NPTEL

NATIONAL PROGRAMME ON
TECHNOLOGY ENHANCED LEARNING

IIT BOMBAY

CDEEP
IIT BOMBAY

Quantum Information and Computing

Prof. D.K.Ghosh
Department of Physics IIT Bombay

Modul No.07

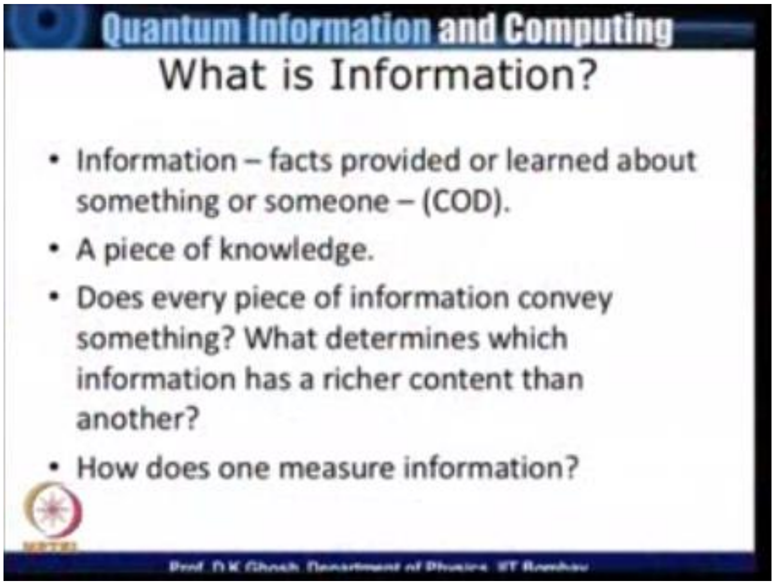
Lecture No.34

Classical Information Theory

In the last several lectures we had talked about elements of quantum computing and we culminated our discussions with discussion of two major algorithms namely a search algorithm due to Grover and the factorizing algorithm due to Shor. Towards the end of the last we have the sessions we also talked how quantum codes are corrected. So now we are ready to talk about the quantum communication various protocols but before we do that let us try to understand what does information actually convey.

And to this lecture and the next I will talking about elements of classical information theory so if you look up.

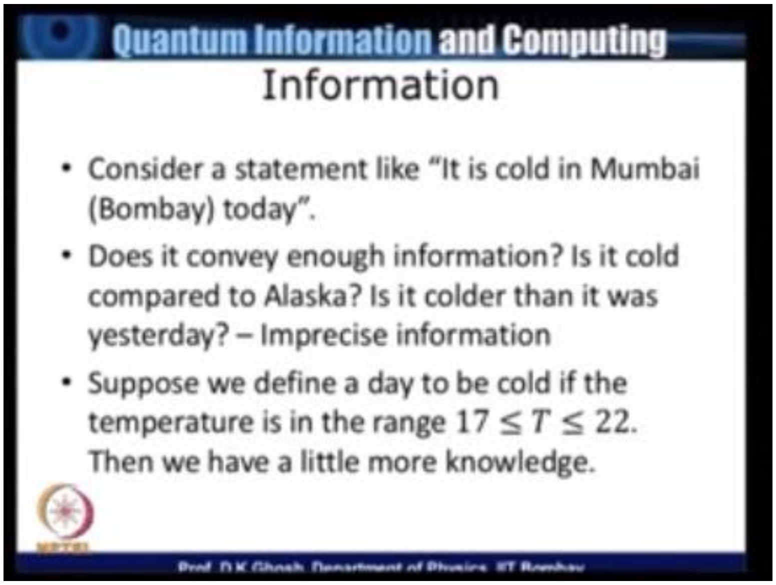
(Refer Slide Time: 01:10)



The definition of the word information in a dictionary we will find it says it is facts provided or learned about something or someone this is quoted from a concise of the dictionary. In a colloquial pair lines by the word information we understand it is acknowledged but the question is does every piece of information convey something now what determines in one piece of information has a richer content that is more important that another piece of information and more important since we are talking about science.

How we quantify it for because in common the usage the word information is not generally not quantified it is understood purely qualitative. So let us look at to be to make up our idea is clear.

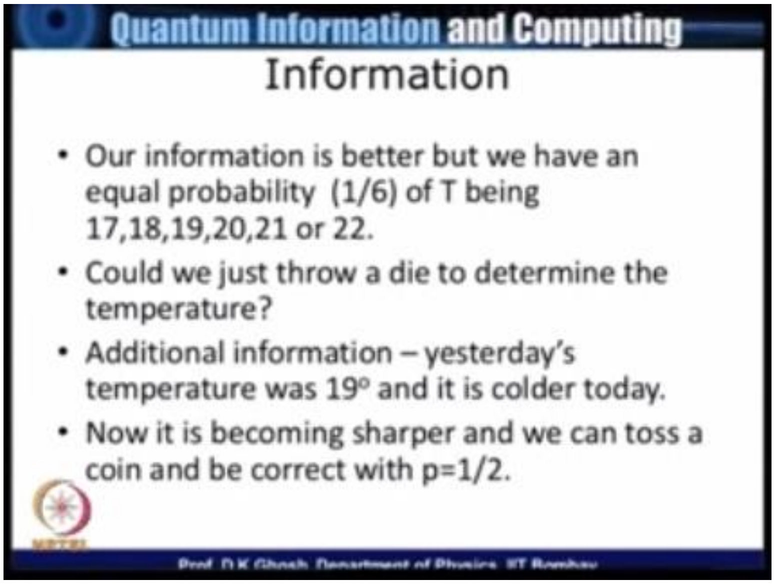
(Refer Slide Time: 02:06)



Let us consider a statement it says "it is cold in Mumbai today" now does that statement inform us something is that good enough information now the point is that look at the statement it says it is cold in Mumbai now what do it mean cold in Mumbai is it colder compare to Alaska or is that a time information saying that it is colder than what it was yesterday. In other words though this statement it is cold in Mumbai could top information you something to in connection with the Bombay resident to sort of roughly knows what you are trying to say.

But if you want to convey this to a wider audience you have to be more precise than what you have done. So first thing that we do is this that let us defined what is cold suppose looking at Indian climates generally we defined that I will say that there to the cold if the temperature range is between 17 degrees to 22 degree centigrade. Now so in that case if you make statement it is cold in Mumbai you have now got some information that the temperature is between these to guidance.

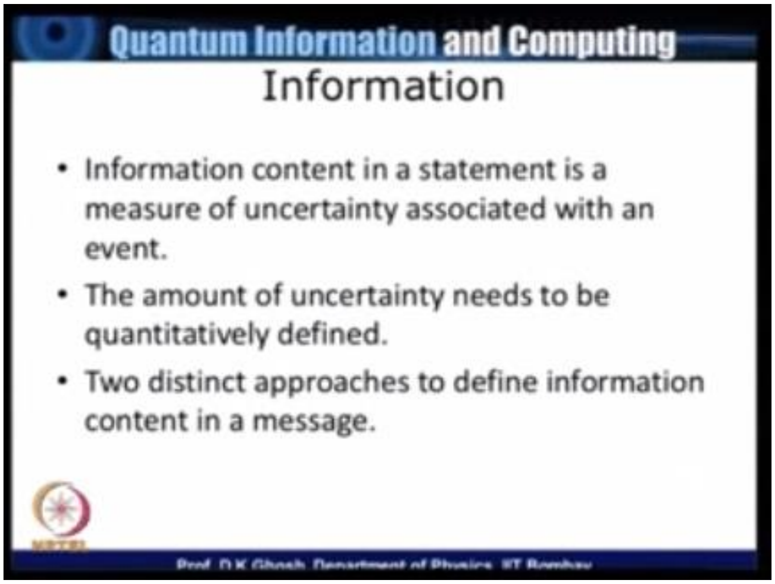
(Refer Slide Time: 03: 21)



But the information though it is better it does not tell you whether it is 17, 18, 19, 20, 21 or 22. So in principle there is an equal probability each one 6 of the temperature being any one of these options. Now so we would then say who like to find out the temperature and there are 6 alternatives which are given here why do not you just throw a die and decided what is the temperature. Now this is one way of doing it of course it will not measure the temperature but you will get some ideas.

Now suppose I made an additional information I said yesterday's temperature was 90˚ degrees and it is colder today, my notice we have gone done something better we have now said that the, our information has become sharper and our temperature range have now become smaller and supposing I had just two alternatives here then I could actually toss a coin and decide what the temperature was to certain degree of what cold was.

(Refer Slide Time: 04:25)

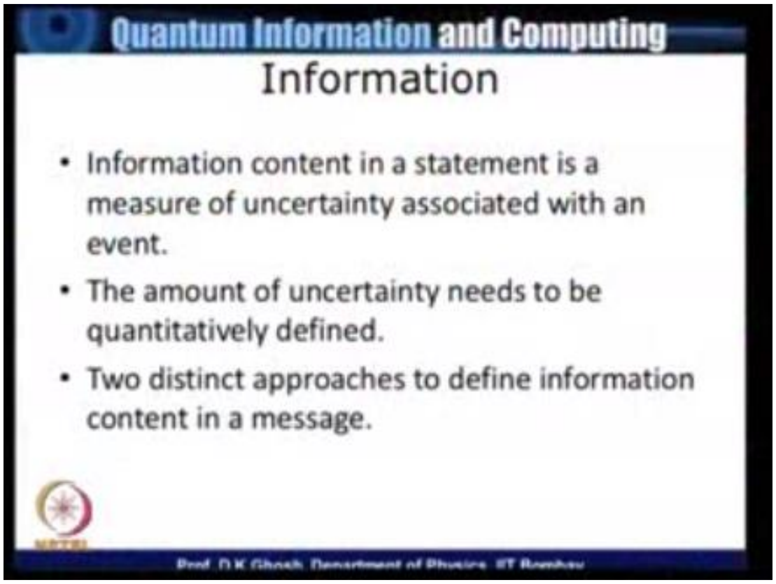


So therefore information in some sense is a measure of uncertainty associated with equipment they notice the difference the, what do I mean by a measure of uncertainty see when you get up definite piece of knowledge it does not remain uncertainty more. So supposing somebody now tells you that yesterday the temperature was actually 21 degree and you had tossed, thrown a die and had got certain number because throwing a die and getting a number between 17 to 22 as we have talked about only give you that particular temperature could be correct with the probability of one six.

But when you get the information of what is the temperate was today your uncertainty is removed. So in others words that I need to some or other quantify this uncertainty and then I had a measure of the information just to give you another example that supposing you find there is a light which is steadily burning on the top of a hill does not give you any information the answer is no, but on the other hand if you see that light is getting switched on and off with a particular frequency that you know may be somebody is trying to communicate something to another person with the cord.

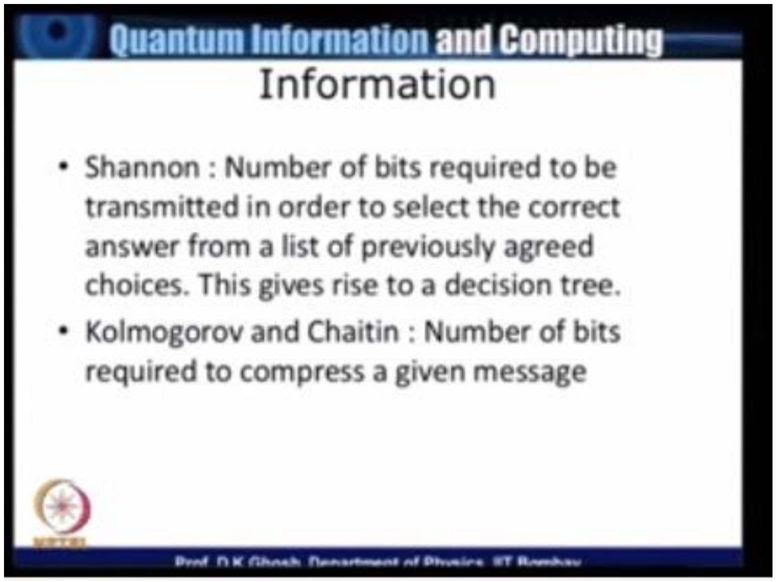
So in other words the uniformity of light was not something which give you an information but the fact that there are from variation of the information gave you some idea about what the, that it could have been a cord.

(Refer Slide Time: 06:23)



Now talking about this the amount of uncertainty which we need to defined quantitatively there are been generally two approaches and we would be actually taking one of them.

(Refer Slide Time: 06:35)

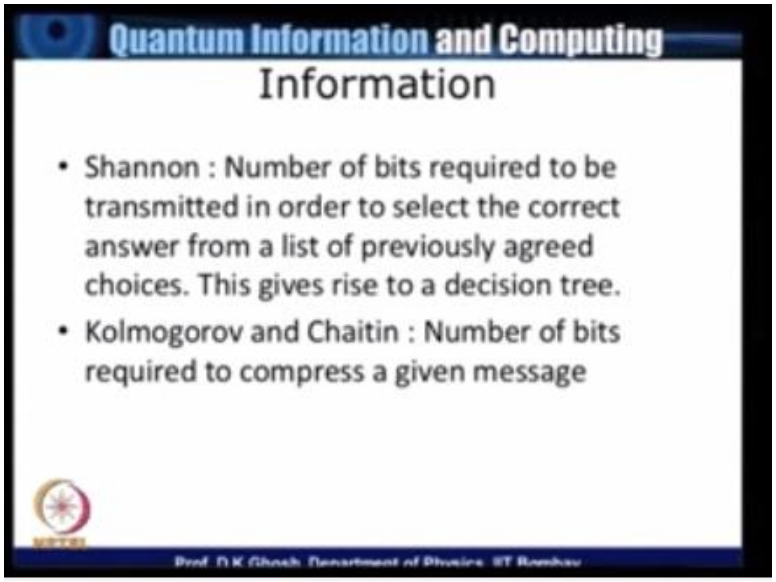


The approach which is due to Shannon it essentially tells you what is the number of bits that needs to be transmitted in order to select the correct answer from a list of previously agreed choices. Now in others what I do is I ask a set of questions and this will give rise to as I can show in an example what is known as decision tree. Now there is an alternative approach to this which is due to Kolmogorov and Chaitin. Now what they look at is the complexity of the message, now just to give you an idea supposing I have a message let us say 200 bit message, now you say 200 bit is fairly strong another long.

But suppose I have that string as 10, 10, 10, 10, etc… up to 200. Now the length of the code that we need to write down in order to communicate this string is a fairly small code that is because we will be simply using the fact that we use a do loop as we say in computer language and say it is 10 we need to divide it is an extremely short cord but all current supposing that bit of 200 string was an absolute random variation of 1’s and 0’s then you need a much longer code. In others words the complexity of the algorithm becomes more if the measured gets more complicated.

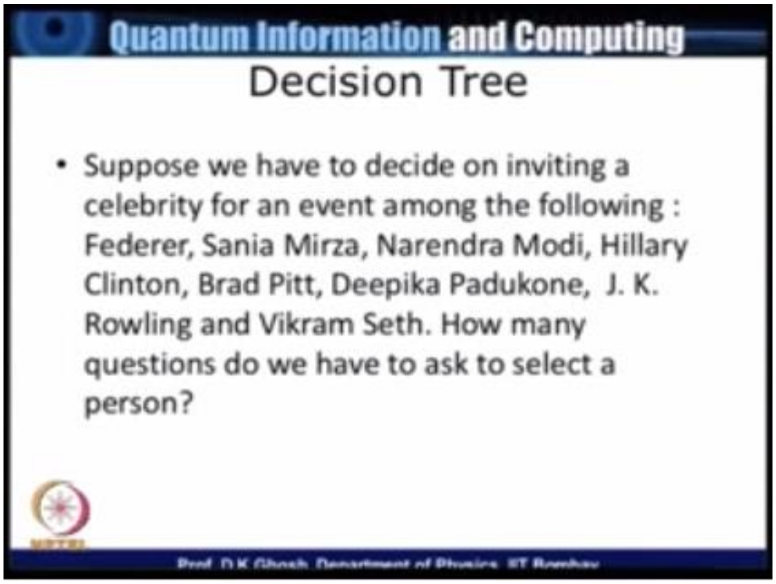
So that is the algorithmic approach due to Kolmogorov and Chaitin which you do more in computer science, but I just thought I should simply mention it.

(Refer Slide Time: 08:30)



Now Shannon thing as I told you is that they have set of choices so I ask a set of questions and your answers should be in yes or no and the way I quantify the information then is how many questions need to be asked in order to arrive to get an answer. Now I will give you a sort of a graph example.

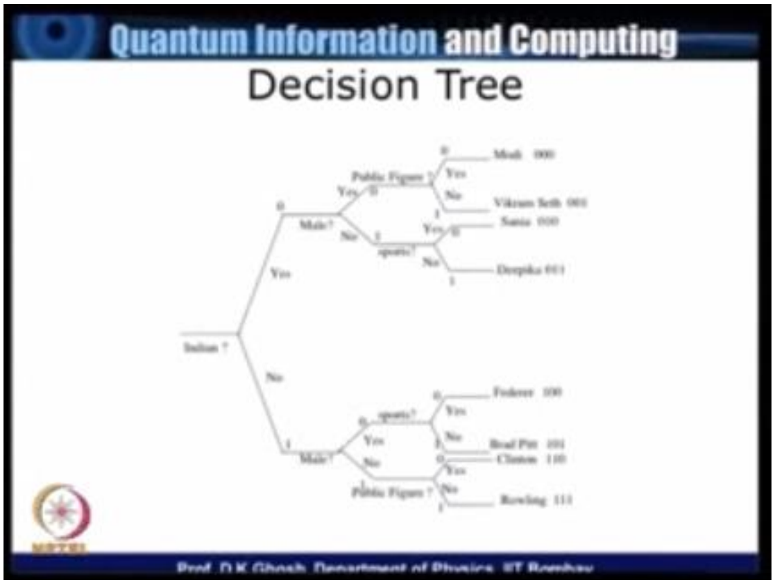
(Refer Slide Time: 08:55)



Supposing we want to decide on inviting a celebrity for an event and our options are these people whose names are very well known and notice that what I have done there is to give you a set of 8 names and not only that this 8 names have certain structure and the structures that are there are 4 of them are Indians and 4 of them are non Indians then I have people belonging to various professions there are people there who are sports people, film people, people in public services and also two of us.

Now what I want to know is that supposing I want to select one person out of this eight by going on asking questions to select a particular question what I get is what is shown here in this slide.

(Refer Slide Time: 09:57)

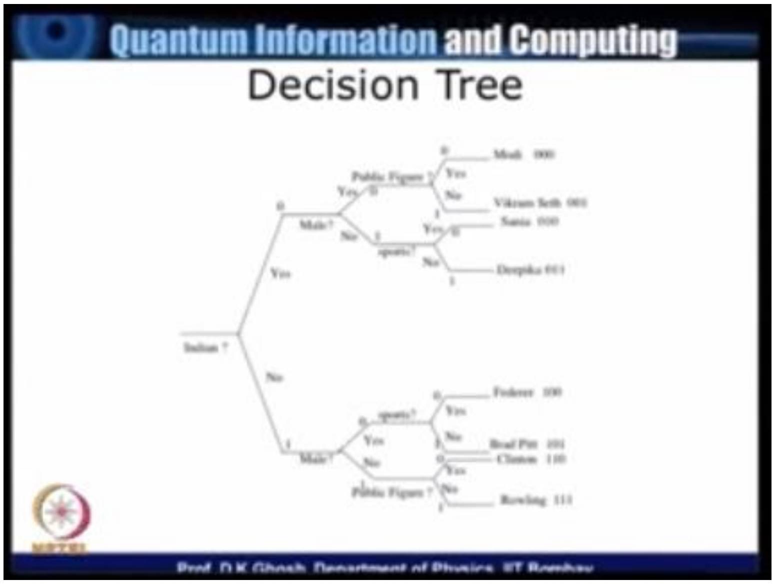


The question that we ask first is the person an Indian now there are two options the answer could be yes or it could be no, alright supposing the answer is yes. Now you see there are 4 Indians in that so therefore the first group with an answer yes it has 4 possibilities in the second group with answer no there are also 4 possibilities. Now I associate with every yes a bit 0 alright so I come to the first bit the left most bit 0.

Now I ask a question is that a male or female if the answer is yes again I pick up a second 0, if the answer is no I pick up a 1 now each one of these groups have two answers. So when I ask the question is that a public figure like the answer is yes then your choice is already known that we are probably talking about Prime Minister Modi and if the answer is no because it was male and they remaining two people are female so the answer has to be author which happens to be the current sates.

Now supposing the answer was no and we do not ask is that a sports personality now you said yes I pick up a third 0 then we would say if it is sports personality and a female so it has to be Sania Mirza and if it is no then it has to be the actress. So you notice now what I did I said I asked a few question now I can count the number of questions and you would look at the number of questions that is there in this case.

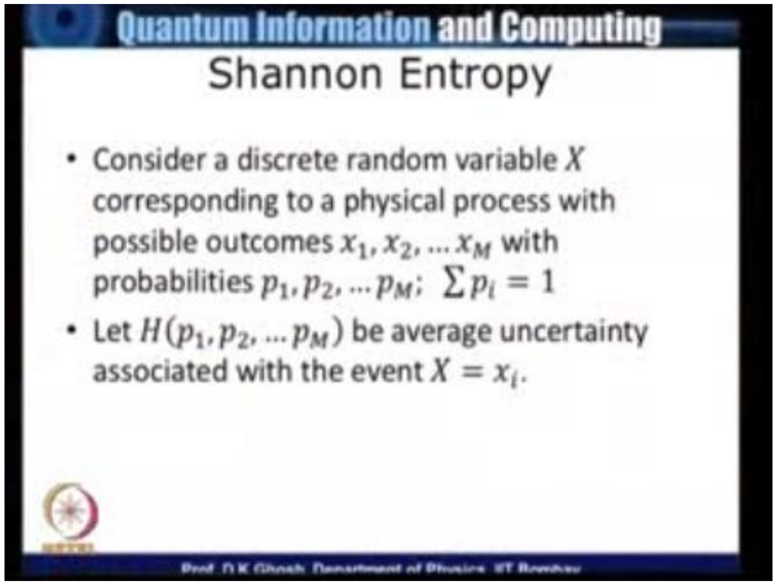
(Refer Slide Time: 11:46)



So you notice here I have question number 1 first with an answer then I asked the second the question which I get a answer then I get a third question which I get an answer. So I need to ask three questions to get answers in this case. And I can associate a bit a unique bit which is for example in the first case Modi it is 0000 for Vikram set it is 0001 etc.…

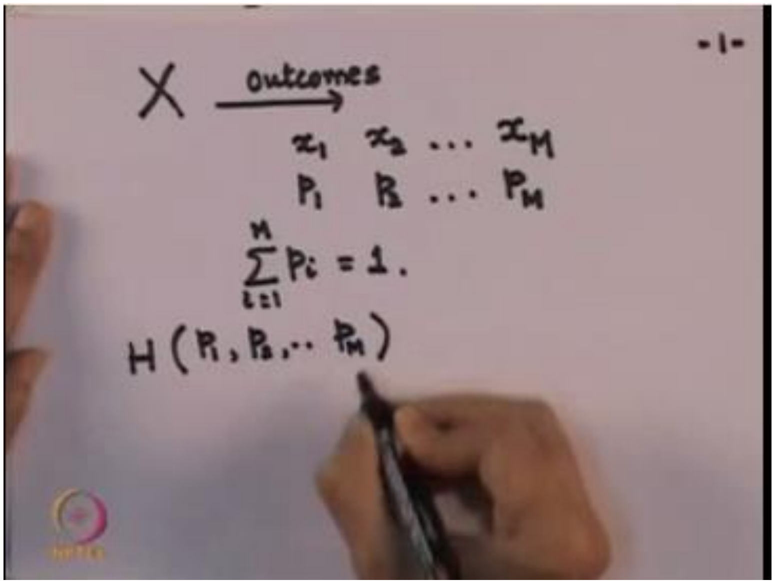
Now let us go to a mathematical quantification so decision tree is one way of quantifying information that is how many bits of communication that I need to convey in order to arrive at a particular answer out of a previously agreed set of alternatives that is one way it is called a decision tree. The second thing is for reasons which will become clear is known as Shannon entropy.

(Refer Slide Time: 12:50)



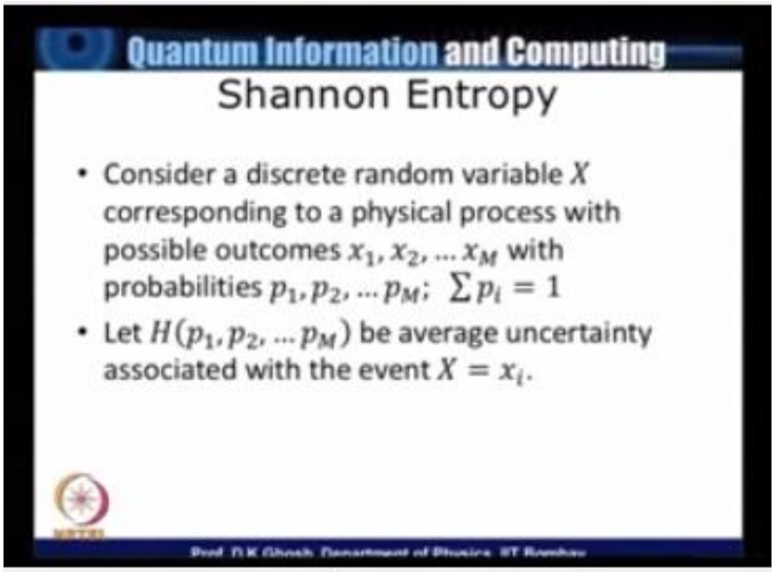
This there is a Shannon's ways of dealing with it and the more common application today in physics is due to Shannon entropy. So we have already identified that information or a measure of information is connected with an uncertainty. So therefore, let us define a random variable supposing I define a random variable x.

(Refer Slide Time: 13:22)



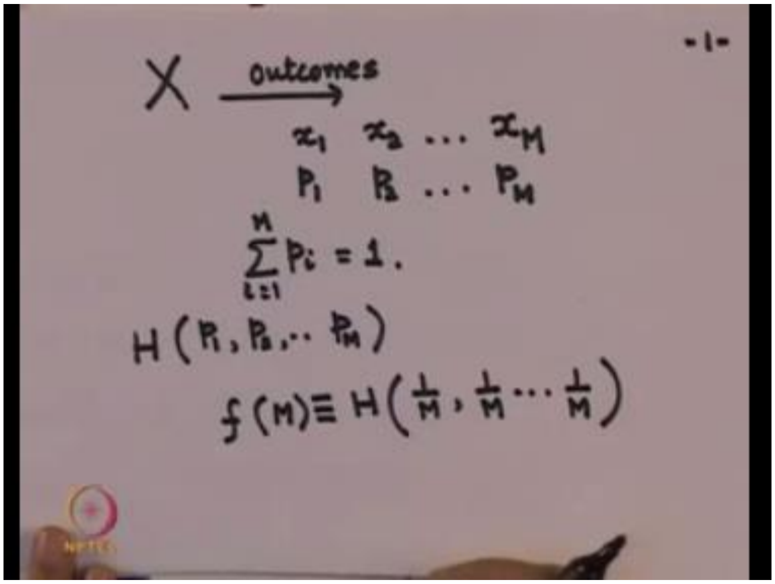
Let me recall for you some standard notations of statistics where the event itself the random variable itself is given by a capital letter and the value is that variable it can take is given by small letters. So we talk about a random variable X corresponding to physical process which have possible outcomes which I will write as x_1 which occurs with the probability P_1, x_2 which occurs with the probability P_2 etc, etc, and let suppose there are m number of events, X_m occurs with a probability P_m. Since the total probability is 1, so I have $\Sigma_i P_{i,}$ I = 1 to m = 1. Now let us define a function, which I will represent by H and say it is a function of the probabilities P_1, P_2 of various events P_M and this is.

(Refer Slide Time: 14:35)



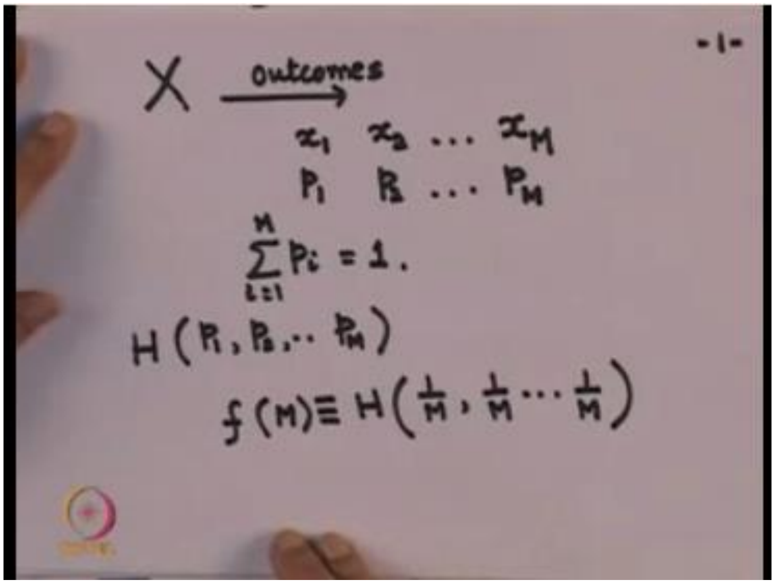
The average uncertainty associated with an event x = x_i. Now you can also interpret it by saying that this is the quantity which represents the amount of uncertainty that is removed when we know the result of an experiment that it is equal to x =x_i, and this obviously the function of all of these. Now I my job will be to have a mathematical expression for this quantity.

(Refer Slide Time: 15:07)



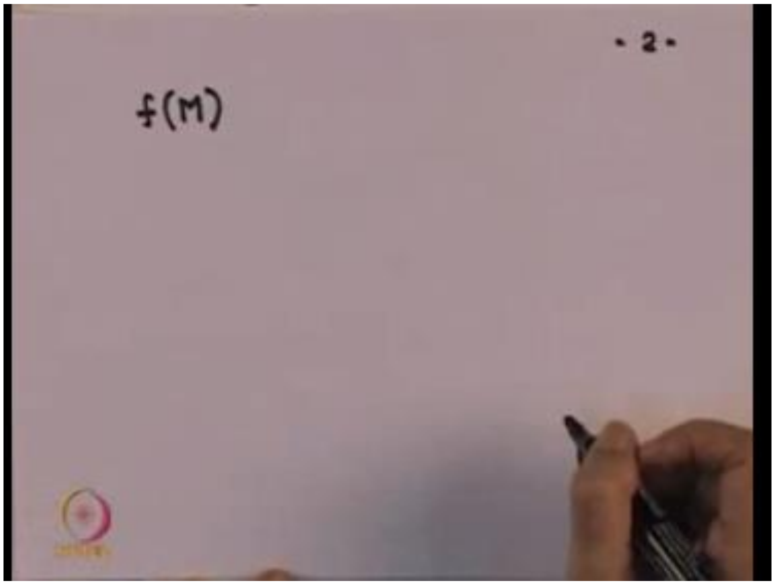
In order to do that let me define certain auxiliary quantum, consider a situation where all my P's are the same, in other words if I have take m events in each one of the P's these 1/M, 1/M, 1/M etc and let me define that by a function f (m) = H, so it is actually a definition 1/M, 1/M etc M terms what does it actually mean? It means that suppose you are doing a coin toss, now I know that in a fare coin toss my I have only Q events P1 = ½ P2 = ½.

(Refer Slide Time: 16:00)



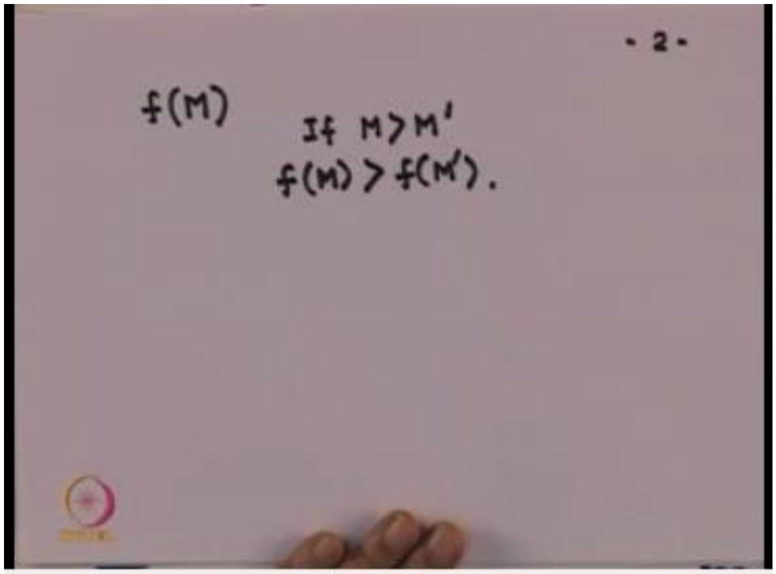
So my f (2) is H (1/2, 1.2) so that is what you meant by this function. Now the other thing which becomes obvious for the property of this function f is that.

(Refer Slide Time: 16:16)



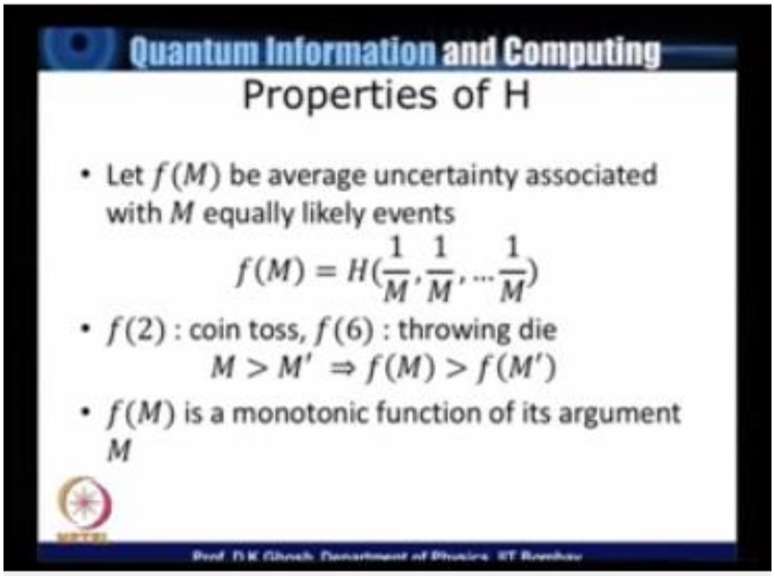
F (m) must be a monotonically increasing function with increasing value of M, now look at what it means see supposing I am talking now about if I am saying.

(Refer Slide Time: 16:31)



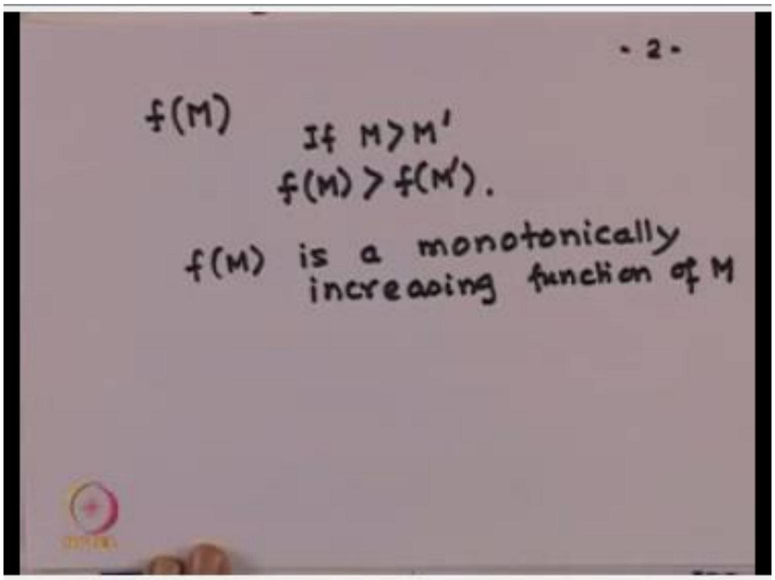
If M>M' then f (M) > f (M') now that is very obvious.

(Refer Slide Time: 16:45)



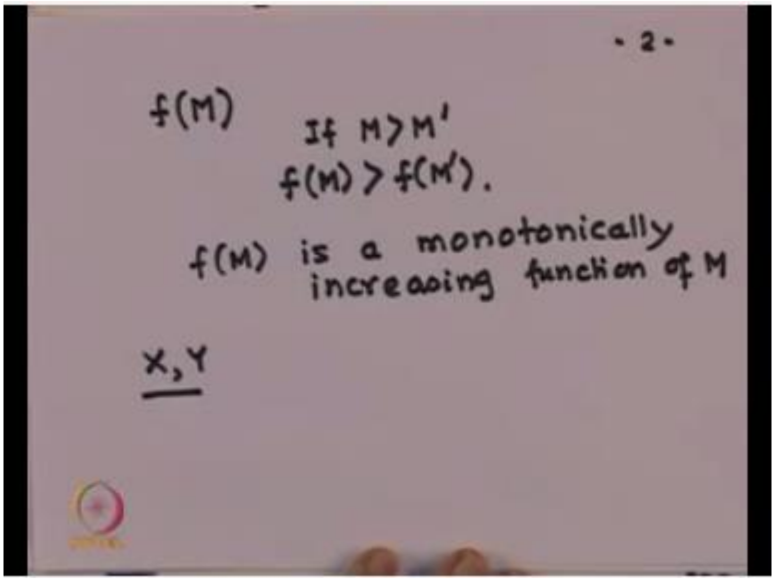
Because take the example of a coin toss, now I have only two alternatives. So if I have two alternatives by corresponding f (M) is ½, ½ and supposing I am talking about throwing of a die then obviously I am talking about 1/6, 1/6, 1/6, in other words f(M).

(Refer Slide Time: 17:09)



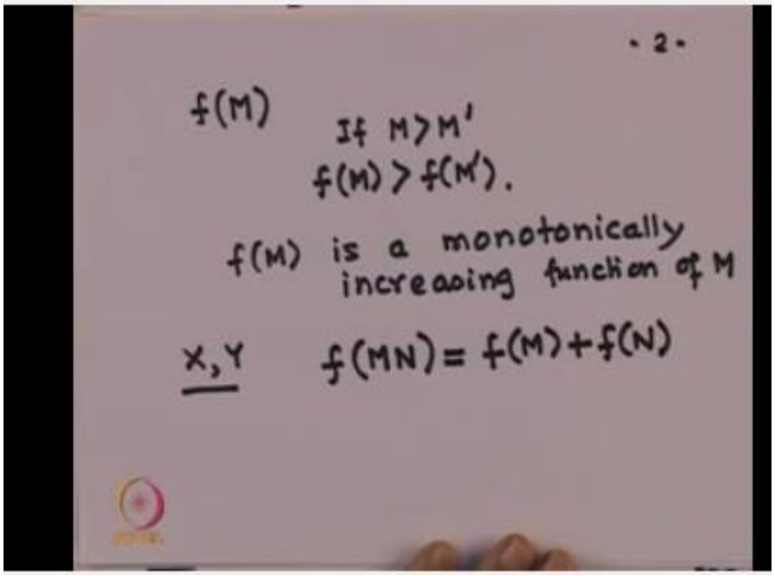
Is a monotonically increasing function, f (M) is a monotonically increasing function of M. Some other interesting f(m). For instance, suppose I consider two random variables one I call as x another I call as y.

(Refer Slide Time: 17:42)



This as to be specific, let us suppose x is a coin toss which has two alternatives and y is showing of a die which is 6 alternatives and supposing I want to do a joint experiment, what does a joint experiment mean? It means I throw a coin and I throw a die and I get various alternatives. So for example, I could get a head with any of the 1, 2, 3, 4, 5, 6 or a tail with any of the 1, 2, 3, 4, 5, 6, so my number of alternatives have gone up. The possibilities then that is there is, that this must have this property.

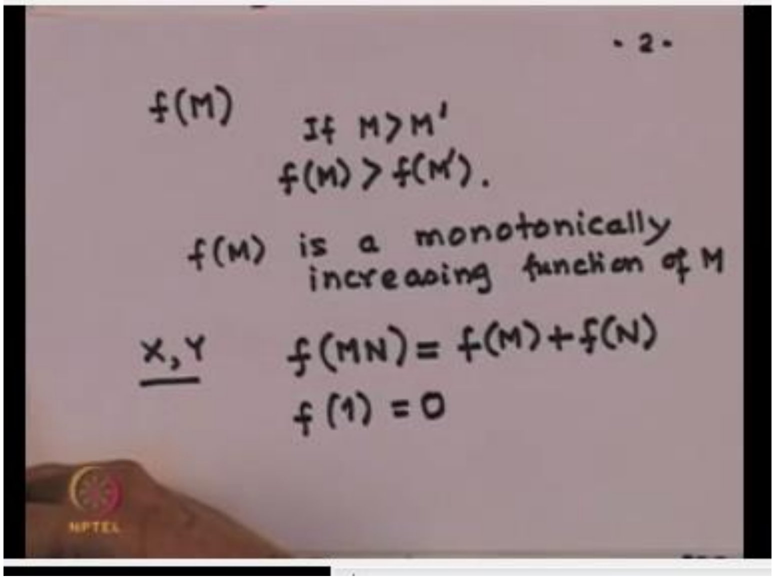
(Refer Slide Time: 18:28)



And f (MN) = f (M) + f (N) now this is actually obvious, because when you do a joint experiment the number of alternatives go on, M times N becomes the number of events, but if you know the result of let us say the first group which had M events that does not help you remove the uncertainty associated with a second events which is given by f(N).

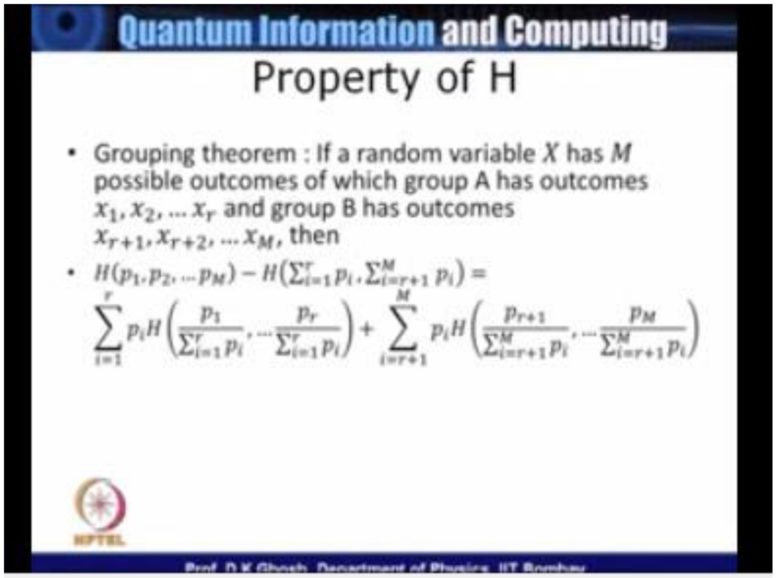
So if I know the result of a coin toss I still have an answer to the function defined by a subset, on the other hand if I know the result of throwing of a die whether it is 1, 2, 3, 4 and 5 or 6 the uncertainty with respect to what happened for the coin toss whether it was a head or tail statement so therefore I must have f (MN) = f(M) + f(N).

(Refer Slide Time: 19:38)



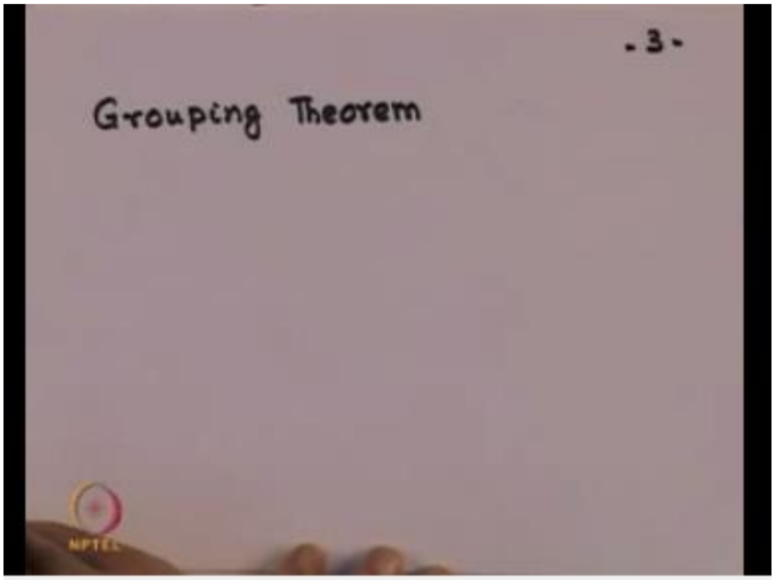
The another obvious property is f(1) = 0 in there just 1 event there is no uncertainty associated it is called a certainty, so since there is no uncertainty f(1) = 0.

(Refer Slide Time: 19:54)



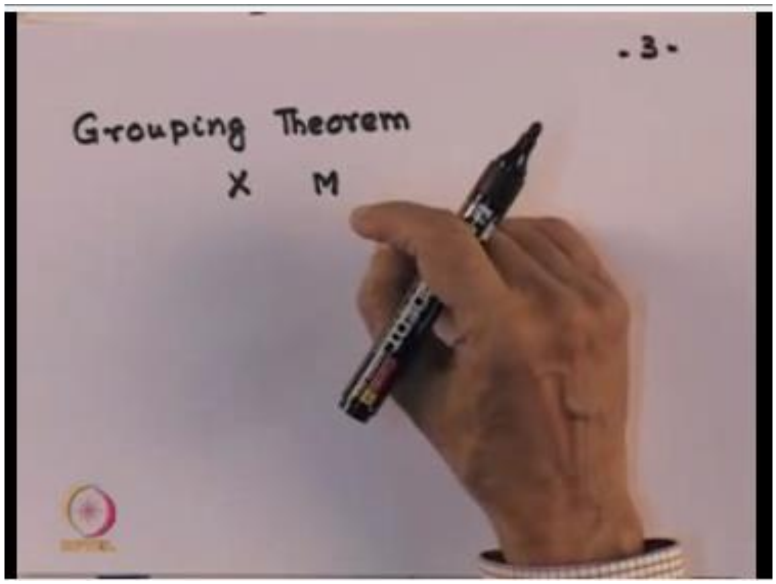
There is another property which is important it takes a bit of a time to work out since I am not going to be giving a course on classical information theory I will not work it out but show it to you some with some examples of how it works and this is what is known as a grouping theorem.

(Refer Slide Time: 20:13)



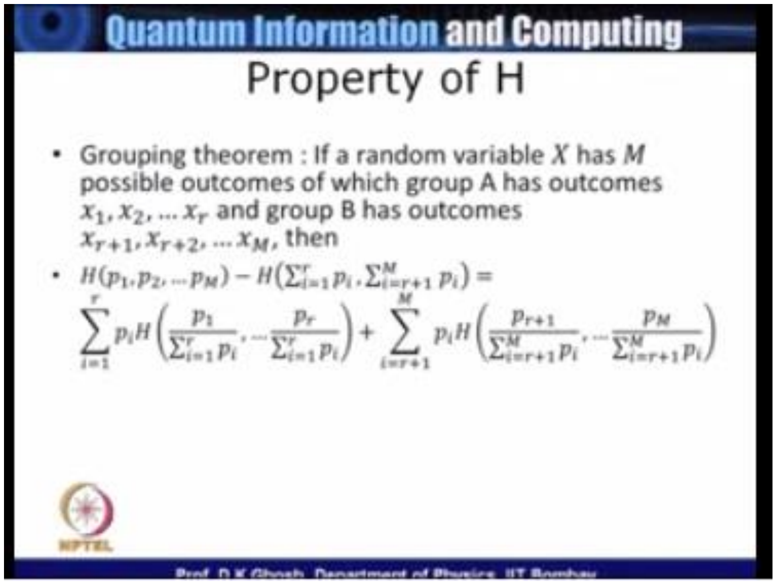
The proof is, on in any classical information theory book but basically a grouping theorem talks about, suppose I have a random variable x.

(Refer Slide Time: 20:31)



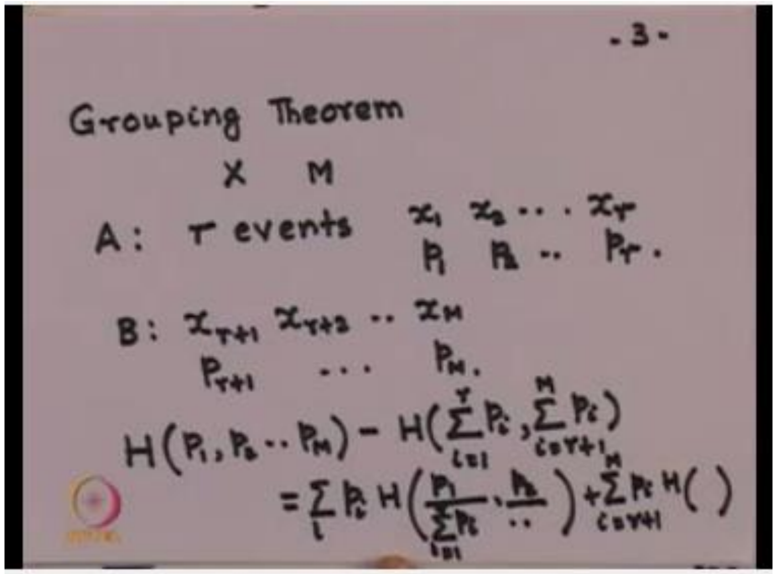
With M different possibilities and whatever we associated probabilities called P1, P2 and P3.

(Refer Slide Time: 20:42)



But let us suppose I divide them into two groups let us call it group A.

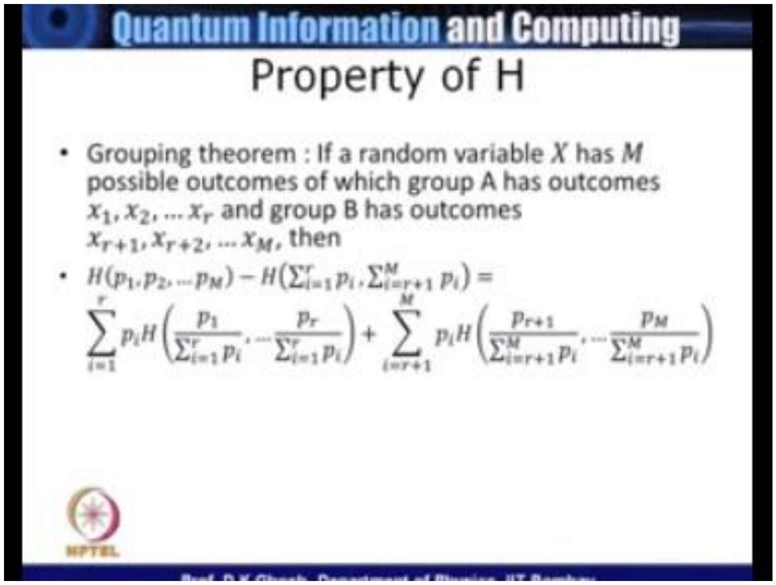
(Refer Slide Time: 20:48)



And group A which have R events and this R these R events are x_1, x_2 ….x_r with probabilities of course P1, P2….Pr and group B is well essentially the remaining events which is x_{r+1} x_{r+2} …… x_m and with corresponding probabilities these are P_{r+1} etc ……P_M. Now what is the grouping theorem does is to say, that this uncertainty which we have said is given by P1, P2…PM = H as a function of the first group probability.

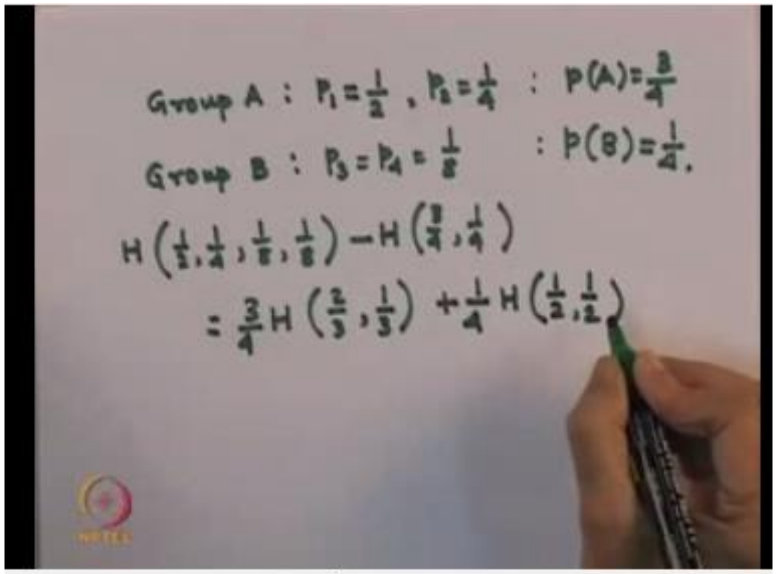
Which means it is i =1 to r of Pi and second group probability I = r+1 to m Pi and that happens to be equal to Σ over I Pi then the H associated with these events now you notice what is this, so I have P1 by now the normalization because I know it is event in group A is Σ over i = 1 to r Pi etc. And a similar term Σ over i = r+1 to m, Pi and similar thing but now is will be in the slide because I do not have writing space here, but now the denominator is from i=r+1 to m look at the slide for the exact structure of the equation..

(Refer Slide Time: 22:49)



This equation is derived by using what is known as base 0 in statistics and as I said take a bit of a time so I will probably put it on the lecture note associated with this set of lectures but proof I will not do it here, but just to illustrate what the grouping theorem meant? So suppose I consider two groups.

(Refer Slide Time: 23:23)



Now I have the group A in which I have two events probability of first event is ½ and the second event is ¼ let us say. So the total group A's probability so let me write it as a P(A) = ¾ there is a group B in which I again have two events with let us say equal probability so P3 = P4 = 1/8 so that P(B) = ¼ so if I look at the total uncertainty for the whole event what the grouping theorem tells me is, H.

So let me first write down all of them so I got ½, ¼, 1/8, 1/8 – H(3/4, ¼) now that would be equal to choose the event A which is ¾ and now I need the probability within a group so you notice that within this group given that the group is A this probability of P1 probability corresponding to P1 would be 2/3rd and this would be 1/3rd because you can see this is half that of this one so total probability within a group B = 1.

The first one has probability 2/3rd the second one has a probability 1/3rd and likewise for the second group whose probability is 1/4th times the probability within a group of the two events there but since these are equal so each must have the probability ½ and ½. Now what I am going to do in the next lecture, is to look at these properties and find out is there a function that I note which satisfy these things, what are these things? F(1)= 0 is a monotonically increasing function of its argument f(MN) = f(M) + f(N) and finally a statement of grouping theorem events lie, we will see next that this is given by a very well known function and I will identify that with what I will designate as entropy or sign of entropy.

NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING (NPTEL)

NPTEL
Principal Investigator
IIT Bombay

Prof. R.K. Shevgaonkar

Head CDEEP

Prof. V.M. Gadre

Producer

Arun kalwankar

Online Editor
& Digital Video Editor

Tushar Deshpande

Digital Video Cameraman
& Graphic Designer

Amin B Shaikh

Jr. Technical Assistant

Vijay Kedare

Teaching Assistants

Pratik Sathe
Bhargav Sri Venkatesh M.

Sr. Web Designer

Bharati Sakpal

Research Assistant

Riya Surange

Sr. Web Designer

Bharati M. Sarang

Web Designer

Nisha Thakur

Project Attendant

Ravi Paswan
Vinayak Raut

NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING (NPTEL)