INDIAN INSTITUTE OF TECHNOLOGY
ROORKEE

NATIONAL PROGRAMME ON TECHNOLOGY
ENHANCED LEARNING
(NPTEL)

Discrete Mathematics

Module-04
Discrete Probability
Lecture-04
Information and mutual information

With
Dr. Aditi Gangopadhyay
Department of Mathematics
IIT Roorkee

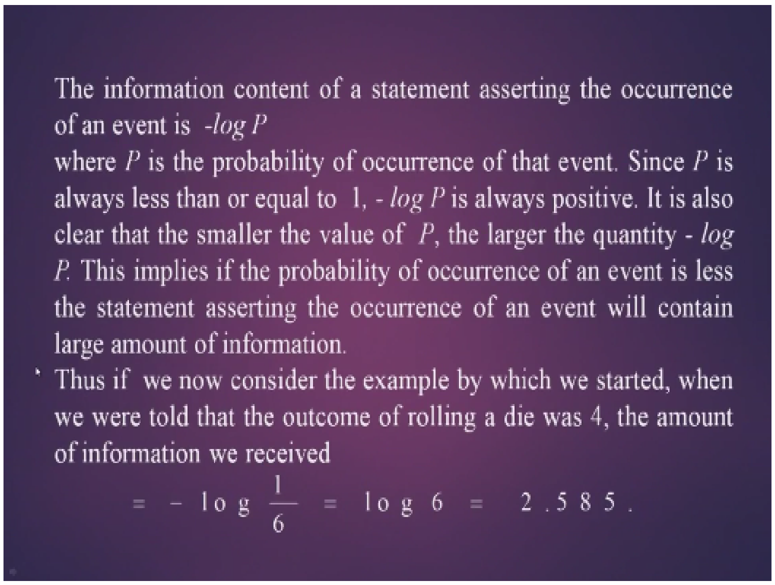
The topic of the lecture is information and mutual information suppose a die has phases 1 and 4 with red color and phases 2, 3, 5 and 6 with black color now suppose this kind of die was rolled and we were told that the outcome was 4 clearly we were given all the information concerning the outcome of the experiment if we were told that the outcome was red we would agree that we were given some information but not all the outcome is narrowed to one of two possibilities on the other hand if we were told that the outcome was black we would feel that we were given some information.

But even less the outcome is narrowed to one of four possibilities another example suppose after 6 weeks of the semester students were told that there will be a one-hour examination clearly such an announcement contains a certain amount of information however if the students were told after 1 week of classes that there will be one hour examination we would say the announcement contains much more information because it is quite honest unexpected that an examination would be she jeweled after only one week of classes so this is a surprising statement so that is why it will contain more information.

So from these examples it is clear that it is very important to measure quantitatively how much information a certain piece of message carries if a statement tells us the occurrence of a certain event that is likely to happen we would say that the statement contains a small amount of information on the other hand if a statement tells us the occurrence of a certain event that is not likely to happen we would say that the statement contains a large amount of information.

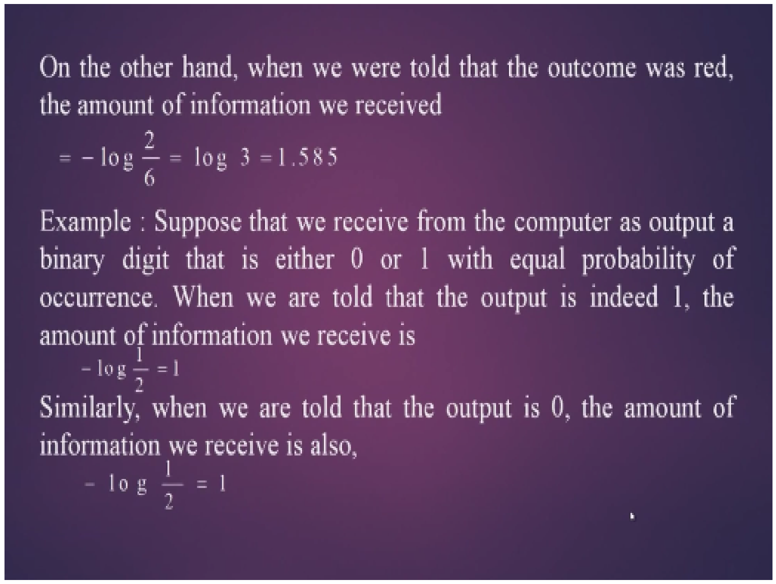
So this observation suggests that the information contained in a statement asserting the occurrence of an event depends on the probability of occurrence of the event the information content of a statement asserting the occurrence of an event is - log P.

(Refer Slide Time: 04:54)



Where P is the probability of occurrence of that event since P is always less than or equal to 1 - log P is always positive it is also clear that the smaller the value of P the larger the quantity - log P this implies if the probability of occurrence of an event is less the statement asserting the occurrence of an event will contain large amount of information thus if we now consider the example by which we started when we were told that the outcome of rolling a die was for the amount of information we received =- log 1 / 6 and that is log 6 which is 2 . 585.

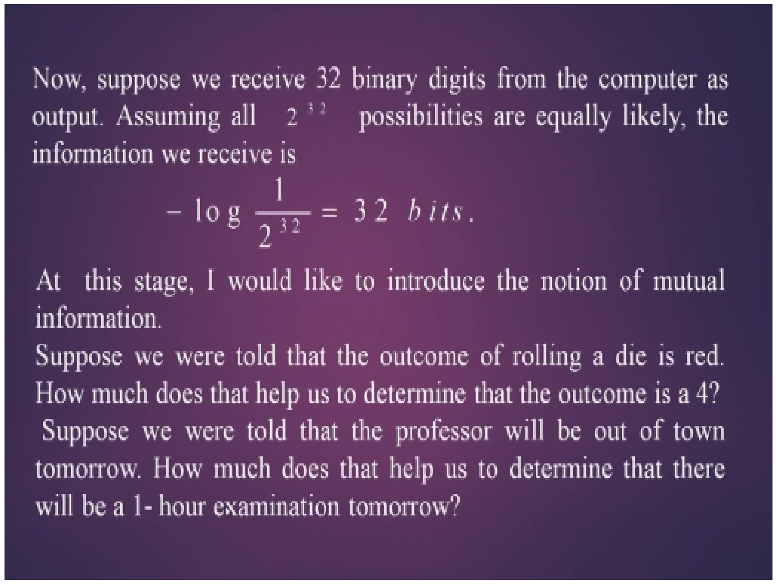
(Refer Slide Time: 06:16)



On the other hand when we were told that the outcome was red, the amount of information we received = - log 2 / 6 because if red is red has occurred then the probability will become 2 / 6 because there are 2 possibilities only 1 and 4 so that is why number of favorable cases will become 2 and that is why the probability will become 2 / 6 and as a result of this amount of information will be - log 2 / 6 and that will be log 3 which is 1. 585 another example suppose that we receive from the computer as output a binary digit that is either 0 or 1 with equal probability of occurrence.

When we are told that the output is indeed 1 the amount of information we receive is - log ½ because there are only 2 possibilities 0 and 1 so that is why probability will be 1 / 2 and the amount of information will become - log ½ that is 1 similarly when we are told that the output is 0 the amount of information we received is also - log ½ so it will be 1 now suppose we receive 32 binary digits from the computer as output assuming all 2, 2^{32} possibilities are equally likely the information we receive is - log $½^{32}$.

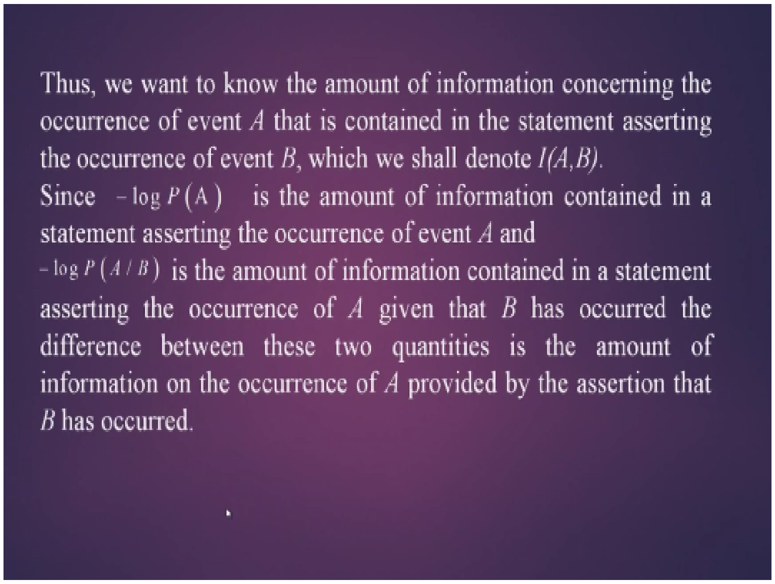
(Refer Slide Time: 08:53)



So there are all total 2^{32} possibilities so the probability and there you equivalence oh so that is why the probability will become $\frac{1}{2}^{32}$ so the amount of information will become - log 1 /2^{32} and that is 32 bits so the amount of information we are getting 32 bits at this stage I would like to introduce the notion of mutual information suppose we were told that the outcome of rolling a die is red how much does that help us to determine that the outcome is a 4 suppose we were told that the professor will be out of town tomorrow how much does that help us to determine that there will be one hour examination tomorrow.

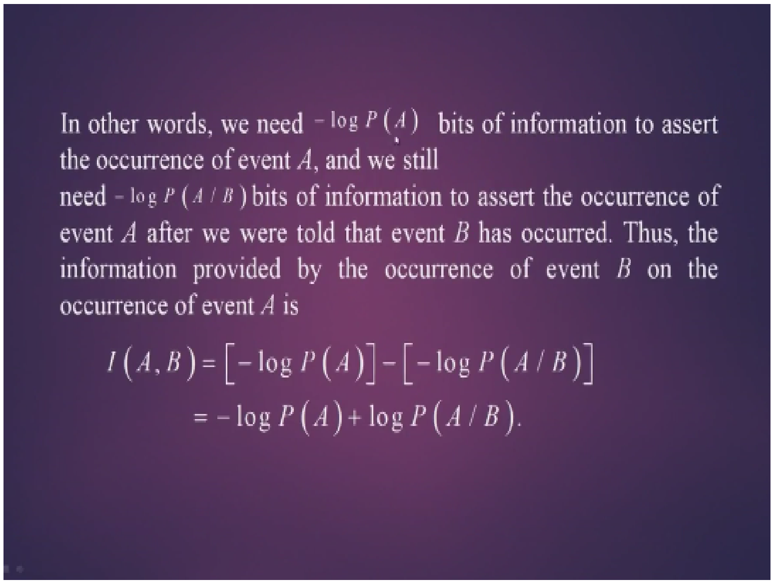
So this kind of questions can be tackled by mutual information thus we want to know the amount of information concerning.

(Refer Slide Time: 10:43)



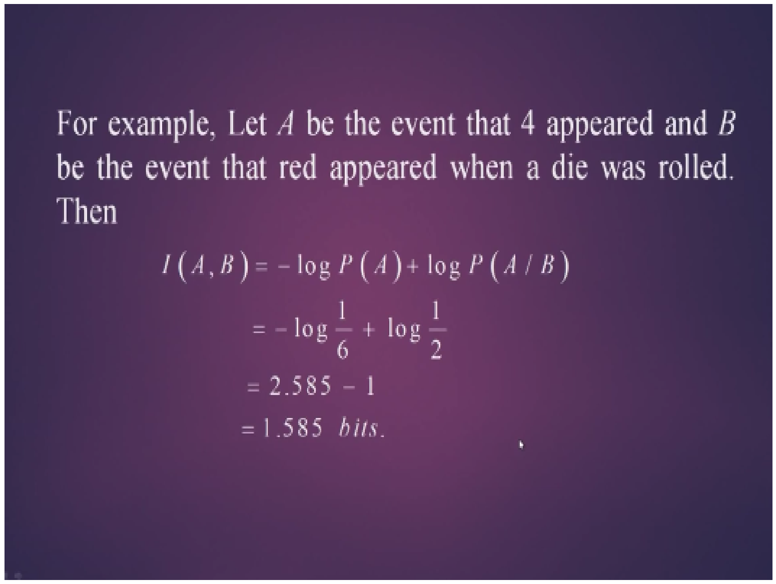
The occurrence of event A that is contained in the statement asserting the occurrence of event B which we shall denote by I (A, B) since - log P(A) is the amount of information contained in a statement asserting the occurrence of event A and - log P(A/B) is the amount of information contained in a statement asserting the occurrence of A given that B has occurred the difference between these two quantities is the amount of information on the occurrence of A provided by the assertion that B has occurred.

(Refer Slide Time: 12:01)



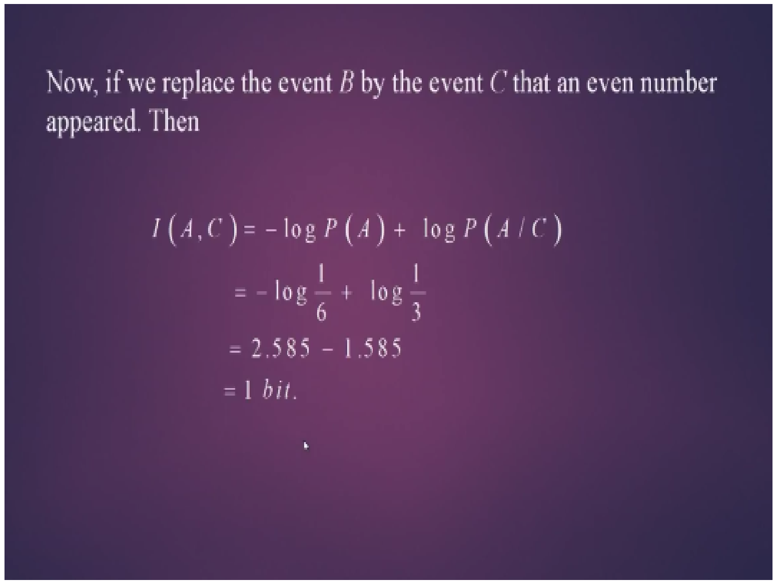
In other words we need - log P(A) bits of information to assert the occurrence of event A and we still need - log P(A/B) bits of information to assert the occurrence of event a after we were told that even B has occurred thus the information provided by the occurrence of event B on the occurrence of event A is I(A,B) that is I (A, B) that =- log P(A) that is - log P(A)-[- log P(A/B)] so this is the conditional probability so this =- log P(A) + log P(A/B).

(Refer Slide Time: 13:31)



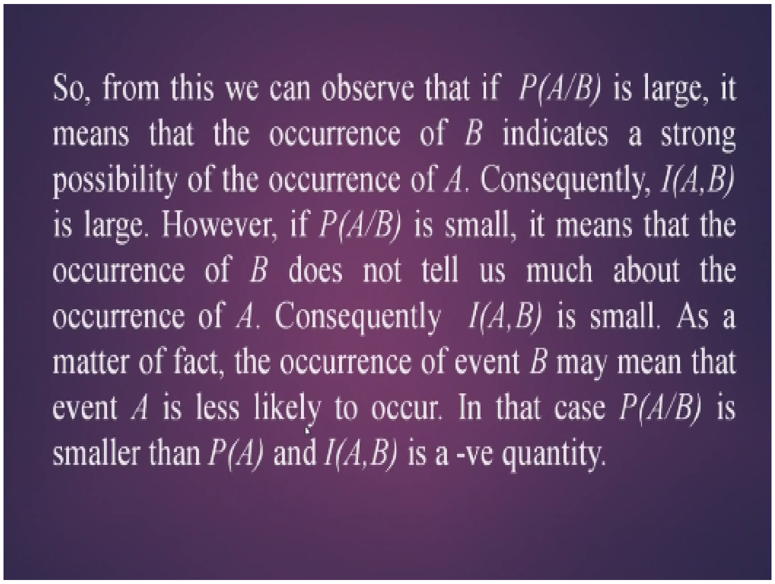
For example let a be the event that 4 appeared and B be the event that red appeared when a die was rolled then I(A, B) =- log P(A) + log P(A/B) so this will be = - log 1 / 6 because here the event A is that 4 has appeared so there are 6 possibilities 1, 2, 3, 4, 5,6 among that 4 has occurred with probability 1 / 6 and log P(A/B) what is this it is given that red has appeared so this event is given so in that case we will have only 2 possibilities 1 and 4 so that is why the probability that A/ B will become 1/2 . So this will be log ½ so we will get 2.585 - 1 and this value will be 1.585 bits.

(Refer Slide Time: 15:18)



Now if we replace the event B by the event C that an even number appeared then what will happen, so B is replaced by C so that is why we have to find I(A, C) and that =- log P(A) + log P(A/C) so - log P(A) that is same that is - log 1 / 6 + now what is P(A/C) if it is given that even number has occurred so that means there are only 3 possibilities 2,4 and 6 so in that case 4 will occur with probability 1 / 3 so that is why this will become log 1 / 3 so we will get 2.585 - 1.585 which is 1 bit.

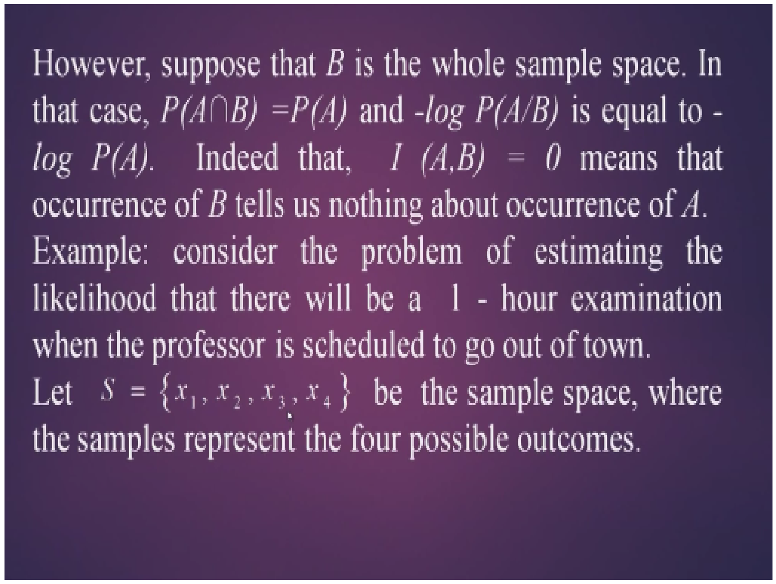
(Refer Slide Time: 16:48)



So from this we can observe that if P(A/B) is large it means that the occurrence of B indicates a strong possibility of the occurrence of a consequently I(A, B) is large however if P(A/B) is small it means that the occurrence of B does not tell us much about the occurrence of A consequently I (A,B) is small as a matter of fact the occurrence of event B may mean that event A is less likely to occur in that case P(A/B) is smaller than P(A) and I (A, B) becomes negative.

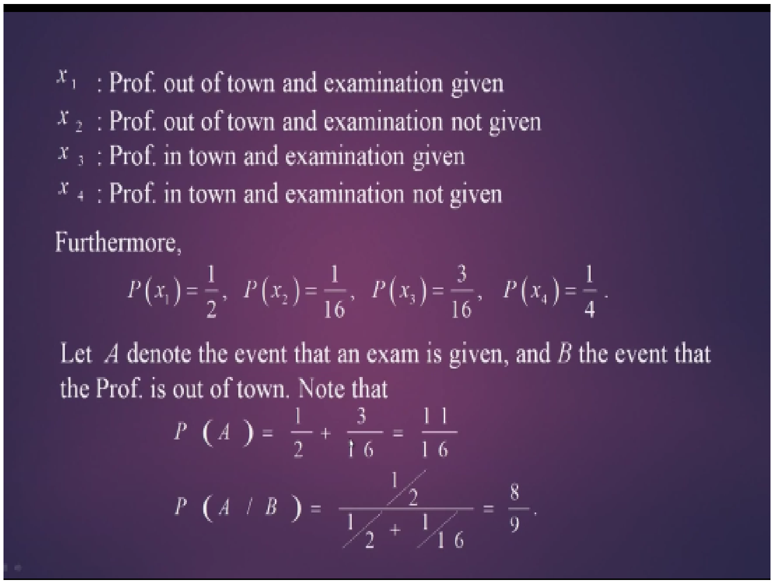
Let us also examine some extreme cases suppose that B is a subset of A in S in that case intuitively the occurrence of B assures the occurrence of a since we have $P(A \cap B) = P(B)$ it follows that $P(A/B) = 1$ and $-\log P(A/B) = 0$ that is the mutual information provided by the assertion that B has occurred on the occurrence of A =the information provided by the assertion that A has occurred however suppose that B is the whole sample space in that case $P(A \cap B) = P(A)$ and $-\log P(A/B) = -\log P(A)$.

(Refer Slide Time: 20:02)



And I (A, B) =0 and this means that occurrence of B tells us nothing about the occurrence of A so here we take one example consider the problem of estimating the likelihood that there will be one hour examination when the professor is scheduled to go out of town let s is the set containing x_1, x_2, x_3 and x_4 this is the sample space where the samples represent the 4 possible outcomes.

(Refer Slide Time: 21:11)



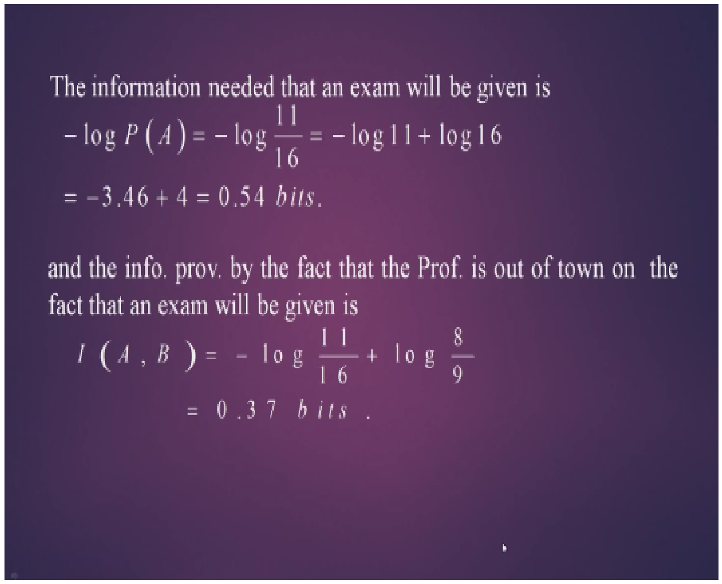
Now this can occur in 4 possibilities only professor is out of town and examination is given so that is considered as event x_1 probability professor or out of state out of town and examination not given so that is considered as x_2 professor in town and examination given that is x_3 and x_4 represents professor is in town and examination is not given so there are 4 cases which are possible here now another thing we are considering the probability is also probabilities of x_1, x_2, x_3, x_4 probability of x_1 is considered as ½ probability of x_2 is 1 / 16 probability of x_3 is 3 / 16 probability of x_4 is 1 / 4.

Let a denote the event that an exam is given and B the event that the professor is out of town note that P (A) = ½ + 3 / 16 so P (A) means probability that the that an exam is given probability what is the probability that an exam is given so that probability is P (A) and we know that we have 2 cases here x_1 and x_3 so that is why we have to add probability of x_1 and probability of 3.

So that is why it is ½ + 3 / 16 which is 11 / 16 now next we are finding the conditional P(A/B) so that is P(A∩ B)/ P(B) now what is P(B) B is the event that professor is out of town and that is occurring in 2 places x_1 and x_2 so that is why to find the P(B) we have to add the probability of x_1 and probability of x_2 so that is here in the denominator and in the numerator it is ½ because it is P(A ∩ B) that is both A and B will occur.

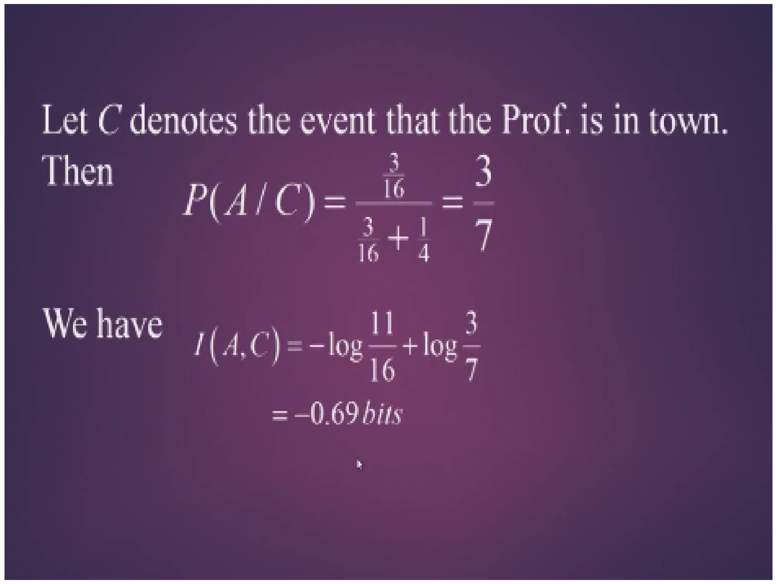
So here A and B that is exam is given and professor is out of town both these events is occurring here so probability of x_1 we have to consider in the numerator and if we simplify this we will get the value 8 / 9.

(Refer Slide Time: 25:17)



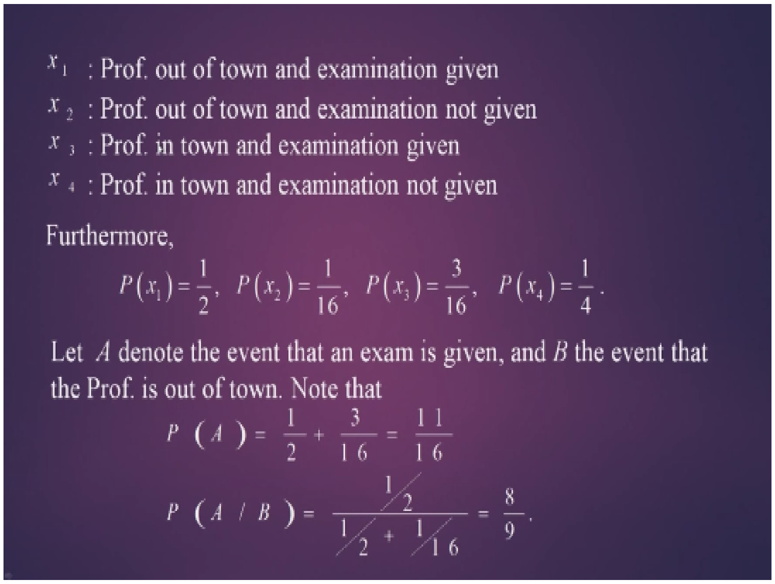
The information needed that an exam will be given is - log P(A) so it is - log 11 / 16 because P(A) is 11 / 16 so this will be - log 11 + log 16 and that will be - 3.46 + 4 which is 0.54 bits and the information that professor by professor is out of town on the fact that an exam will be given is I(A, B) =- log 11 / 16 + log 8 / 9 how is it coming here 8 is the event that exam and exam is given and B is the event that Professor is out of town so we are trying to find the mutual information of A and B.

So that is I(A, B) which is - log P(A) + log P(A/B) now - log P(A) will become - log 11 / 16 now what about P(A/B) that we have already found that is 8 / 9 so it will be log 8/9 and that is why the resulting value will be 0.37 bits.
(Refer Slide Time: 27:46)



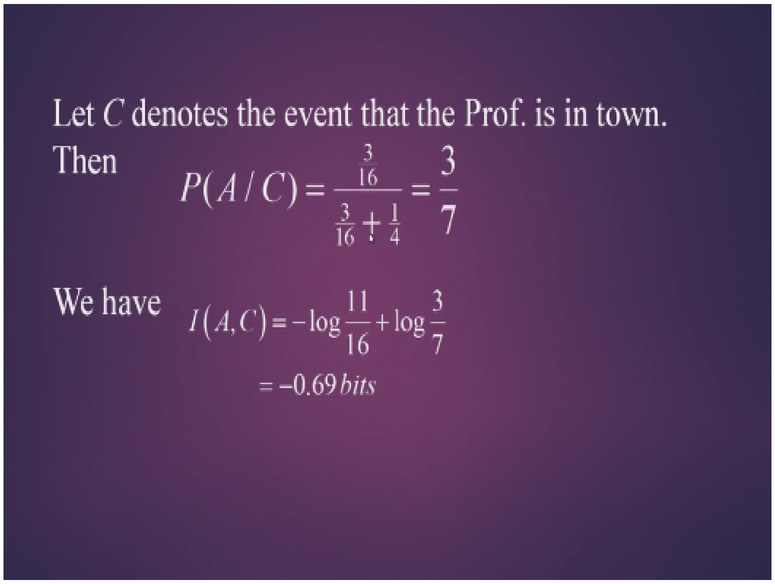
Let C denotes the event that the professor is in town then P (A/C) will become P (A∩C) / P(C).

(Refer Slide Time: 28:22)



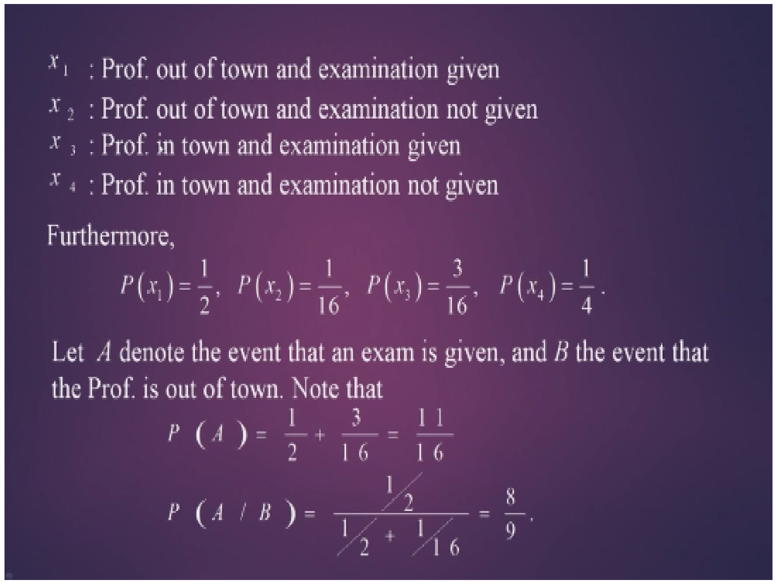
Now P(C) is the probability that the professor is in town and if we see the cases professor is in town in two cases x_3 and x_4 so we have to find the probability that professor is in town and for finding this we have to add $P(x_3)$ and $P(x_4)$ and that is what we are doing here.

(Refer Slide Time: 28:50)



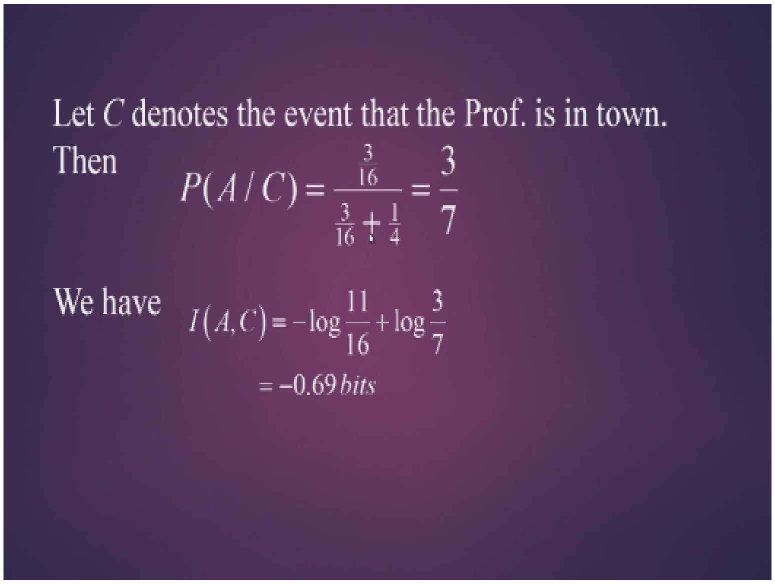
So it will be 3 / 16 + 1 / 4 and in the numerator both the events will occur that probability so probability that A and C will occur so exam is given and professor is in town that probability if we consider the cases again.

(Refer Slide Time: 29:16)



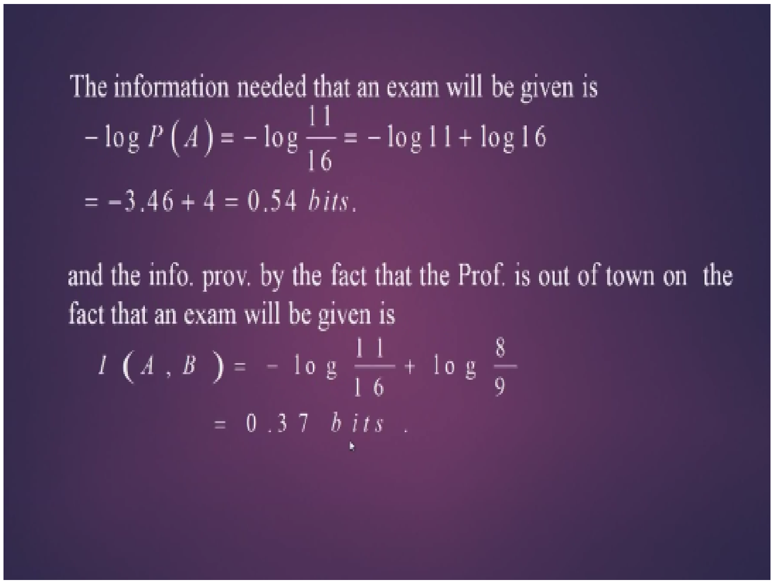
So that is occurring here professor is in town and examination is given so x_3 so that means we have to take the probability of x_3

(Refer Slide Time: 29:31)



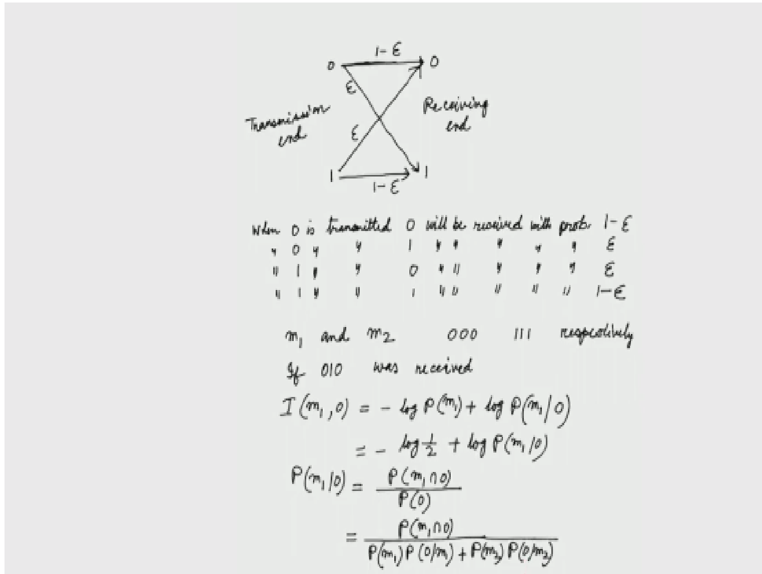
That will be 3 / 16 so if we simplify this we will get 3/7 so we will have I (A, C) that is the mutual information on taking this to events concerning this two events that is - log 11 / 16 + log 3 / 7 and that is - 0.69 bits the fact that the professor is in town makes it less likely that an examination will be given so if we see so P (A/C) this is 3/7.

(Refer Slide Time: 30:37)



But P (A/B) it is 8 / 9 so from this we can conclude this that Professor is in town makes it less likely that an examination will be given consequently the mutual information provided by the presence of the professor on the occurrence of an examination is a negative quantity so that we have already observed that is – 0.69. So the mutual information is very less now let us consider one example here.

(Refer Slide Time: 31:43)



Let us draw the figure first so this is the transmission into and this is the receiving end this is a simple model of a communication channel known as the binary symmetric channel at the transmission end either 0 or 1 is transmitted and at the receiving end either 0 or 1 is received specifically when 0 is transmitted 0 will be received with probability 1 - ε when 0 is transmitted 1 will be received with probability ε when 1 is transmitted 0 will be received with probability ε when 1 is transmitted 1 will be received with probability 1 – ε.

So this kind of communication channel we are considering suppose we have 2 equally likely messages m_1 and m_2 that will be transmitted over the channel using the representations 000 and 111 respectively if 010 was received we can compute the mutual information transmitted and the event that either part or the whole of the sequence 010 was received so if 010 was received we are trying to find the mutual information between the event that message m_1 was transmitted and event that either part or the whole of the sequence 010 was received and that will be I (m_1, 0).

Now what is this will be - log the probability that m_1 is transmitted + log the probability that 0 was received given that 0 was received the probability that m_1 was transmitted so this value we can calculate in this way this - log P (m_1) that will be - log ½ because m_1 and m_2 are equally likely so there are two possibilities m_1 or m_2 so that is why it will be ½ probability will be ½ so that is why this value will be - log 1/2.

Now what about this conditional probability this we have to find so probability that m_1 / 0 and this we know this probability that (m_1 ∩ 0) both this events will occur m_1 is transmitted as well as 0 is received this probability and in the denominator probability that 0 is received and this can be written as probability that (m1 ∩ 0) /P(m_1) into P(0 / m_1)+P (m_2) into P(0/ m_2) because this is this is P(0) this is nothing but the total probability so because there are 2 possibilities m_1 and m_2 so that is why either m1 or m2 will occur.

So P (m1) into P (0/ m_1) so this is the probability that m_1 is transmitted that is given then what is the probability that 0 is disap and next is P (m_2) into P (0, m_2) now P (0, m_2) is the probability that it is given that m_2 is transmitted then what is the probability that 0 will be received.

(Refer Slide Time: 40:18)

$1 \xrightarrow[1-\varepsilon]{} 1$

When 0 is transmitted 0 will be received with prob. $1-\varepsilon$

" 0 " " 1 " " " " " ε

" 1 " " 0 " " " " " ε

" 1 " " 1 " " " " " $1-\varepsilon$

m_1 and m_2 000 111 respectively

If 010 was received

$$I(m_1, 0) = -\log P(m_1) + \log P(m_1/0)$$

$$= -\log \tfrac{1}{2} + \log P(m_1/0)$$

$$P(m_1/0) = \frac{P(m_1 \cap 0)}{P(0)}$$

$$= \frac{P(m_1 \cap 0)}{P(m_1)P(0/m_1) + P(m_2)P(0/m_2)}$$

$$= \frac{P(m_1)P(0/m_1)}{P(m_1)P(0/m_1) + P(m_2)P(0/m_2)}$$

$$= \frac{\frac{1}{2}(1-\varepsilon)}{\frac{1}{2}(1-\varepsilon) + \frac{1}{2}\varepsilon}$$

$$I(m_1, 0) = -\log \tfrac{1}{2} + \log \frac{\frac{1}{2}(1-\varepsilon)}{\frac{1}{2}(1-\varepsilon)+\frac{1}{2}\varepsilon} = 1 + \log(1-\varepsilon)$$

So this numerator can be written as probability of m_1 into P (0, m_1)/ P (m_1) into P (0, m_1) + P (m_2) into P (0, m_2) so these probabilities we have to find. Now the numerator if you consider it will be P(m_1)into P(0 , m_1) P(m1)is ½ and P(0/ m_1)that will be 1 - ε so m_1 given that m_1 is transmitted 0 will be received with probability 1 - ε okay so this will be ½ into 1 - ε now in the denominator we will have P(m_1)so same whatever we are writing in the numerator this one is also same now the other one that will be ½ into ε so we will get I(m_1, 0) 0 =- log ½ + log ½ 1 - ε / ½ into 1 - ε + ½ ε so finally we are getting the value 1 + log 1 - ε .

(Refer Slide Time: 43:09)

$$I(m_1, 01) = -\log P(m_1) + \log P(m_1/01)$$

$$\cancel{P(m_1/01) = \frac{P(m_1 \cap 01)}{P(m_1)P(01)}}$$

$$= \frac{P(m_1 \cap 01)}{P(01)}$$

$$= \frac{P(m_1)\,P(01/m_1)}{P(m_1)P(01/m_1) + P(m_2)P(01/m_2)}$$

$$= \frac{\frac{1}{2}(1-\varepsilon)\varepsilon}{\frac{1}{2}(1-\varepsilon)\varepsilon + \frac{1}{2}\varepsilon(1-\varepsilon)}$$

$$= \frac{1}{2}$$

$$I(m_1, 01) = -\log\frac{1}{2} + \log\frac{1}{2} = 0$$

$$I(m_1, 010) = -\log\frac{1}{2} + \log\frac{\frac{1}{2}(1-\varepsilon)^2\varepsilon}{\frac{1}{2}\varepsilon(1-\varepsilon)^2 + \frac{1}{2}\varepsilon^2(1-\varepsilon)}$$

$$= 1 + \log(1-\varepsilon)$$

0 or 010 was received tells us exactly the same amount of information on the transmis of ~~msge~~ m_1

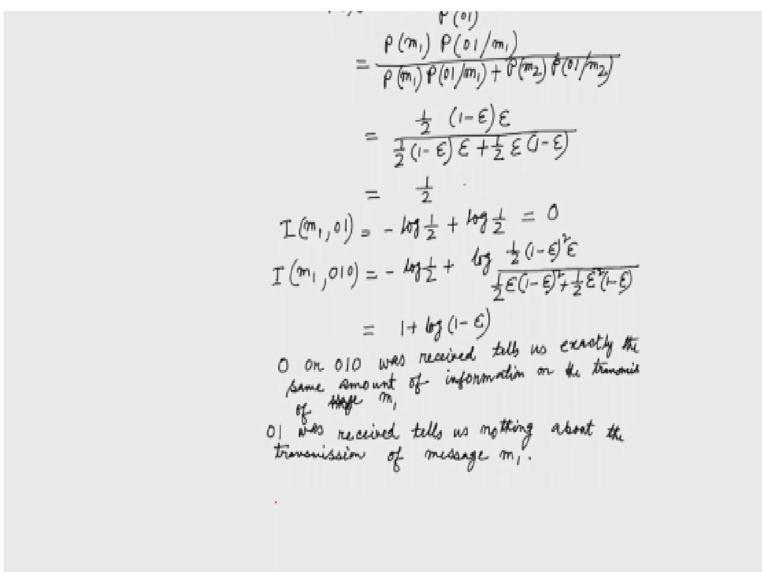
Next we are considering I (m_1, 0 1) because we are trying to find the mutual information between the event that m_1 is transmitted and the event that the whole sequence 010 or part of it is received so one part we have already considered that is 0 now we are considering another part that is 01 so I(m_1, 01) this will be - log probability that m_1 is transmitted + log probability that 0 1 is received then the probability that m_1 is transmitted so this conditional probability log of this conditional probability.

So let us find m_1 / 01 this will be probability ($m_1 \cap$ 01) / P(m_1)into P(0/01) 1 sorry 0 / 01 given so let us write it freshly here m_1 / 01 that will be probability that ($m_1 \cap$ 01) / P(01) which is P(m_1)into P(0, m_1) / P(m_1)into P(01 /m_1)+ P(m_2)into P(01/m_2) now the numerator P(m_1)is again ½ and this conditional probability will be 1 - ε into ε so m_1 given that m_1 is transmitted that means 000 that is transmitted so from 0 will occur with P(1 – ε) and from 0, 1 will be received with P(ε).

So that is why this we are getting in the denominator we will have ½ into 1 - ε into ε + this also will give the same expression ½ into ε into 1 - ε so this value will become 1/2 and as a result of this I (m_1 , 01) will be - log ½ + log ½ so this mutual information =0 in the same way if we find I (m_1 , 010) that will be – log ½ + log 1/2 into 1 - ε whole square into ε /1/2 into ε into 1 - ε whole square + ½ into ε^2 into 1 - ε so this one will give 1 + log (1 – ε).

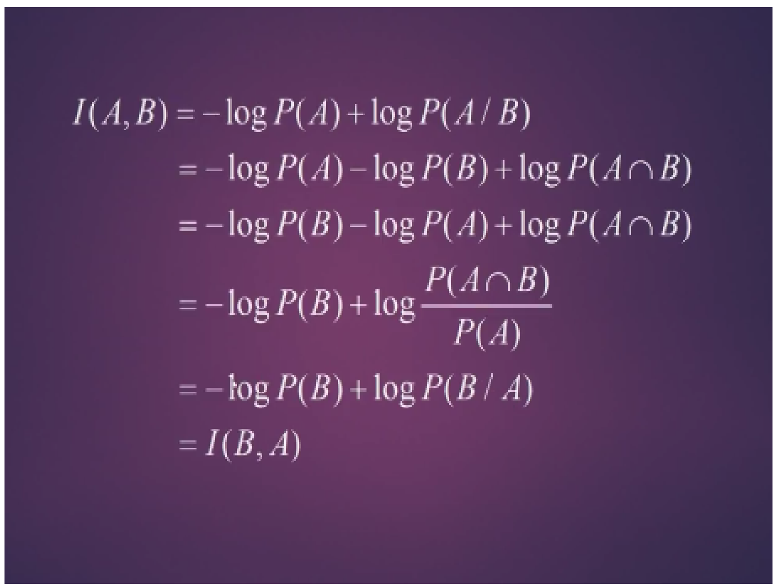
So knowing that either 0 or 010 was received tells us exactly the same amount of information on the transmission on the transmission of message m_1 however knowing that the sequence 01 was received.

(Refer Slide Time: 49:43)



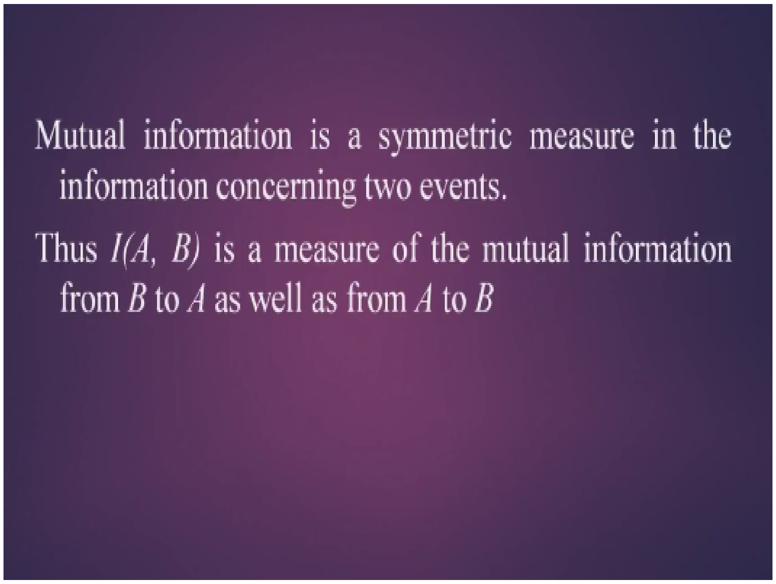
Tells us nothing about the transmission off message m_1 and this is why what we expect this is actually expected because the transmission of either m_1 or m_2 would yield the sequence 01 at the receiving end with the same probability, so now let us discuss one result of mutual information.

(Refer Slide Time: 51:08)



I(A, B) =- log P(A) + log P(A/B) and that =- log of P(A) - log P(B) + log P(A∩B) because P(A/B) is P(A∩B) / P(B) so we can write this as - log P(B) - log P(A) + log P(A ∩ B) so it will be - log P(B) + log P(A ∩ B) / P(A) and that =- log P(B) + log P(B / A) and this is nothing but I(B, A).

(Refer Slide Time: 52:28)



So this implies mutual information is a symmetric measure in the information concerning 2 events thus I(A, B) is a measure of the mutual information from B to A as well as from A to B that is all thank you.

Educational Technology Cell
Indian Institute of Technology Roorkee

Production for NPTEL
Ministry of Human Resource Development
Government of India

For Further Details **Contact**

Coordinate, Educational Technology Cell
Indian Institute of Technology Roorkee
Roorkee-247667
E Mail: etcell@iitr.ernet.in, etcell.iitrke@gmail.com
Website: www.nptel.iim.ac.in

Acknowledgement
Prof. Pradipta Banerji
Director, IIT Roorkee

Subject Expert & Script
Dr. Aditi Gangopadhyay
Dept of Mathematics

IIT Roorkee

Production Team
Neetesh Kumar
Jitender Kumar
Pankaj Saini
Meenakshi Chauhan

Camera
Sarath Koovery
Younus Salim

Online Editing
Jithin.k

Graphics
Binoy.V.P

NPTEL Coordinator
Prof.Bikash Mohanty

An Educational Technology Cell
IIT Roorkee Production

WANT TO SEE MORE LIKE THIS
SUBSCRIBE