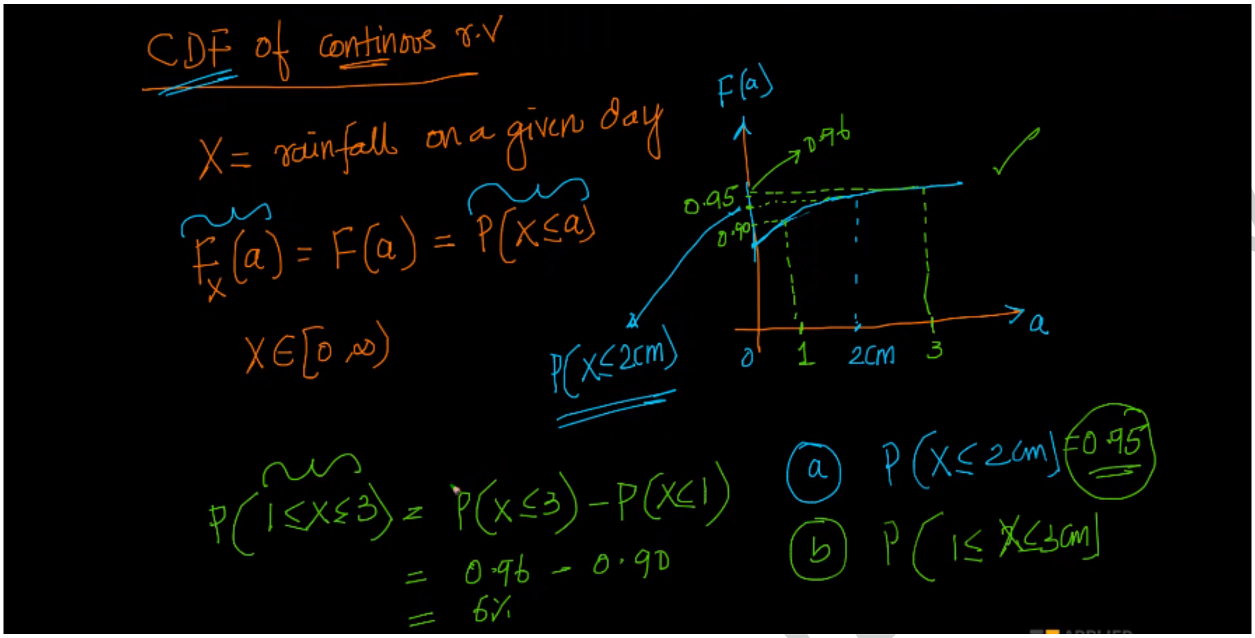


- Cumulative distribution function of a Random variable X at a point a ,where a is one of the values random variable X can take is probability that X can take value which is less than or equal to a as shown above



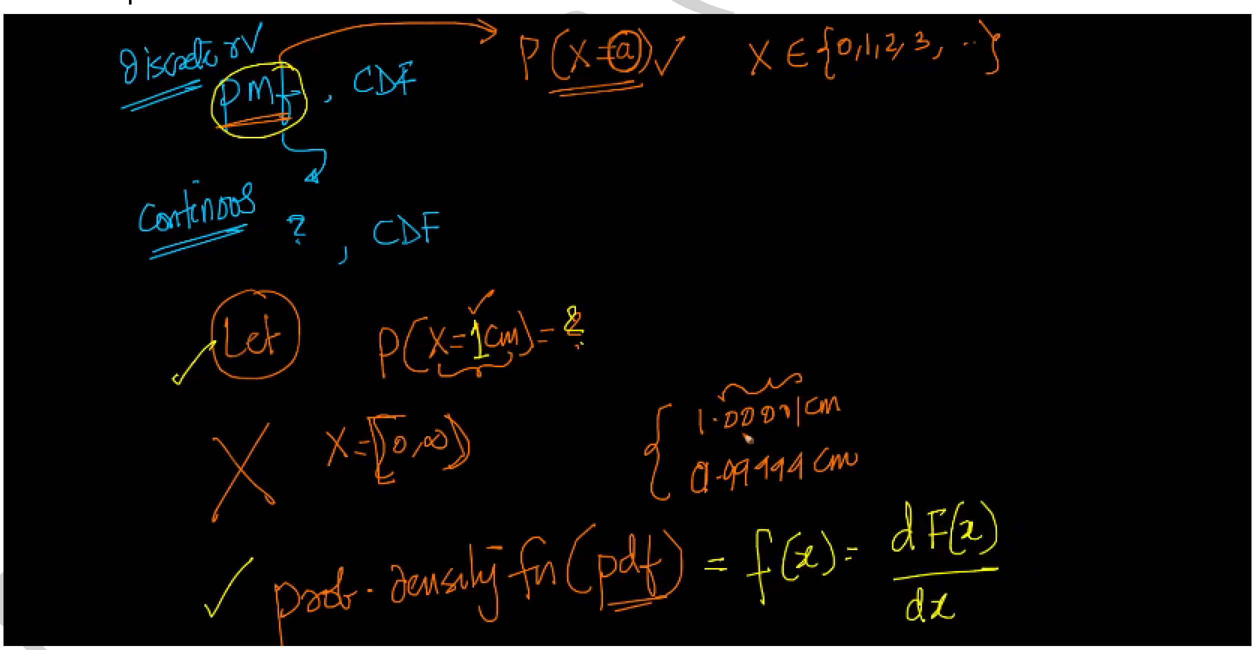
- CDF of a discrete random variable can be plotted as shown above .

Timestamp 30.34

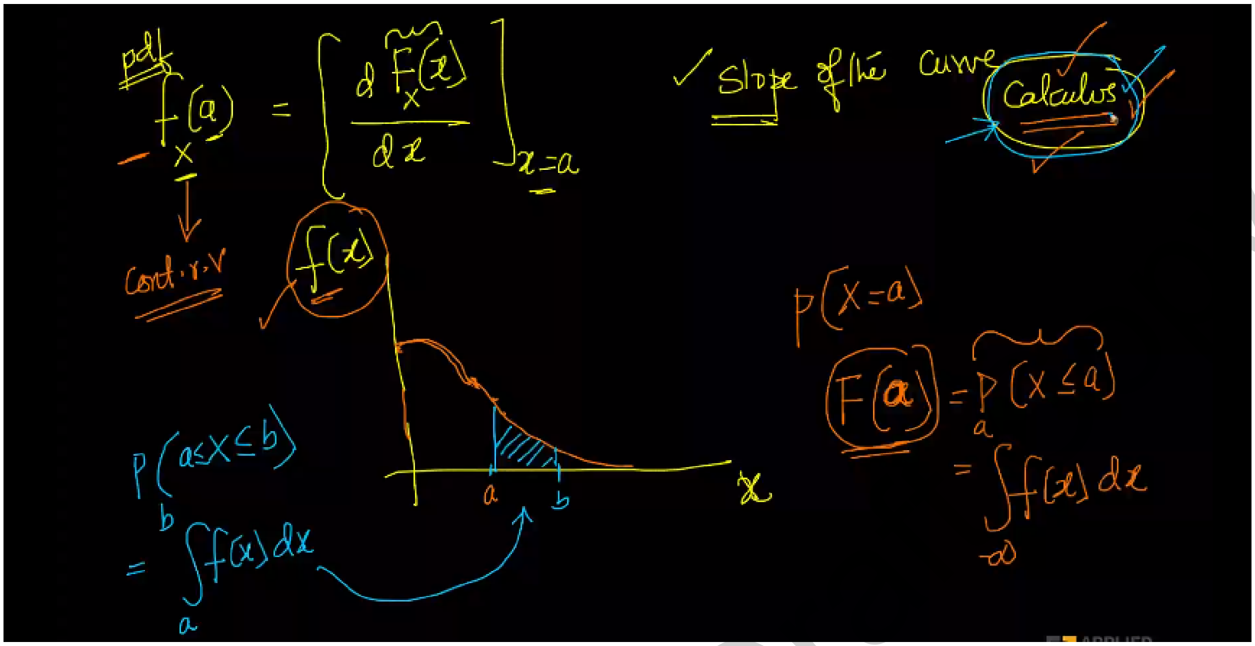


- CDF of a continuous random variable is explained above.We can draw a plot for cdf function and using it we can easily find probabilities as shown.

Timestamp 30.17



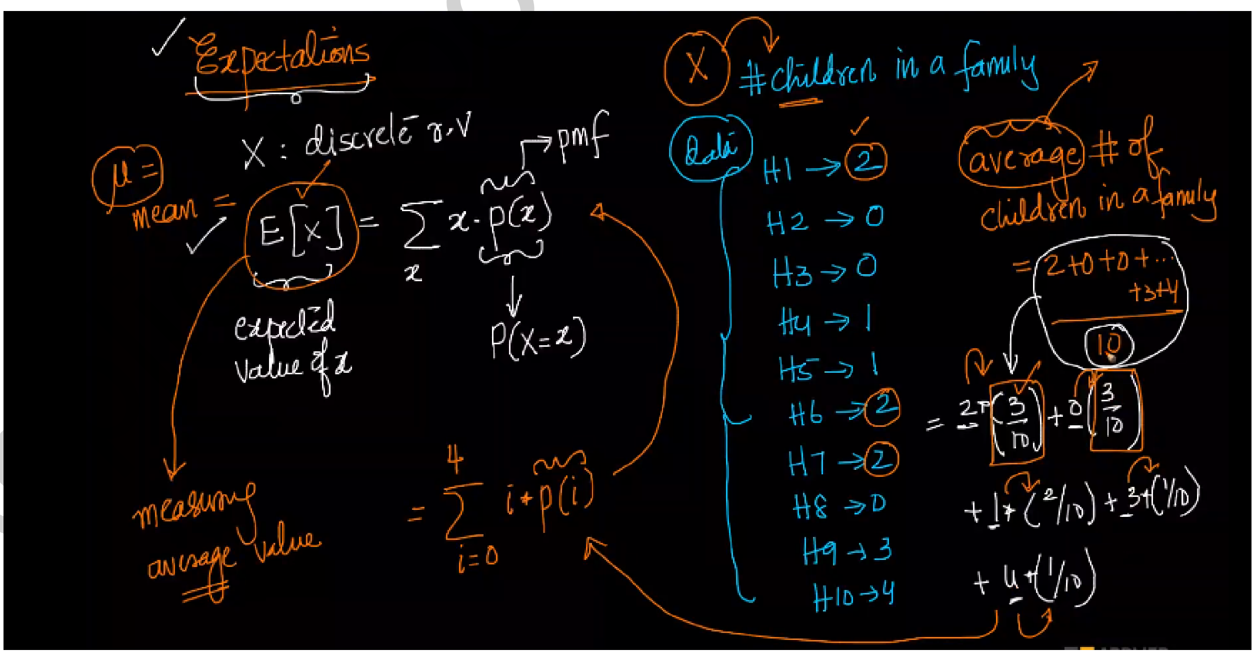
- We have CDF defined for both discrete (curve is not smooth)and continuous random variables(has smooth curve) but PMF is defined for only discrete random variables and for continuous random variables we have PDF (Probability density function).



- We define the PDF as shown above .

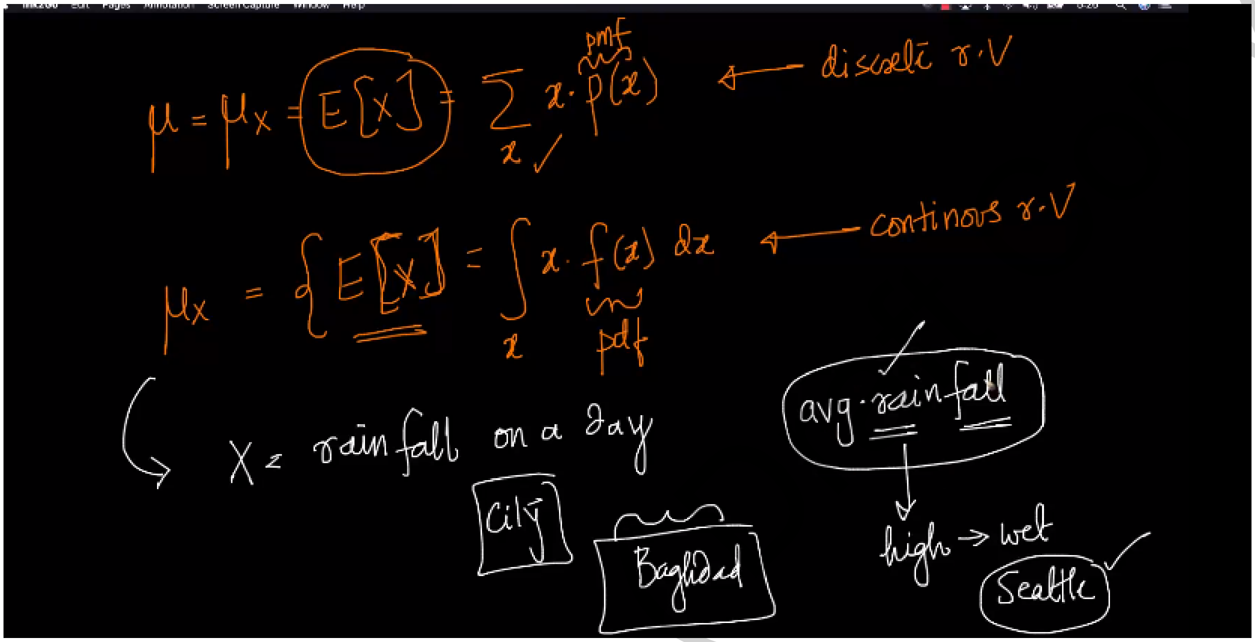
21.14 Expectation

Timestamp

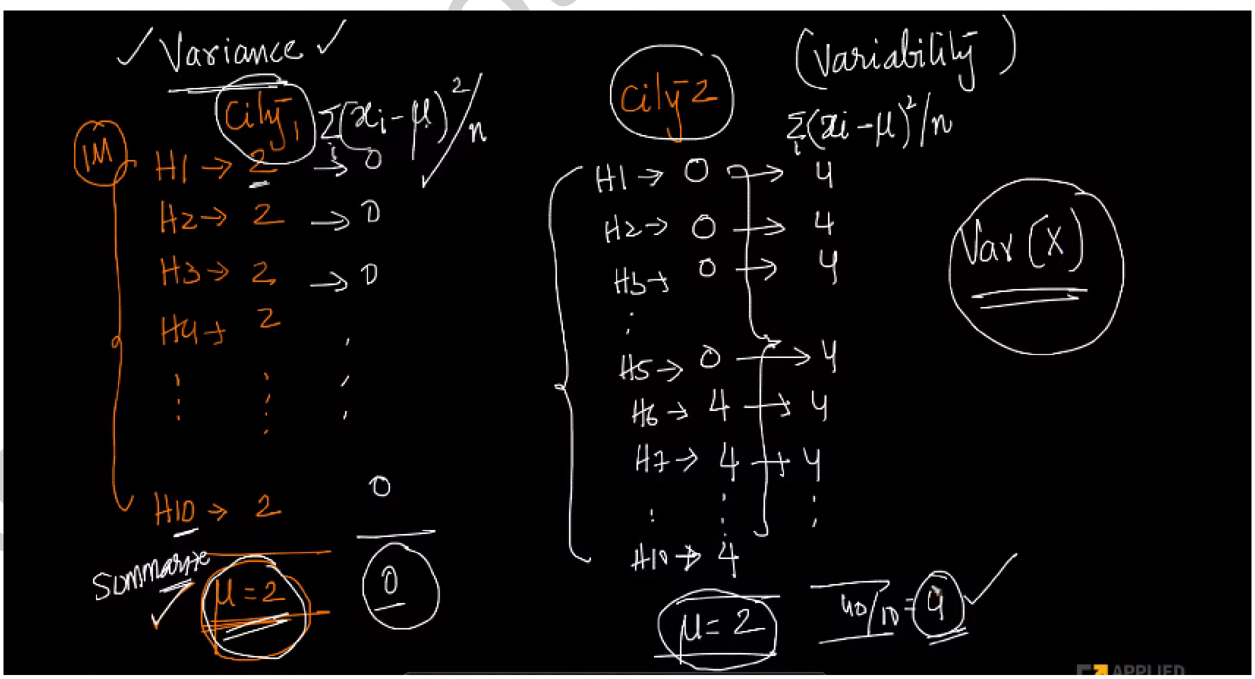


- Expectation can be defined for a discrete random variable.Expectation of x can be written as shown above ,we can understand it using an example shown above.
- Intuitively expectation is measuring the average value of random variable x based on some data.

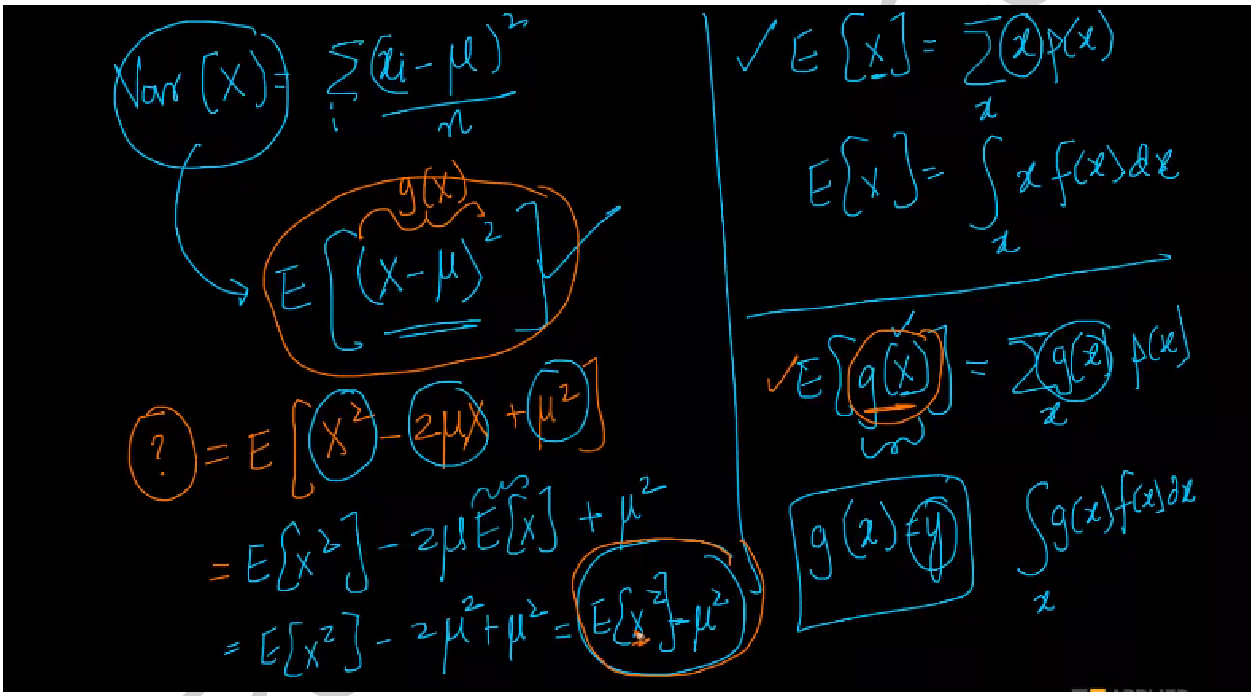
Timestamp 11.21



- Expectation for a continuous random variable can be obtained as shown above .

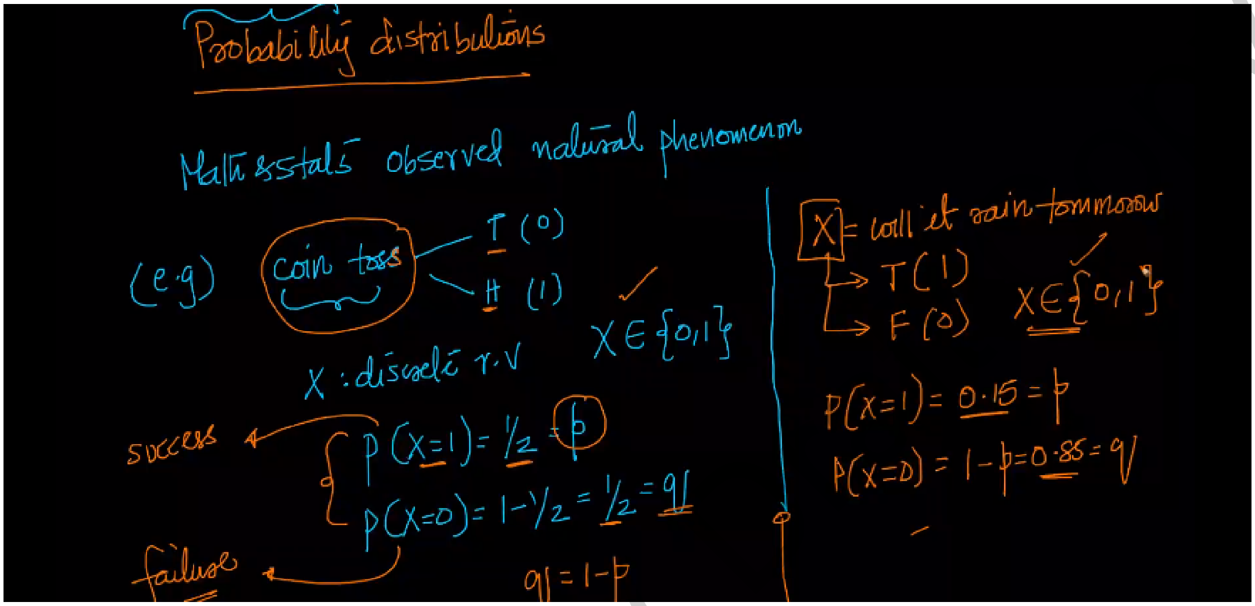


- Variance can be explained using the example shown above.
- If we are given average of number of children in household in two different cities ɥ,ɥ^. we can clearly notice there is significant variability between number of children in the households in two cities, only when we look into the data we can notice the variability.
- Clearly only mean cannot describe the data perfectly ,so to describe such a phenomenon in the data we use variance as a measure.
- As shown variance of the data of City1 is 0 and the variance of City 2 is 4 Which makes us understand that there is a huge variability in the data of city 2.
- Variance measures variability in the data



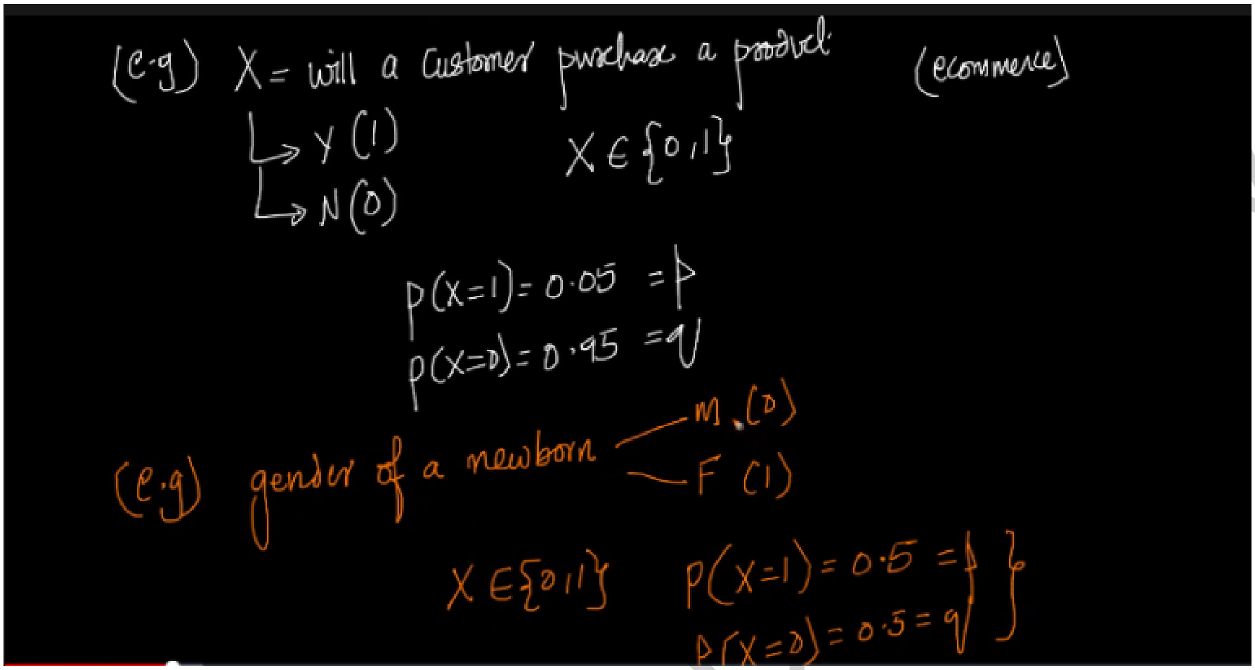
21.15 Probability Distributions: Bernoulli and Binomial

Timestamp 2.05

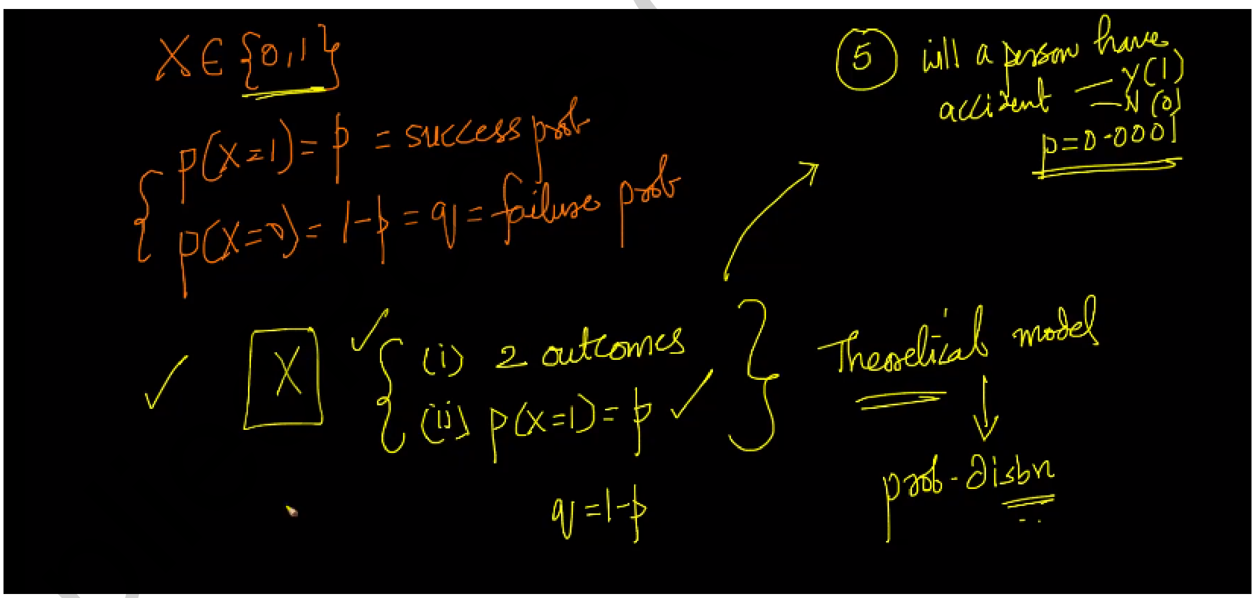


- Let's consider the above examples and understand the probability distributions
- As shown in the experiment of coin toss X is a random variable and probability of x taking a value 1(getting head) is p and probability of x taking a value 0(getting tail) is q . Here q=1-p
- Similarly we can understand the example of random variable x where it describes whether it will rain tomorrow as p. Probability of not raining is q=1-p
- Intuitively p,q are nothing but probability of success and failure.
- Similar to the above examples we can notice the same pattern in the below example as well.

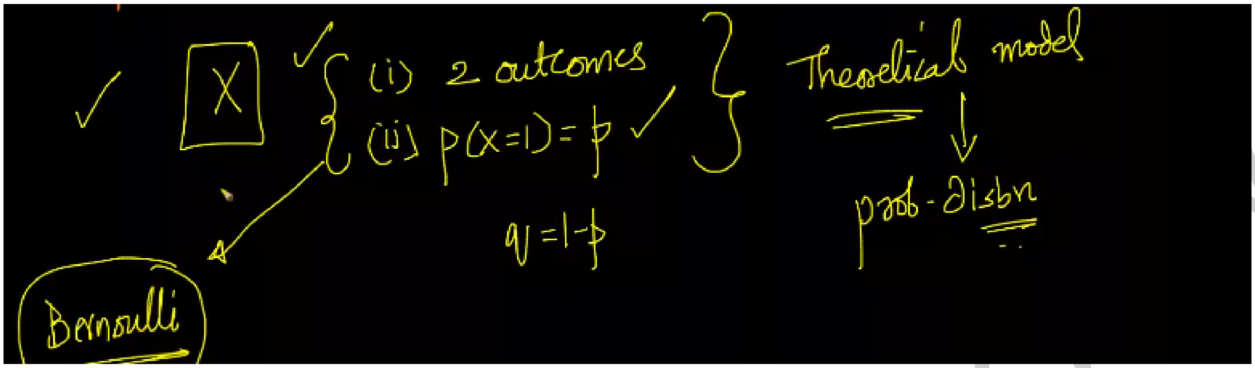
Timestamp 7.14



Timestamp 11.18

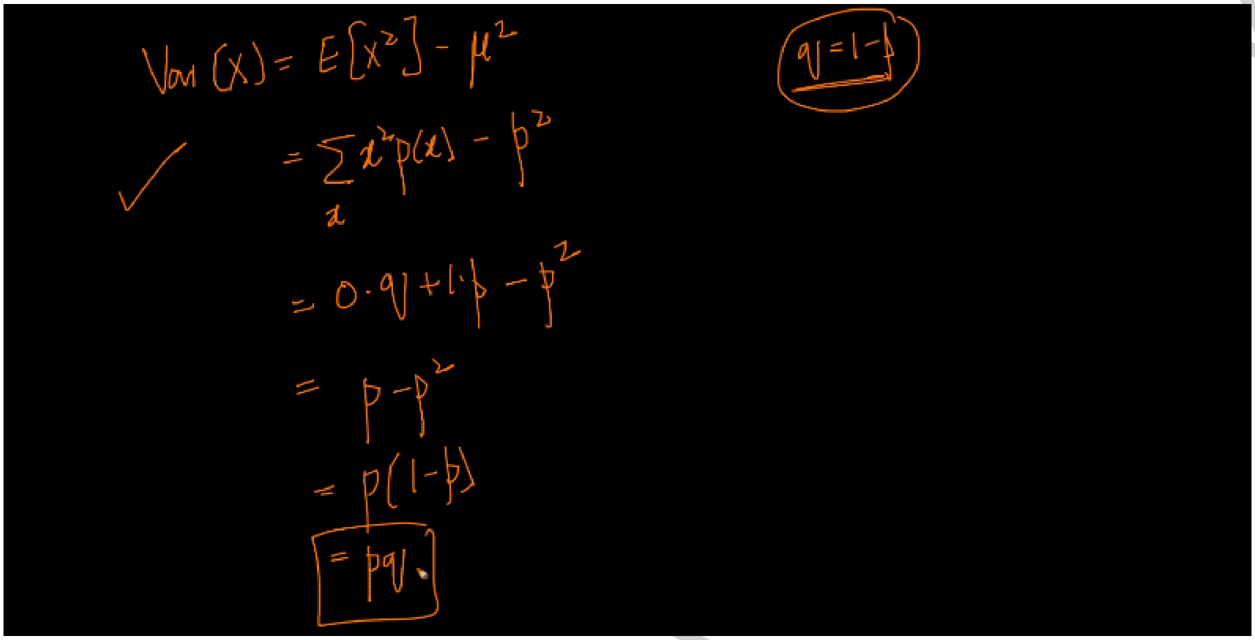


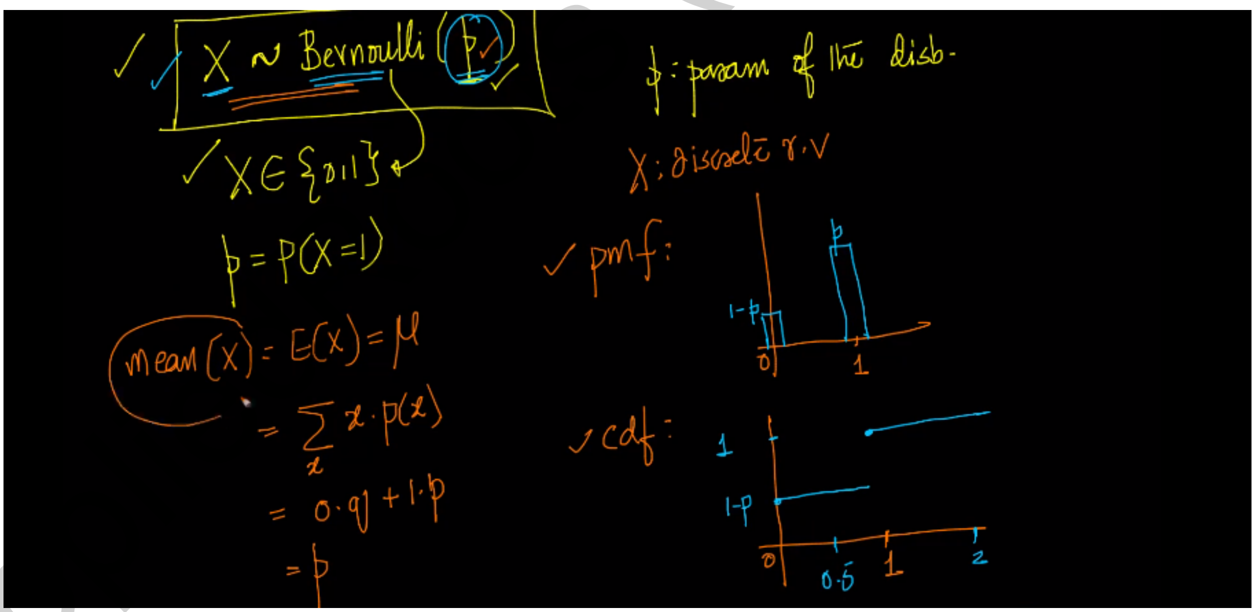
- The pattern we have observed in all the above examples is put together as shown above .
- Statisticians thought that if we can study the patterns i.e)probability distributions and come to certain conclusions(ideas and outcomes),we can apply all those conclusions to any of the examples we have discussed.



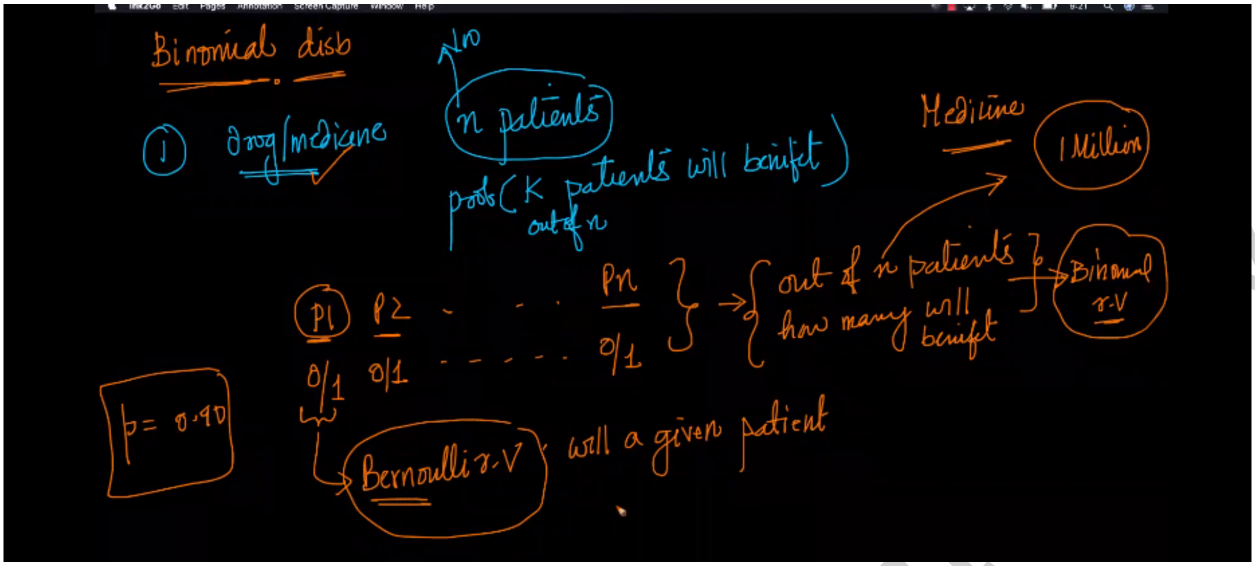
- If the random variable follows above two properties then it belongs to Bernoulli distribution.
- When we say a random variable belongs to the Bernoulli distribution it will have two outcomes and if we get probability of success P then we can say everything about this random variable

Timestamp 15.26



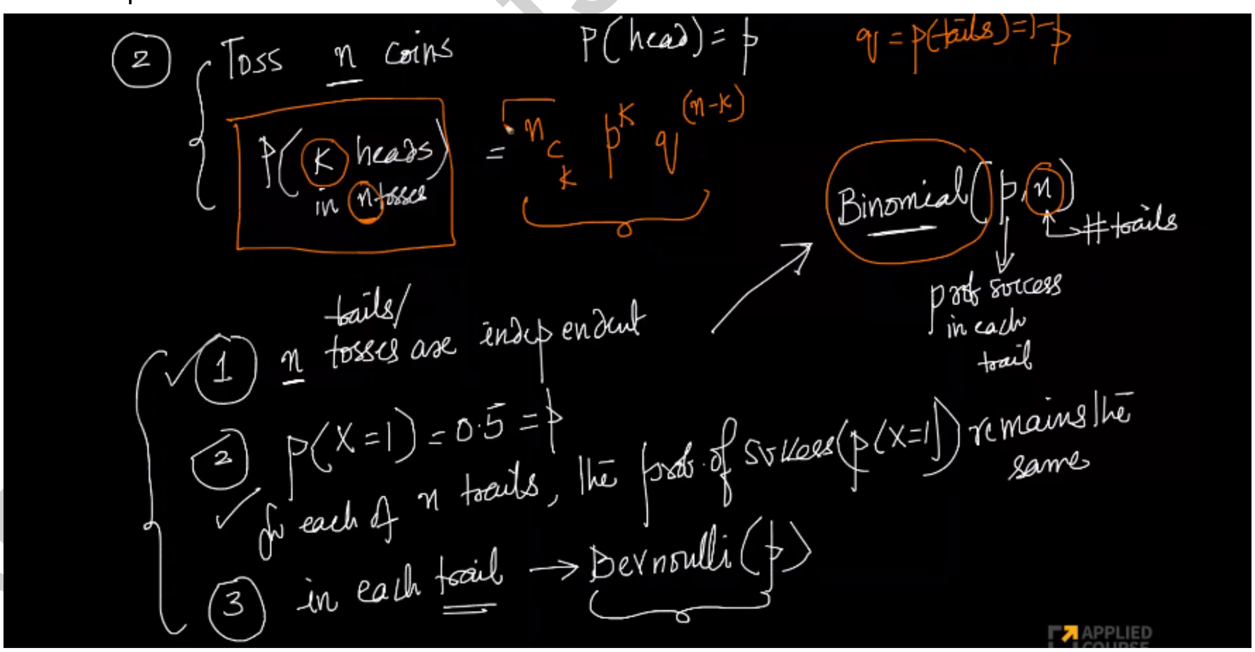


- The PMF,CDF, mean,variance of a random variable belonging to Bernoulli distribution as shown above.



- Let's understand Binomial distribution using the above example
- Let's say we have manufactured a drug and we want to find the probability that how many patients will benefit out of n patients.We know that the outcome of drug on patient 1 is independent of the outcome of drugs on patient 1.For each patient there are two outcomes 0/1 (p/q)drug didn't help the patient and drug did helped the patient this is nothing but Bernoulli distribution.

Timestamp 31.46



- If a real world phenomenon satisfies above three conditions then it is said to belong to Binomial Distribution as shown above.

Timestamp 33.41

$$X \in \{0,1,2,\dots n\}$$

$$P(X=k) = p(k) = {}^nC_k \, p^k (1-p)^{n-k} \longleftarrow \text{pmf}$$

$$P(X \le k) = \sum_{i=0}^{k} {}^nC_i \, p^i (1-p)^{n-i} \longleftarrow \text{cdf}$$

$$\mu = E(X) = E(X_1 + X_2 + \dots + X_n)$$

$$= E(X_1) + E(X_2) + \dots + E(X_n)$$

$$= p + p + \dots \text{ n-times}$$

$$= np$$

$$\mu = E(X) = \sum_{k=0}^{n} k \cdot p(X=k) = \sum_{k=0}^{n} k \cdot {}^nC_k \, p^k (1-p)^{n-k}$$

$$Var(X) = Var(X_1 + X_2 + X_3 + \dots + X_n)$$

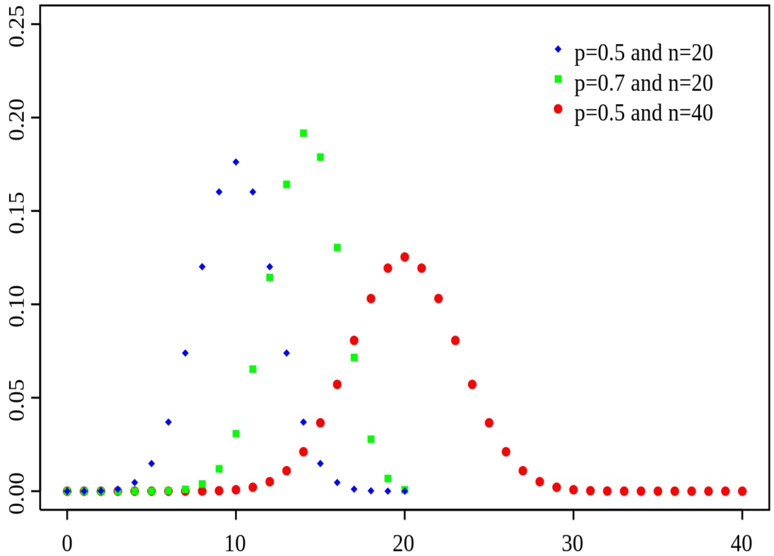
$\{X_1, X_2, X_3 \dots X_n\}$ are indep

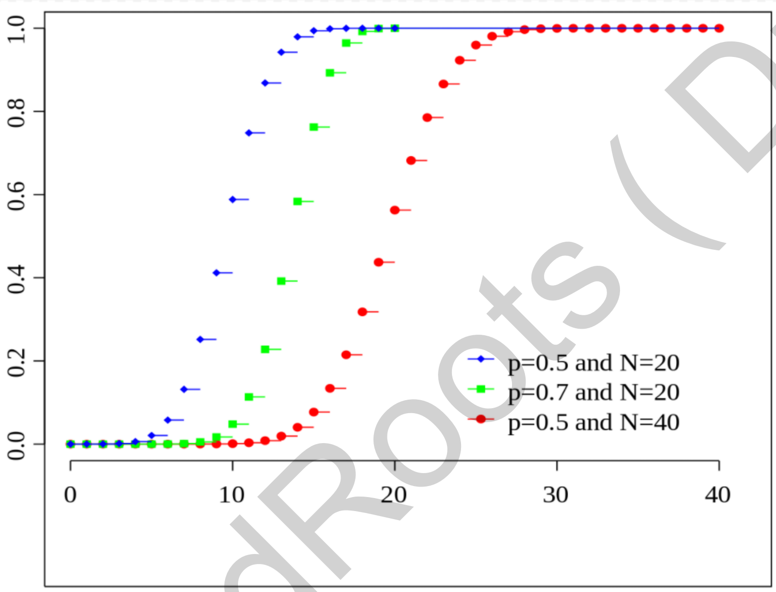
$$= Var(X_1) + Var(X_2) + \dots + Var(X_n)$$

$$= pq + pq + \dots \text{ n times}$$

$$= npq = np(1-p)$$

- PDF,CDF,mean,variance of a Binomial distribution can be derived as shown above

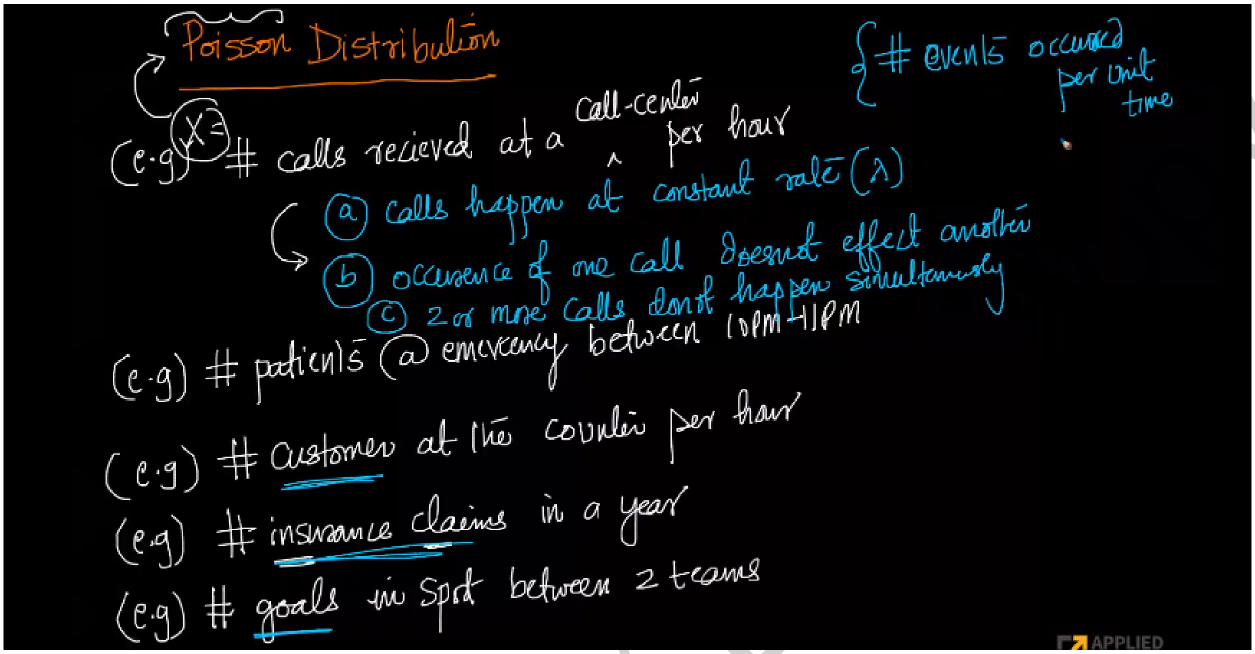


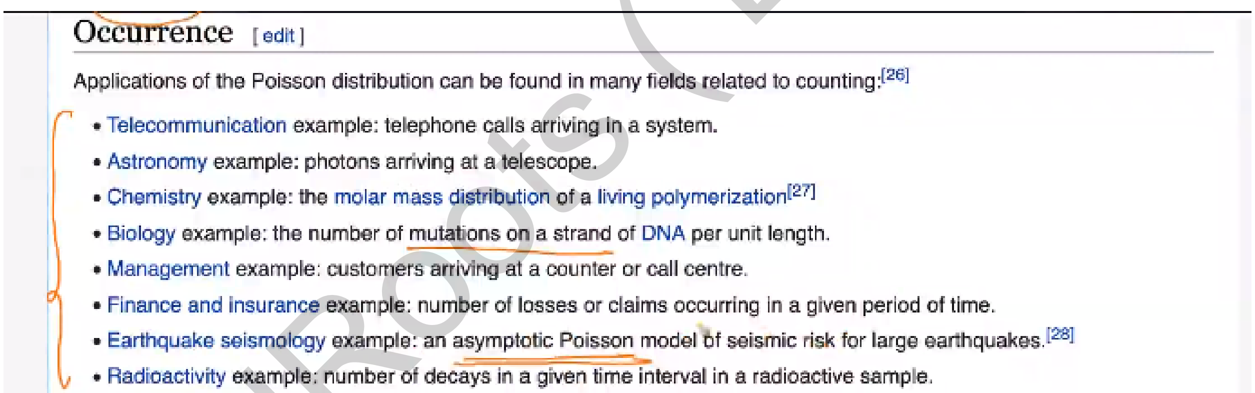


- Above are the pdf and cdf curves of a binomial distribution ,the curves are not continuous ,they are discrete.

21.16 Poisson Distribution

Timestamp





Occurrence [edit]

Applications of the Poisson distribution can be found in many fields related to counting:[26]

- Telecommunication example: telephone calls arriving in a system.
- Astronomy example: photons arriving at a telescope.
- Chemistry example: the molar mass distribution of a living polymerization[27]
- Biology example: the number of mutations on a strand of DNA per unit length.
- Management example: customers arriving at a counter or call centre.
- Finance and insurance example: number of losses or claims occurring in a given period of time.
- Earthquake seismology example: an asymptotic Poisson model of seismic risk for large earthquakes.[28]
- Radioactivity example: number of decays in a given time interval in a radioactive sample.

- We can understand poisson distribution using above examples
- Poisson distribution measures the number of events occurred per unit time

$X \sim \text{poisson}(\lambda)$ ← rate $\qquad X = \{0, 1, 2, 3, 4, \ldots\}$

pmf: $P(X=k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$ $\qquad \underbrace{\sum_{x=0}^{\infty} \dfrac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} e^{\lambda} = 1}$

CDF: $P(X \le k) = \displaystyle\sum_{x=0}^{k} e^{-\lambda} \frac{\lambda^x}{x!}$

- Given λ the rate parameter of the poisson distribution PMF,CDF can be derived as shown.(the derivations are obtained from lot of observations empirically)

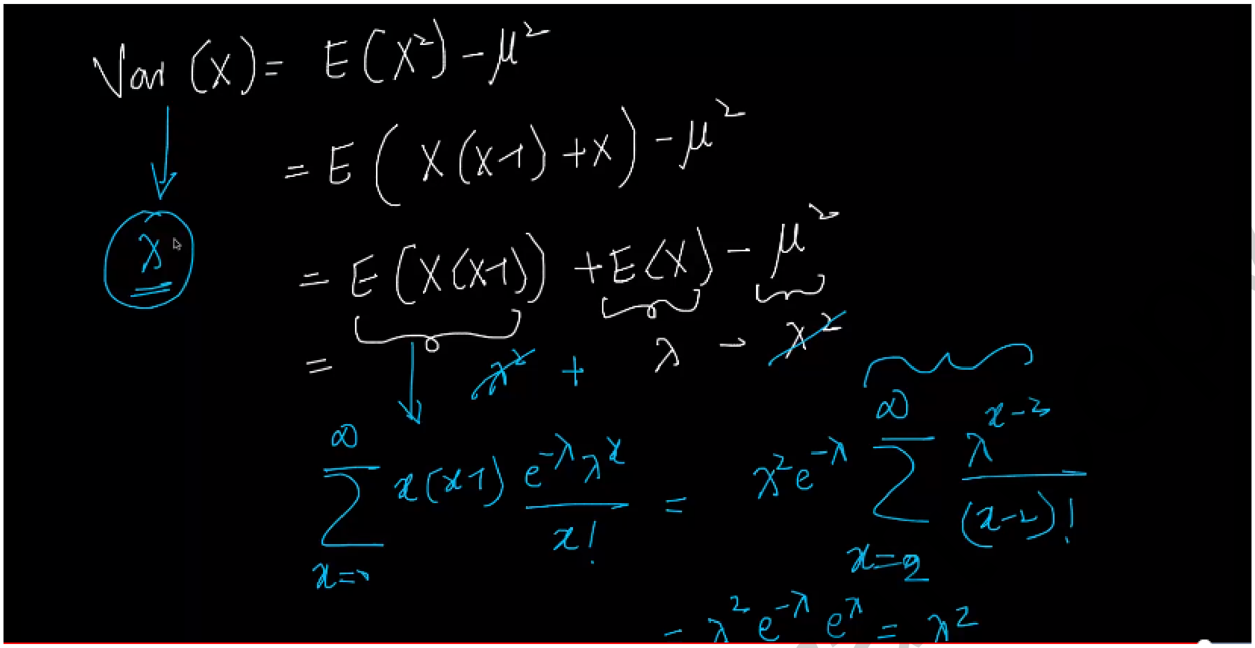
Timestamp 18.32

$$\mu = E[X] = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda} \lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda} \underbrace{\sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}}$$

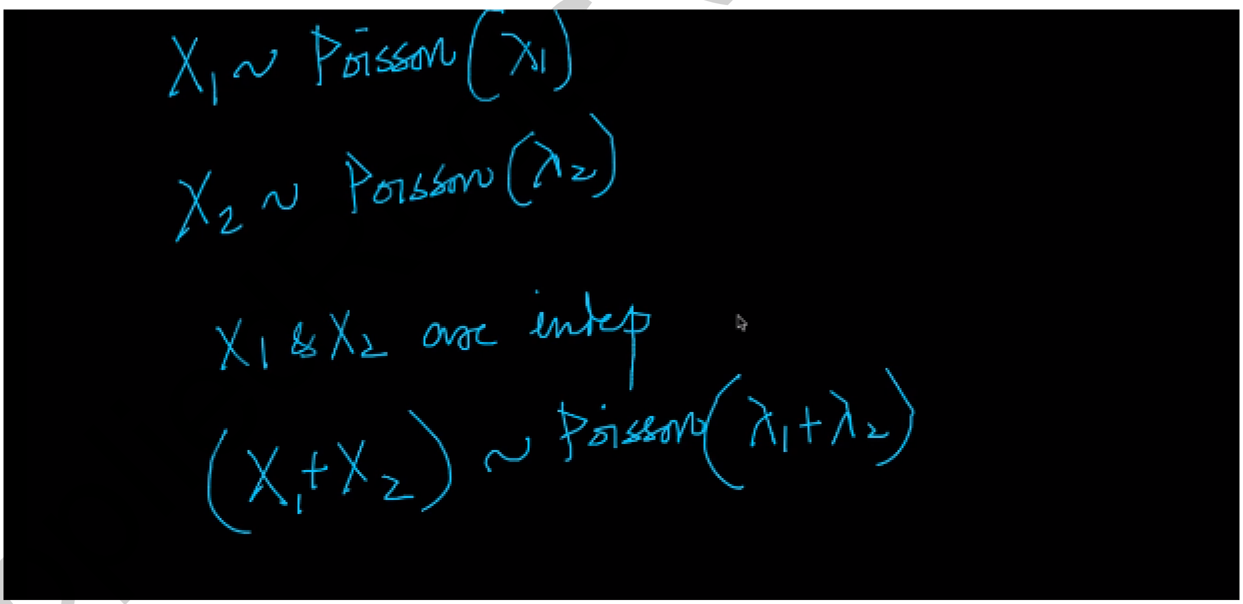
$$= \lambda e^{-\lambda} \left\{ \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \ldots \right\}$$

$$= \lambda \cancel{e^{-\lambda}} \cancel{e^{\lambda}} = \lambda$$



The mean and variance of poisson distribution is λ

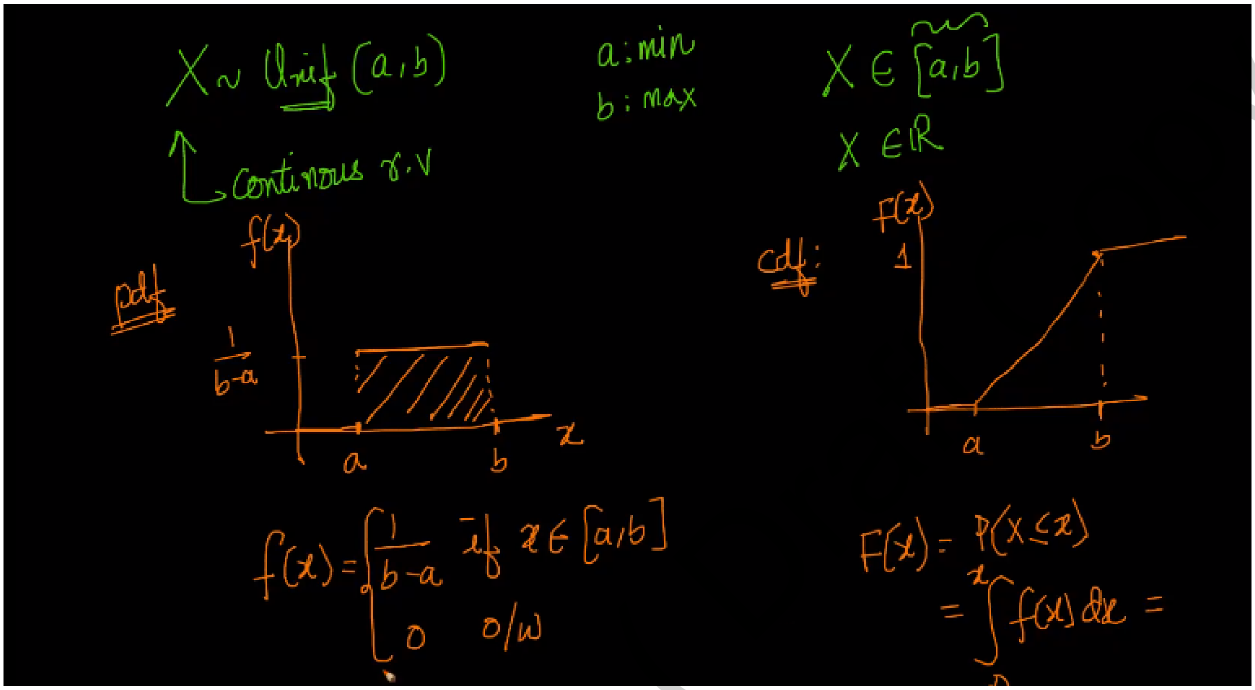
Timestamp 21.59

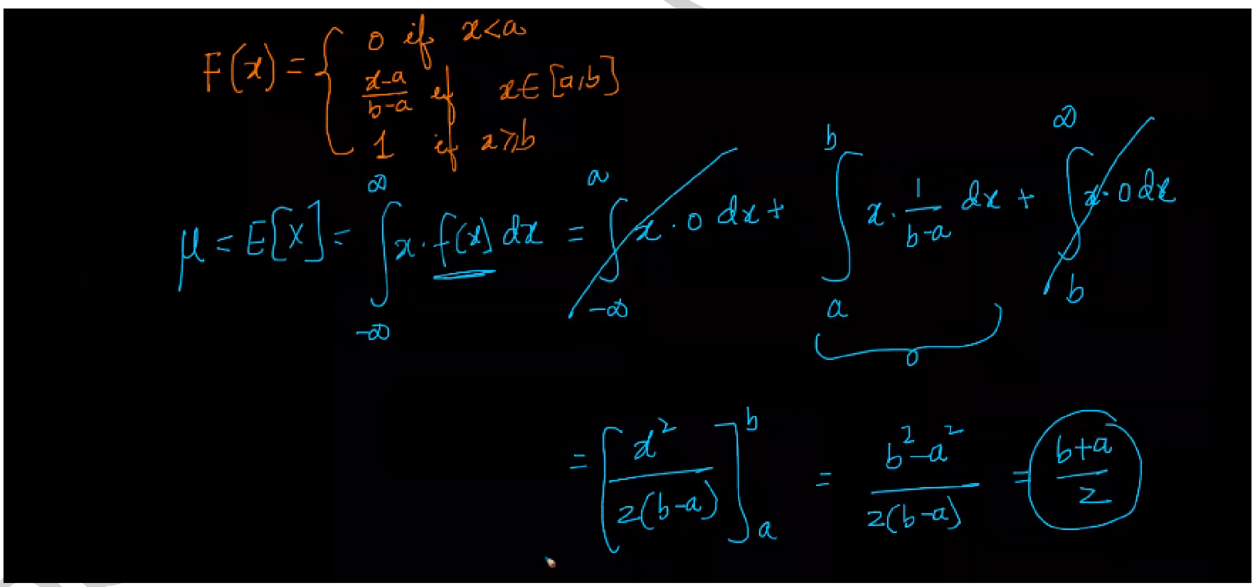


- Interesting property of a poisson variable is If X1 X2 are two random variables belonging to poisson distribution with λ1 and λ2 as rates and X1 and X2 are independent then the random variable X1+X2 also belongs to poisson distribution with rate λ1+ λ2

21.17 Uniform (continuous) distribution

Timestamp 9.17





estimation

$$\mathrm{Var}(X) = E(X^2) - \mu^2$$

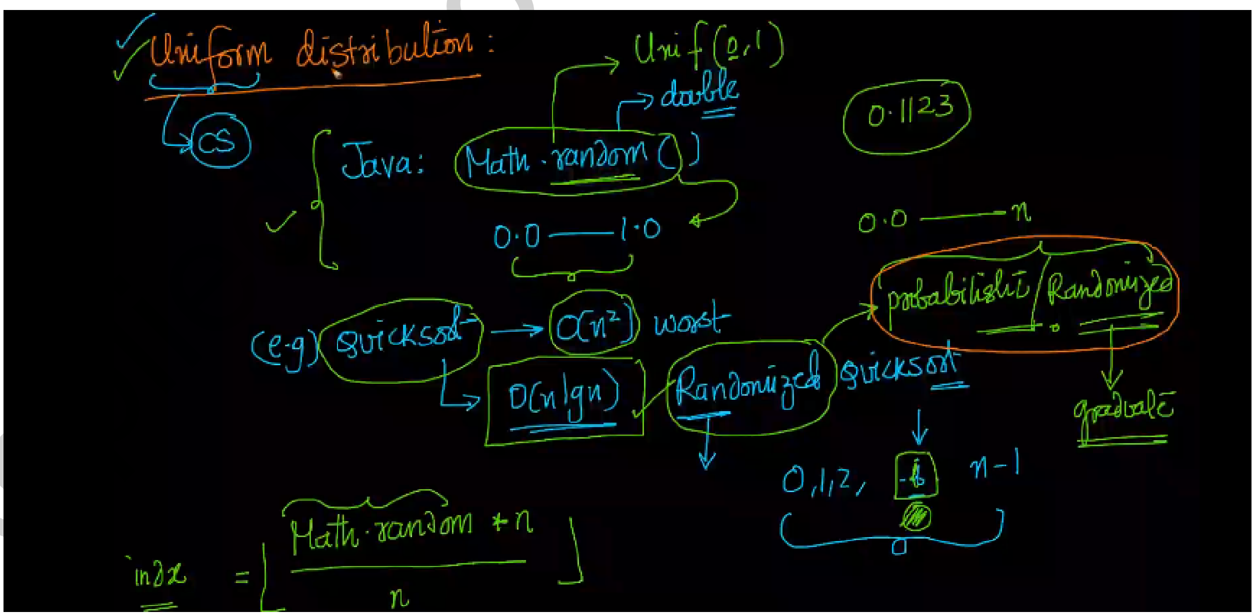
$$= \int_a^b x^2 \frac{1}{b-a}\,dx - \frac{(a+b)^2}{4}$$

$$= \left[\frac{x^3}{3(b-a)}\right]_a^b - \frac{(a+b)^2}{4}$$

$$= \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4}$$

$$= \frac{1}{3}\left(b^2+ab+a^2\right) - \frac{(a+b)^2}{4} \rightarrow \frac{1}{12}(b-a)^2$$

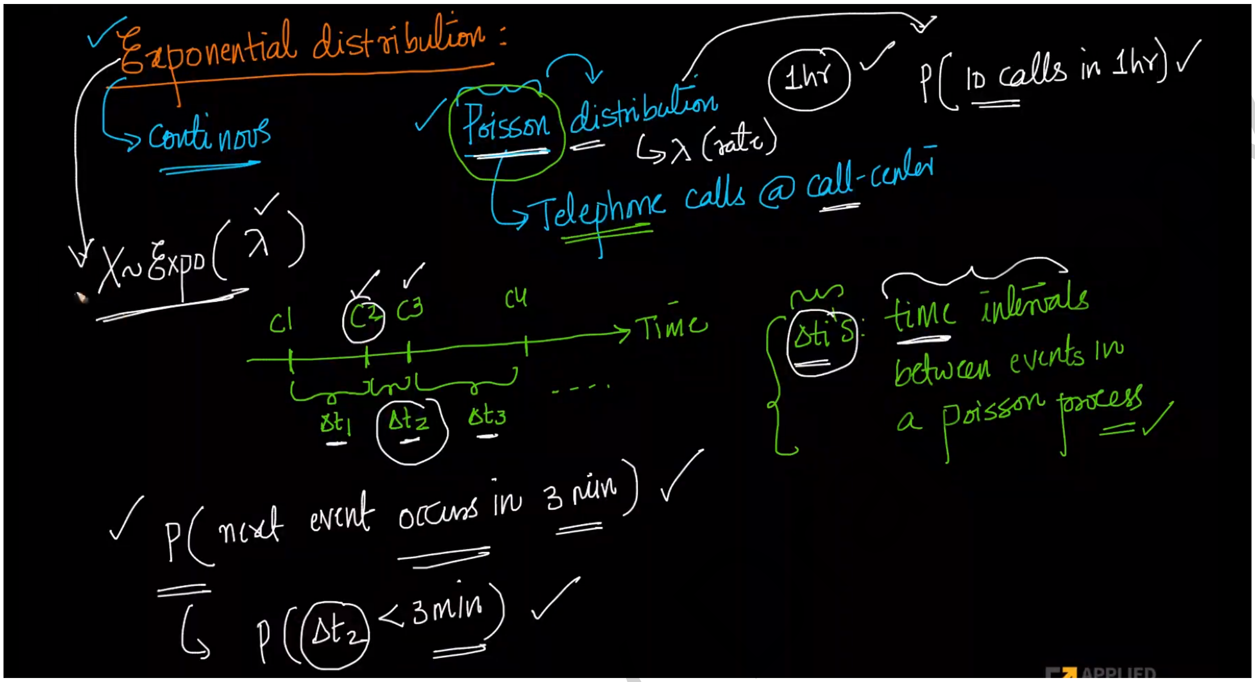
- A uniform random variable X can take value in between a,b including a,b .The PDF,CDF,mean,variance can be defined as shown above



- There are lots of applications for uniform random variables in computer science especially in probabilistic and randomized algorithms.

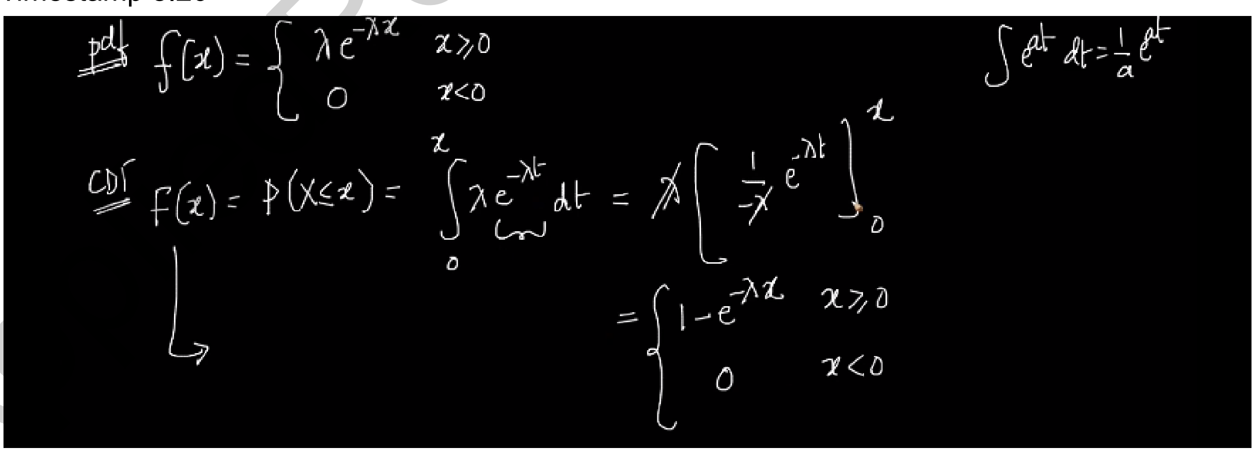
21.18 Exponential Distribution

Timestamp 6.22



- In the example of Telephone calls in a call center ,using poisson distribution we can answer questions like probability of getting 10 calls in one hour if we know λ,but we cannot answer questions like probability that the next event occurs in three minutes.
- Δt ‘s are the time intervals between poisson process.Exponential distribution knows about Δt so we'll be able to answer such questions.

Timestamp 8.20



- The PDF and CDF of Exponential distribution are as shown above

Timestamp 14.20

$$\underline{\underline{\mu}} = E(X) = \int_0^\infty x\cdot(\lambda e^{-\lambda x})\,dx$$

$\longrightarrow$ substitution ✓

$\longrightarrow$ Integration by parts

$$\int u\,dv = uv - \int v\,du$$

$$u = x$$
$$du = dx$$
$$dv = \lambda e^{-\lambda x}$$
$$v = -e^{-\lambda x}$$

$$\frac{1}{e^{\lambda x}}$$

$$= \left[-\underline{x}\, e^{-\lambda x}\right]_0^\infty + \int_0^\infty + e^{-\lambda x}\,dx$$

✓ $\lim_{x\to\infty}$

$$= 0 + \left[\frac{1}{-\lambda} e^{-\lambda x}\right]_0^\infty = \boxed{\frac{1}{\lambda}}$$ ✓

$$Var(X) = \underbrace{E(X^2)}_{} - \mu^2 \qquad \mu = 1/\lambda$$

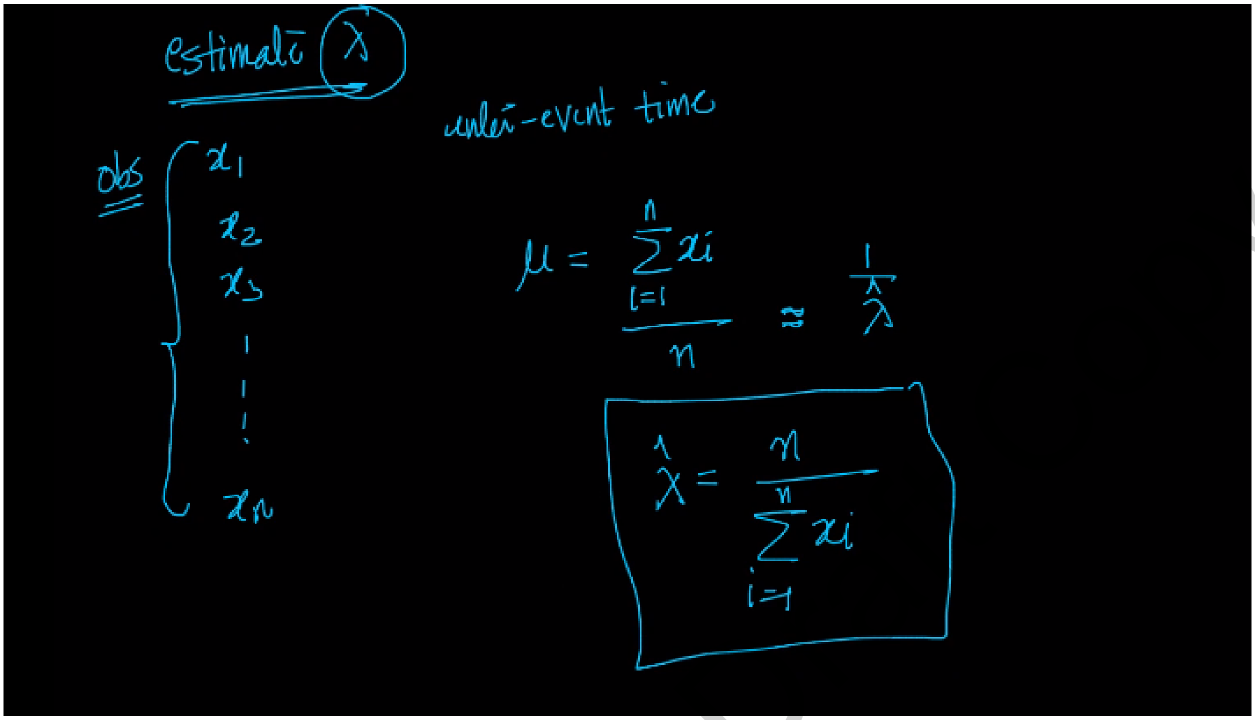
$$\int_0^\infty \underbrace{x^2}_{u}\ \underbrace{\lambda e^{-\lambda x}}_{dv}\,dx$$

Integration by parts

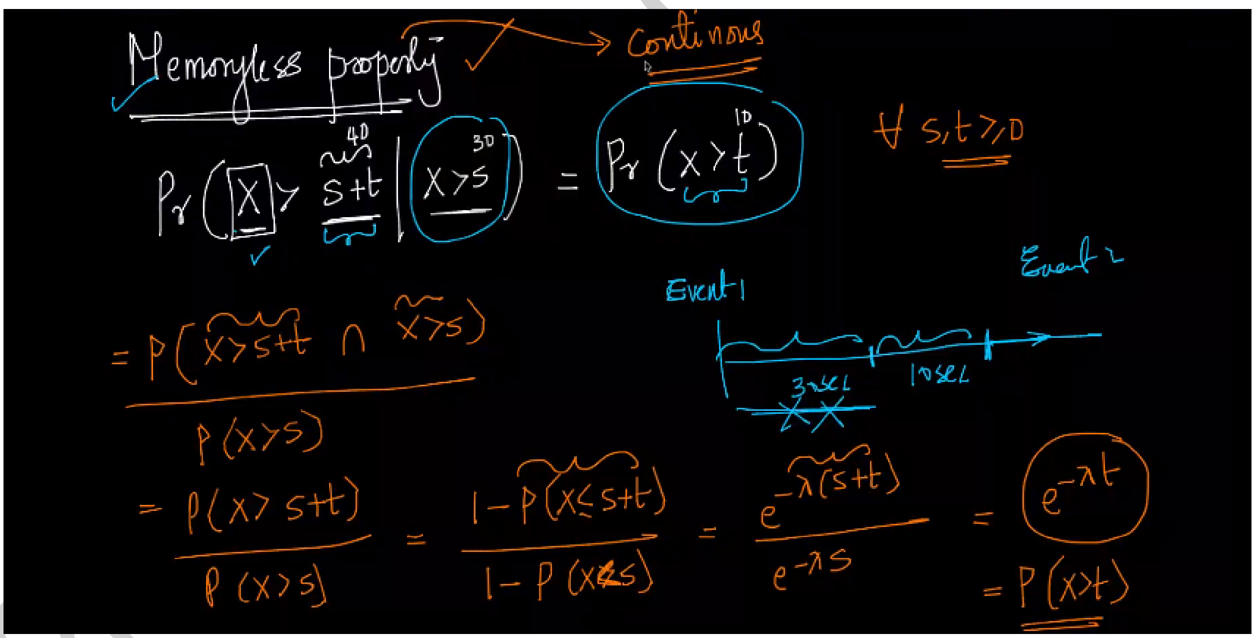
$$= \boxed{\frac{1}{\lambda^2}}$$

- Mean and variance of exponential distribution can be derived as shown above

Timestamp 16.52



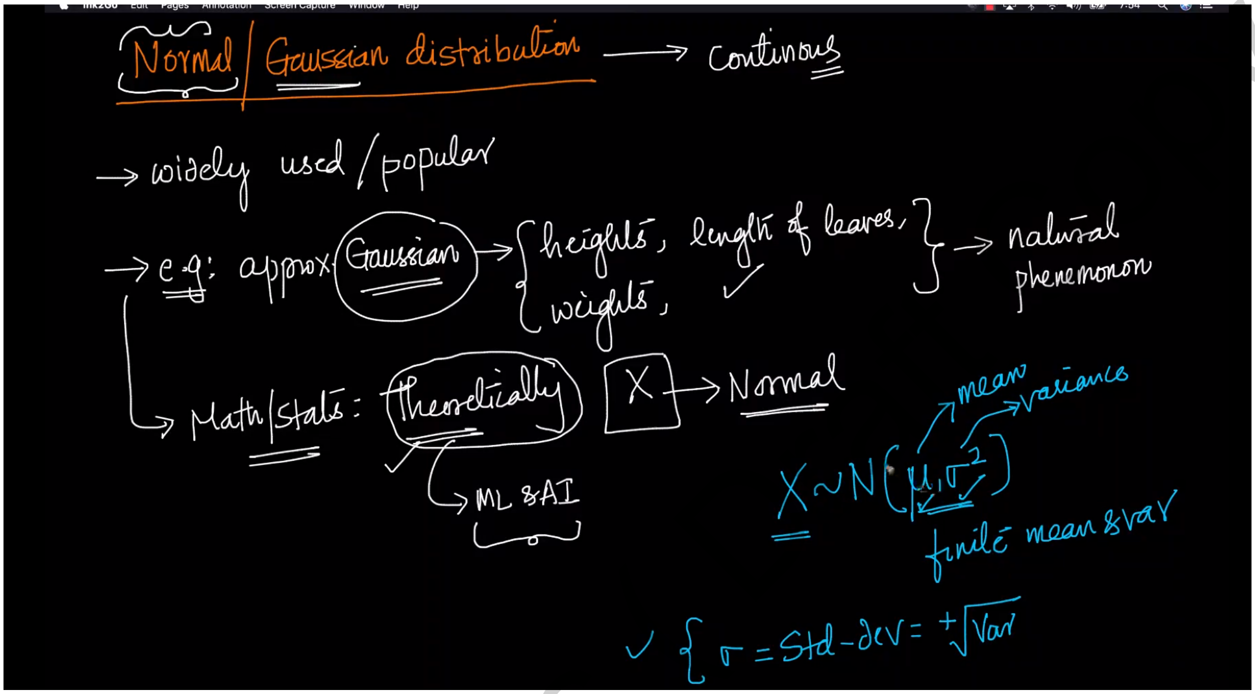
- We can estimate lambda using inter event time as shown above



- Above is an important property of the exponential distribution

21.19 Normal/gaussian Distribution

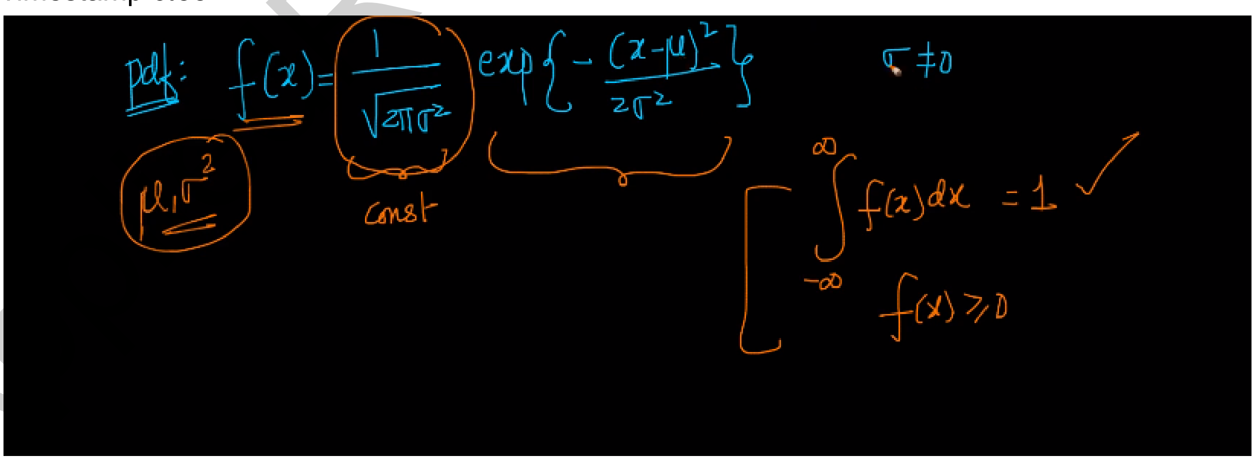
Timestamp 4.00

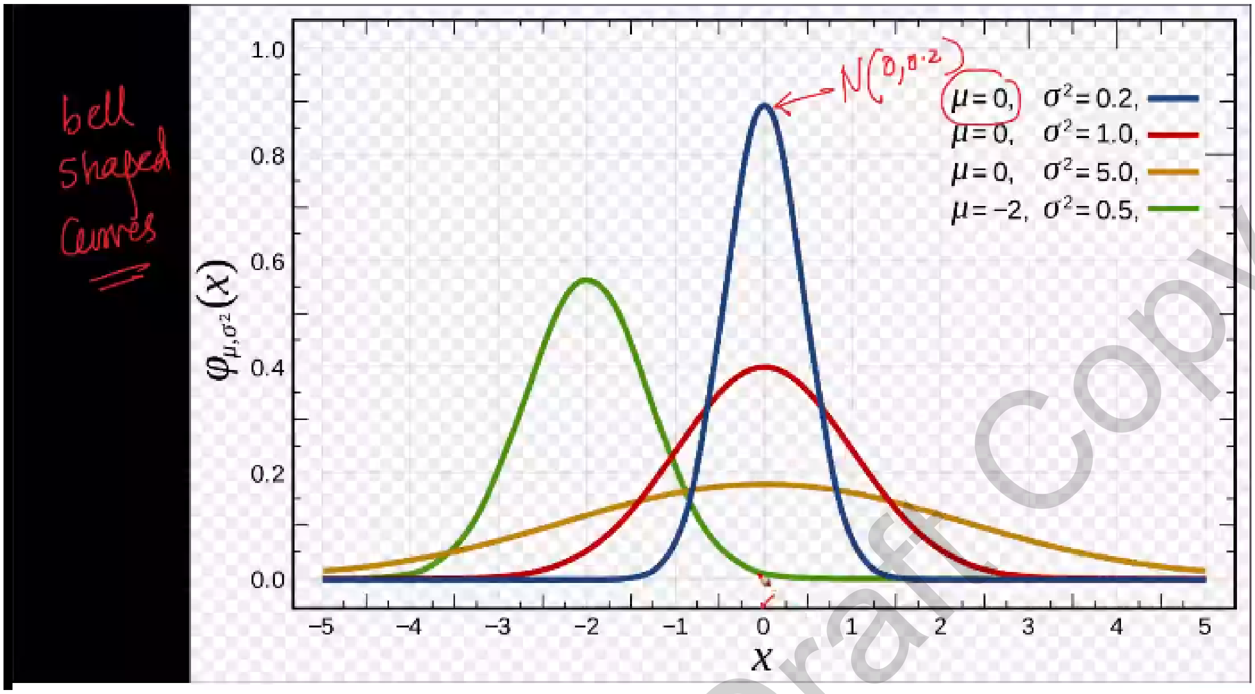


- Normal distribution is a widely used distribution by statisticians and mathematicians

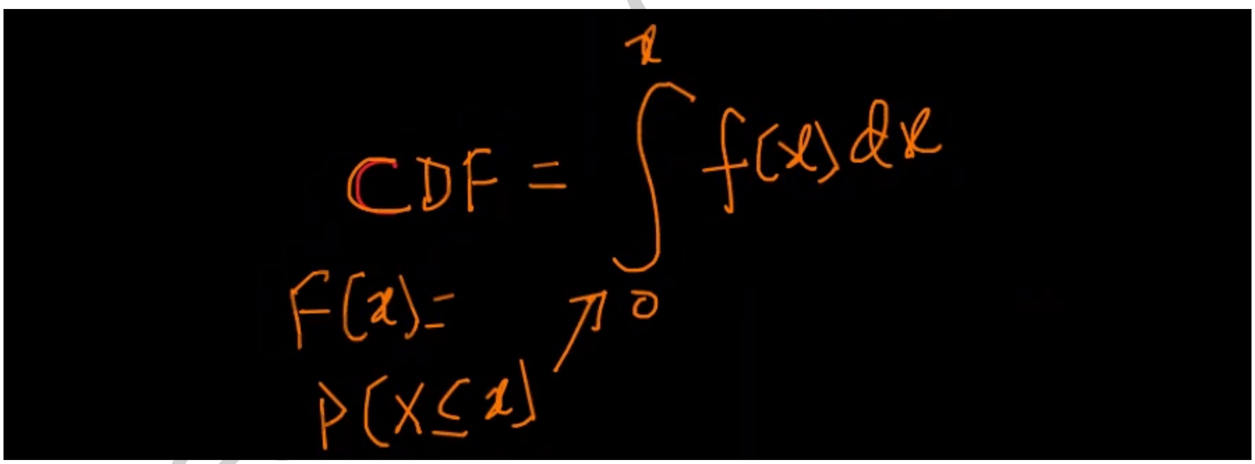
For a random variable to be normally distributed it needs to have a finite mean and finite variance

Timestamp 6.55

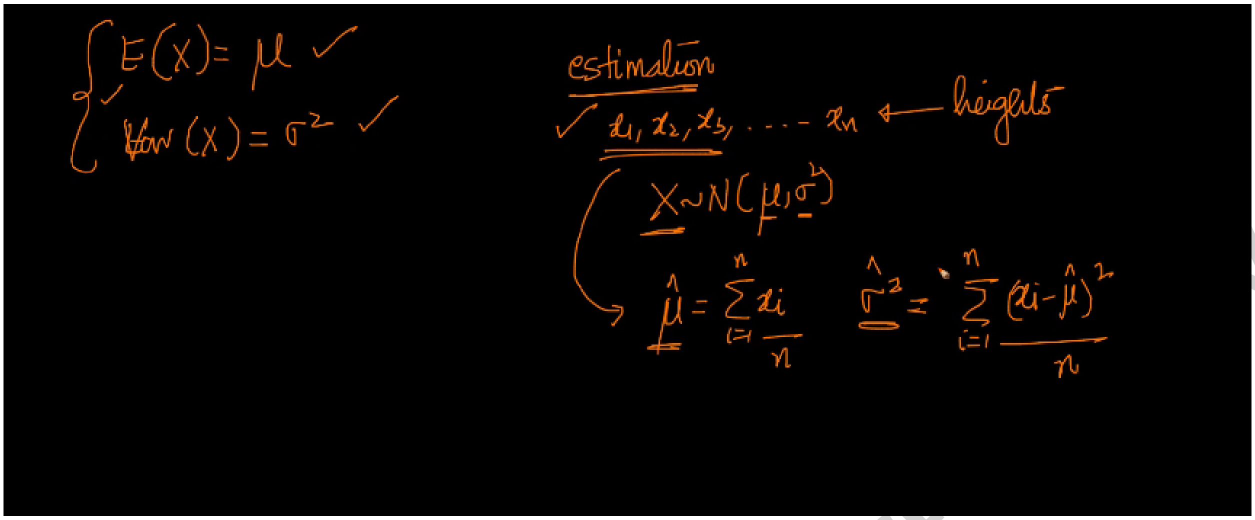




- We will have mean and variance of Normal Distribution as shown above we can easily calculate the PDF function
- PDF Curve of Normal Distribution are bell-shaped

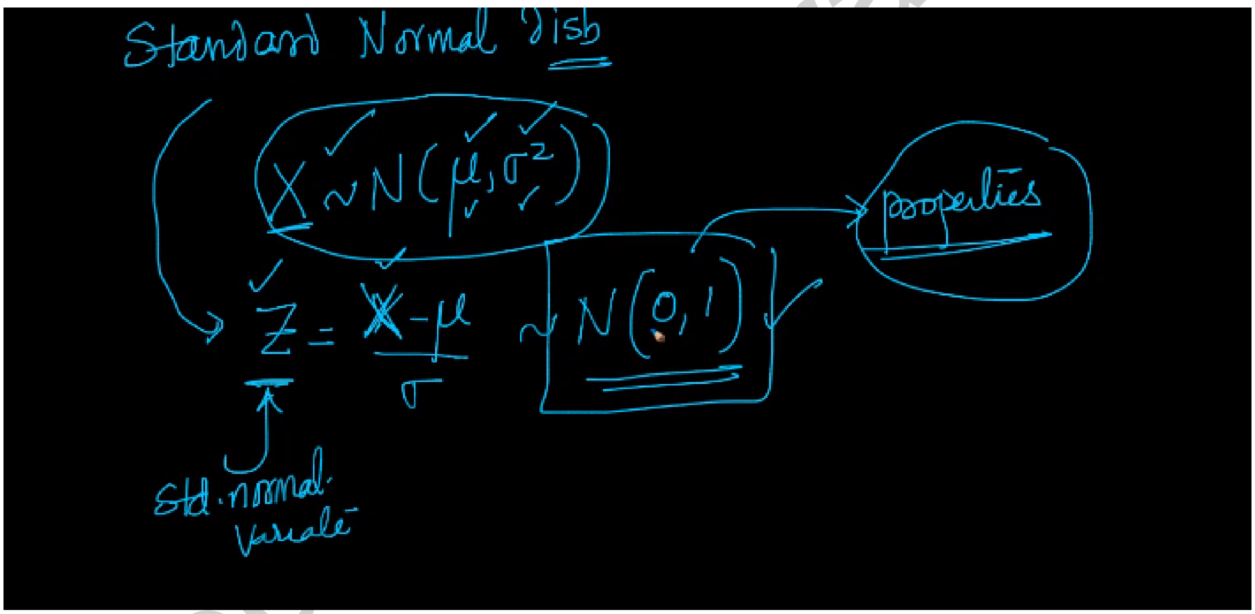


- CDF of the normal distribution



- As shown above you can estimate the mean and variance of the normal distribution

Timestamp 17.02



- We construct a new variable Z using the existing variable by subtracting the variable from mean and dividing it by variance.It is called a standard normal variable belonging to standard normal distribution. Z will have approximately a mean value 0 and variance 1.