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

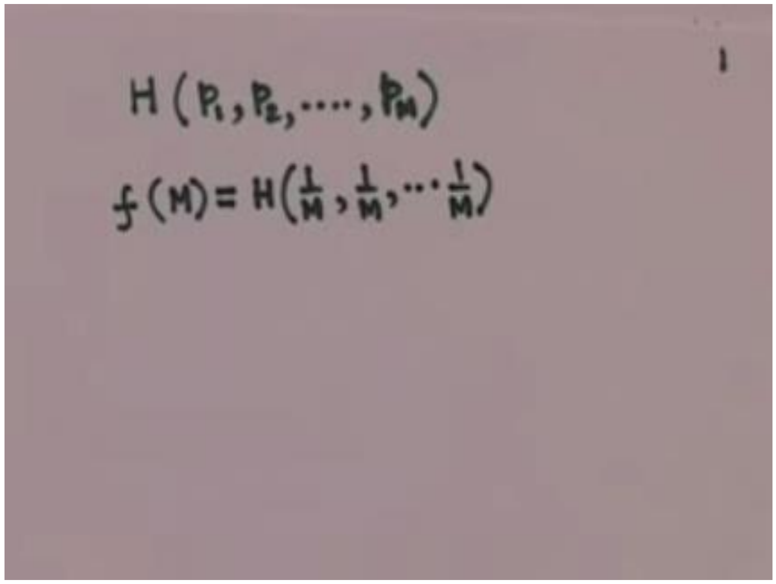
Lecture No.35

Shannon Entropy

In the last lecture we had started talking about the information and we tried to understand what does classical information mean and what does the phrase information actually communicates to us. Now we what we conclude it from that lecture is that the word information at least in the context of physics and technology indicates not what we understand normally qualitatively but it indicate a quantity which is a measure of uncertainty associated with a new and when out of various possibilities of an event a particular event occurs the amount of uncertainty that gets removed that is a measure of the information.

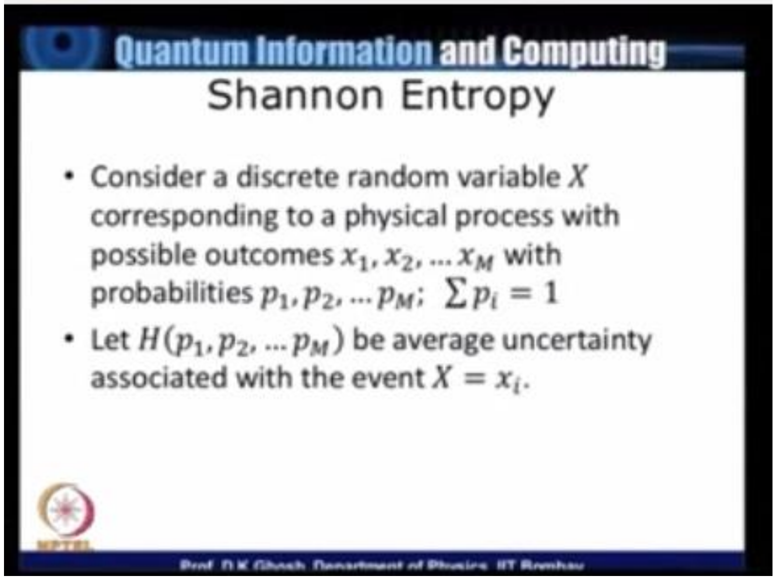
So information was defined in some way in a negative sense about the amount of uncertainty associated with occurrence of the event. We define a function to make a as a measure of such information.

(Refer Slide Time: 01:37)



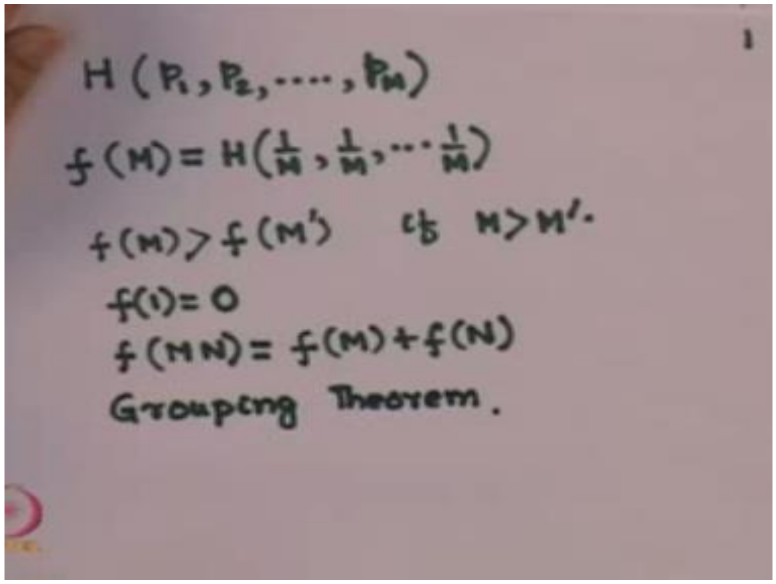
So what we said is that H m events with Probabilities P_1, P_2 ….P_N so we wanted to get an expression for this quantity and we said that this quantity has certain properties. Now what we did is to first defined a new function called f which was simply nothing but H with each one of the probability is the same so that is equal to supposing there are N events so f(M) is defined to be H (1/M, 1/ M) etc now we said that.

(Refer Slide Time: 02:32)



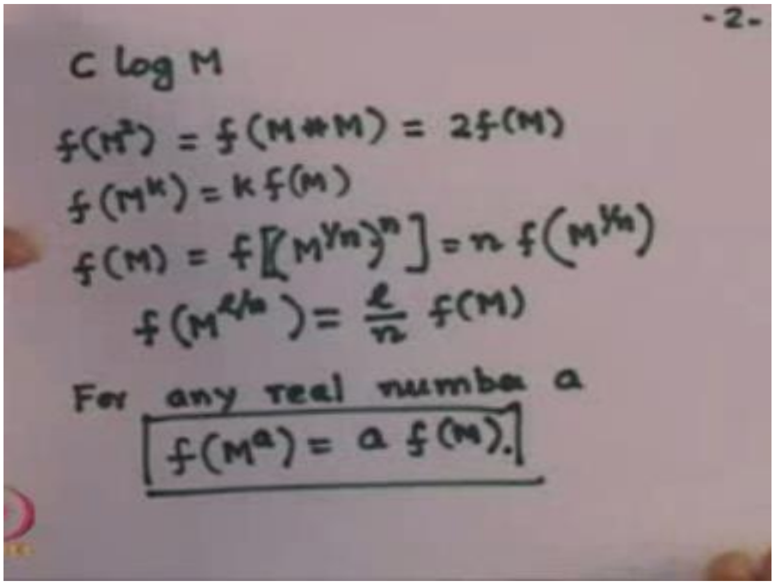
This function P(M) must satisfy certain algebraic property and they were that it is an monotonically increasing function of it is argument so f(M) is a monotonically increasing function.

(Refer Slide Time: 02:50)



So f(M) > F(M)´ if M > M´ the other properties where that it is value for f(1)= 0 this arose because in there is just 1 event there is no uncertainly associated with it. The third property was f(M N) if I am talking about a joint experiment then this must be equal to f(M)+ f(N) and the fourth one which had a longer expression was called a grouping theorem. Now we claim that a function which is constant times logarithm of N.

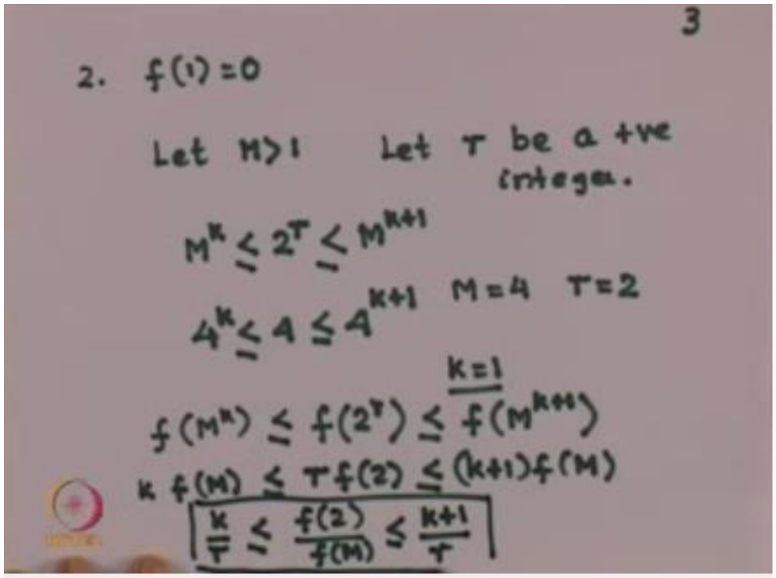
(Refer Slide Time: 04:00)



Base of logarithm is unimportant c times c is a constant times logarithm of M satisfies these 4 properties that I have talked about. Let us examine them one by one firstly you notice $f(M^2)$ I will take c to be a positive constant $f(M^2)$ is by definition f(M x M) and since f (MN) is equal to f(M) +f(M) so this is f(M) + f(M) which means it is equal to 2 f(M) and you can do an iteration and find in general that $f(M^k)$ = k f(M) not only that I can write f(M) as equal to $f[(M^{1/n})^n]$ is just an identity actually so that must be equal to n times $f(M^{1/n})$.

If I rare it to the power L then I can show that $f(M^{l/n})$ = L/n times f(M) that is because $f(M^{1/n})$ is 1/n times f(M) and you raise books they argument to the power and since we have said this would generally true no matter what powers we take so we say for any real number A I get $f(M^a)$ =a times f(M) now notice that this function c log M obviously satisfies because the left hand side would be c log (M^a) and as you know logarithm as this property that log (x^a) is a log (x). So therefore this is satisfied by this log the second point is we wanted f(1) = 0.

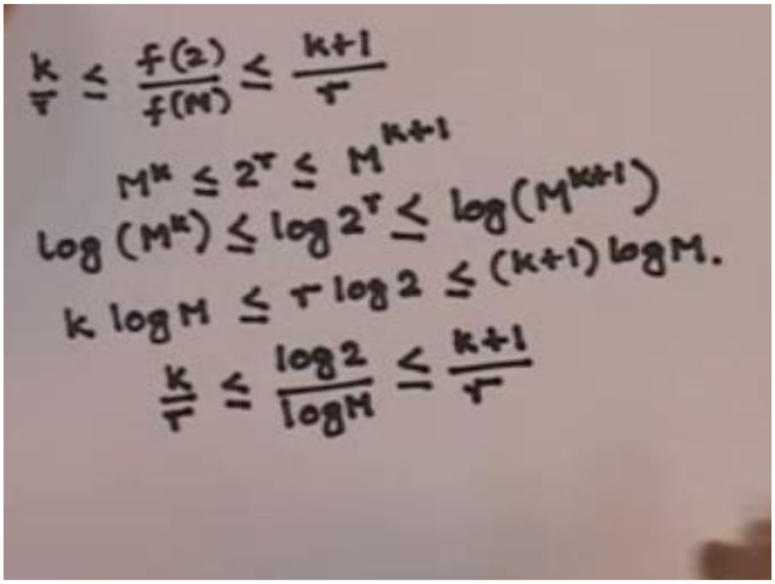
(Refer Slide Time: 06:46)



This is trivially satisfied by the logarithmic formula because log of 1= 0. Now let us see why this is a good function and the way it works as the following let M> 1 and let r be a positive integer now for any M I can find for any M > 1 I can find an r which satisfies this relationship $M^K \leq 2^T \leq M^{K+1}$ I am I claim that for given M and r I can always find k which satisfies unique k which satisfies this unique M.

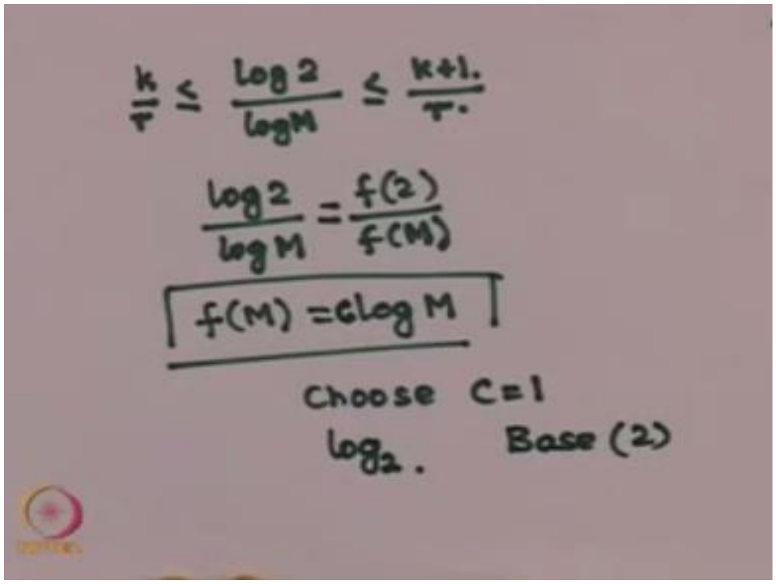
Now let us look at what it means by realistic supposing M just happens to be equal to 4 and I have taken r = 2 so I am looking for a number k which is $4^K \leq 2^2 \leq 4^{K+1}$ you notice that there is a unique k it tells you k= 1 because $4 \leq 4 \leq 4^2$. Now since M is a monotonic function it tell me that $f(M^k)$ remember I said if M> that M´ than f(M) > f(M)´ and once I have this identity then $f(M^k) \leq f(2^r) \leq f(M^{K+1})$ but by property of f these identities can be written as k f(M) because $f(M^k)$ k times f(M) ≤ r f(2) ≤ k+1 f(M) dividing appropriately I get k/r ≤ f(2)/f(M) ≤ k+1/ r now note one thing so we had this relationship.

(Refer Slide Time: 09:59)



$k/r \leq f(2)/f(M) \leq k+1/r$ this arrow is because of the fact f(M) is a monotonic function. Now I have a very similar satiation with respectable logarithmic function and that is primarily because I have $M^K \leq 2^r \leq M^{k+1}$ and since logarithmic function is also a monotonically increasing function I can write $\log (M^k) \leq \log 2^r \leq M^{k+1})$. So since $\log (M^k) \leq$ k times log(M) so I can write $k \log M \leq r \log 2 \leq (k+1) \log M$ thus even for the logarithm function I have this relationship $k/r \leq \log 2/\log M$ which is $\leq k+1/r$ both f(2)/ f(M).

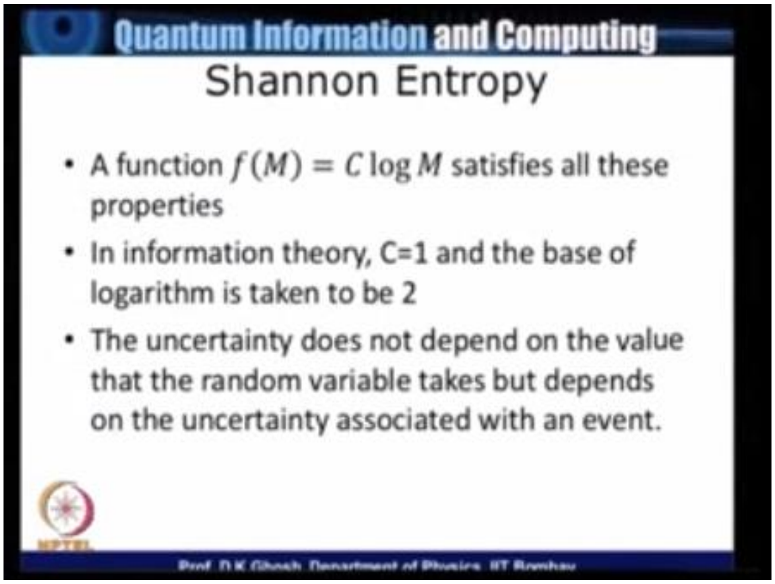
(Refer Slide Time: 11:44)



And log2/ log M they lie between these two limits and I have not said what is r all that I have said is let r be a possibility, so I can take r as large as I can or as large as I wish and then show that these two functions log 2/ log M and f(2)/ f(M) are basically identity which tells me that f(M) is logarithm of M well actually this equation simply shows f(M) is equal to c times logarithm M where c is constant. It is traditional in information theorem to choose c =1 and take the base of the logarithm to equal to 2.

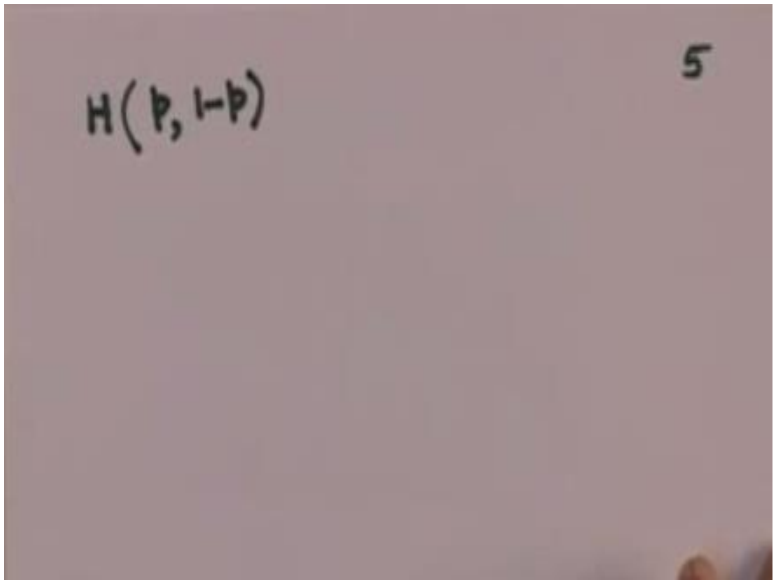
So I have used properties of logarithm in a way that what is the base it did match so therefore I am free to use it and because of the fact we deals with bits in the computing it is traditional to take the base of the logarithm to be equal to 2 so with this.

(Refer Slide Time: 13:17)



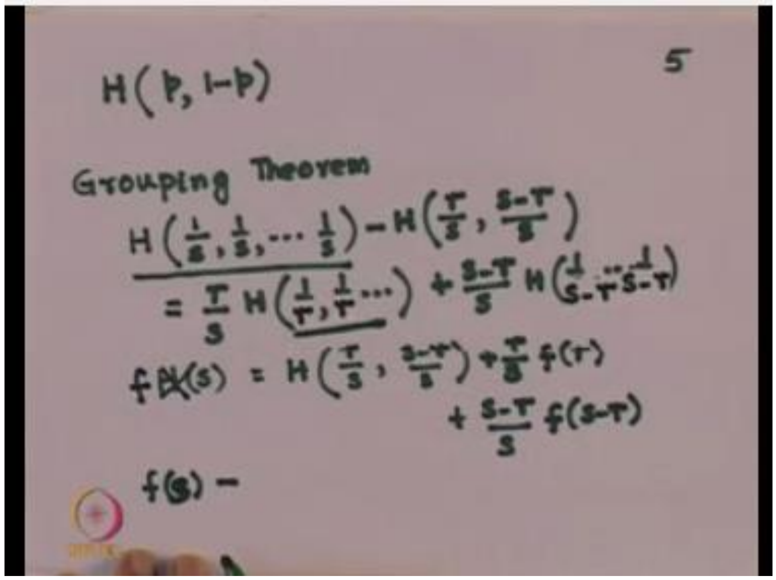
Let us look at the slide. So we have now said that the uncertain that wont actually depend upon the value that the random variable is but depends up on the uncertainty associated with it. Now what we do is the following remember our idea is to primarily find an expression for the uncertainty function H what we have done is to get a relationship for the f(M) function which is the same function as H but where the probabilities of possible events are all equal, and we have seen the logarithm is a good function for that. Now what we do is the following that suppose I consider just 2 just to make it simple.

(Refer Slide Time: 14:16)



So my function information function uncertainty function will be H (P 1-P) by definition. Now this quantity one do you recall our definition of the grouping theorem which I will show in the slide here.

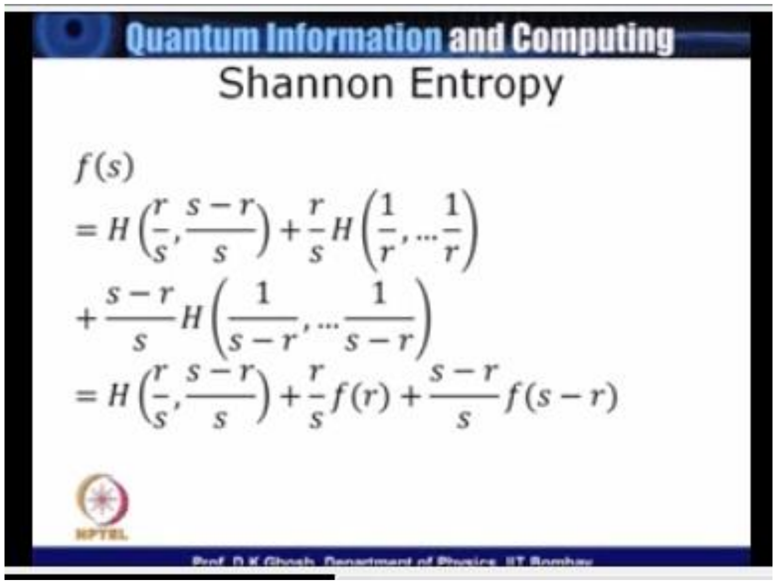
(Refer Slide Time: 14:34)



So if you view as the grouping theorem and just two events one with probability P and one with probability 1-P. So we say that H(P1-P) how do you write using grouping theorem. Now the grouping theorem it is shows like this supposing I have got S equally likely events, so we had seen that this is (1/s, 1/s…1/s) – there are two groups of events first group has r events so it is r/s, second group obviously as s-r events so it is s-r/s, and this quantity was shown to be equal to r/s times H(1/s, 1/s ….) + s-r/s H(1/s, 1/s). Now let us look at what these are, in terms of F this has you remember there are S number of events.

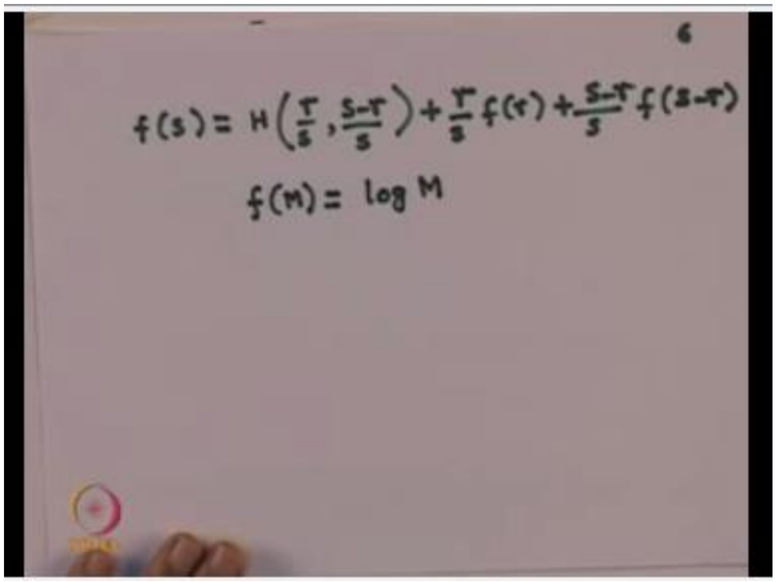
So therefore this is nothing but H(s) this is well I will take this term to the other side equal to H of I keep it understand r/s, s-r/s + r/s. Now these has r number of terms so therefore this can be written as f(r) because remember the argument of f is the number of events that are there having equal to r and the other one is s-r/s f(s-r). Rearrange them to write down what if f (r) or f(s) is this case. Because I have got f(s) there, this is actually f(S) not H.

(Refer Slide Time: 17:13)



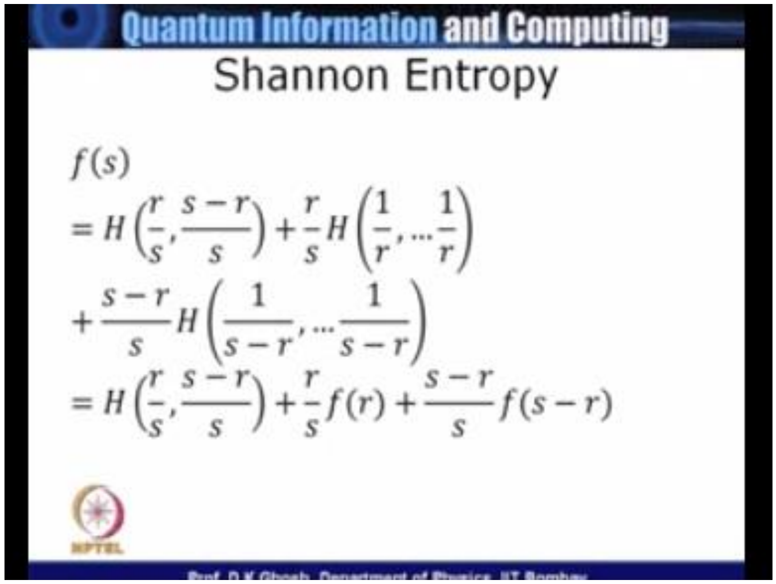
So my f(s).

(Refer Slide Time: 17:22)



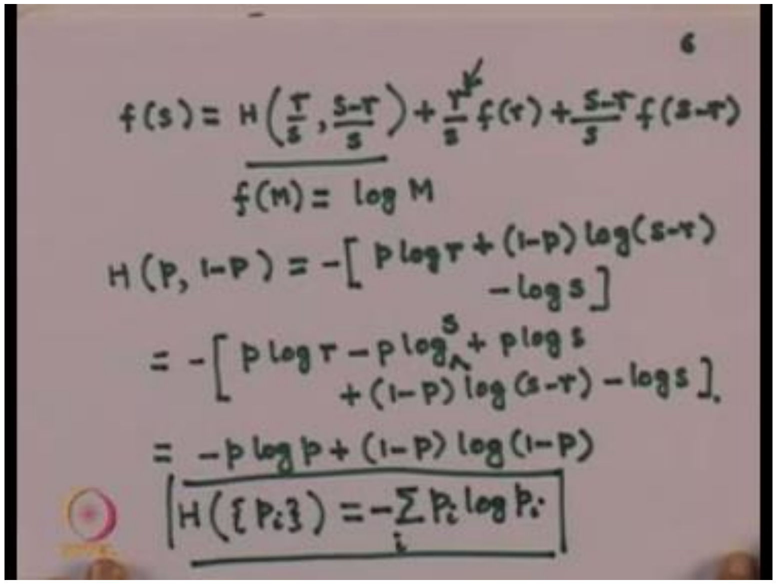
Is H(r/s, s-r/s) and you have to re-writing it r/s f(r) + s-r/s f(s-r), now let us substitute our expression that f(M) = log M and let us suppose that I have essentially two groups.

(Refer Slide Time: 18:02)



So what I get here is this.

(Refer Slide Time: 18:06)



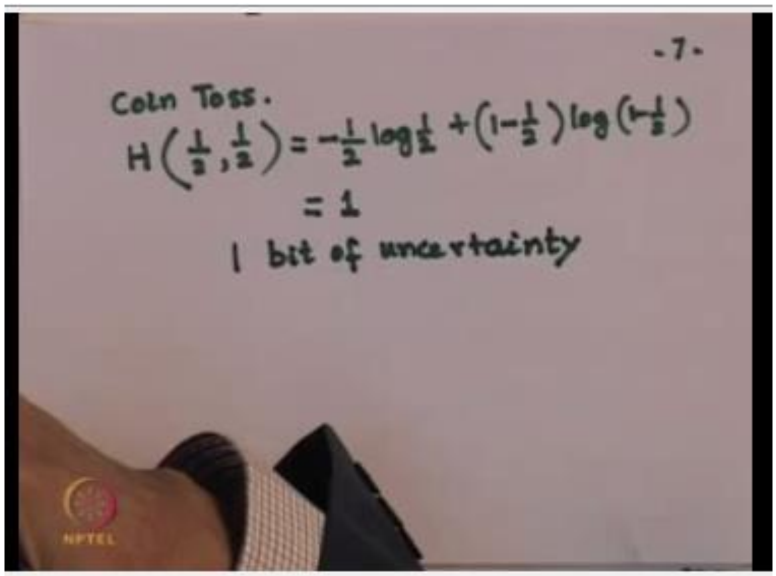
That this is because r/s supposing I have just two things so r/s is nothing but P and this is 1-P because I have just two groups there. So I have said f(s) is logarithm so if I do that I get this term f(P, 1-p) = minus because everything is minus now this is P and f(r) is log (r) so I get P log r + this is (1-p) log (s-r) and of course the term which is there on this side since I would take in overall minus sign is – log s.

So this one can obtain purely by application of grouping theorem that I have a group A having probability P and a group B having probability 1. So this I can do a bit of an algebra and get the following, so this is – [p log r – let us just add P log s and subtract P log S and I have the remaining term 1-P log (s-r) so these two terms and then of course I have got –log s. So look at this thing, this term and the last term.

I will write as P log (r/s) but r/s if you recall its P, so therefore I get –P log P + this is P log s because I added and subtracted. So then I get the remaining terms which are there is remember that I can combine these terms that I have a these two I have already taking take care of these two I can add and write it as (1- p log s) so I have got (1-P) log s-r/s, so this is equal to (1-P) log (1-P). So this is my expression for H[P(1-p)].

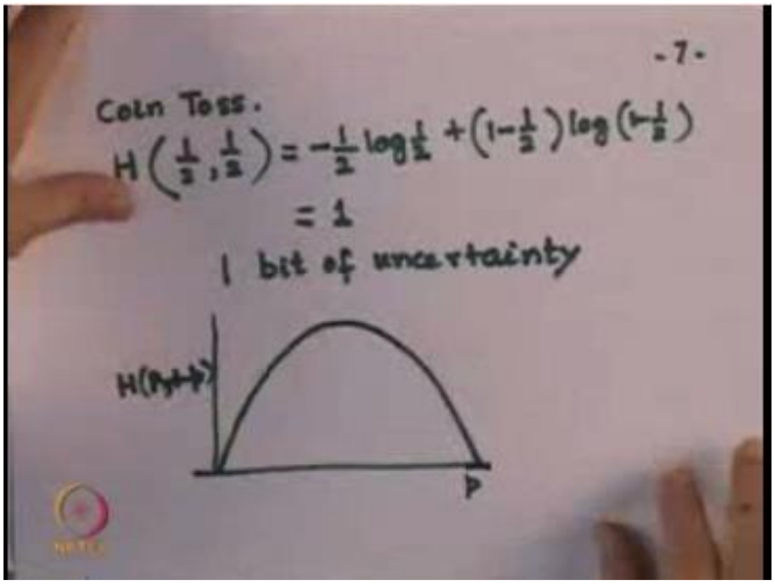
Now if I had more events then it is readily identified we say H supposing I am talking about a collection of probabilities Pi so this is simply equal to $-\Sigma_i P_i \log P_i$. So this is my measure of the information uncertainty. Look at some simple ideas supposing I am talking about a coin toss.

(Refer Slide Time: 22:00)



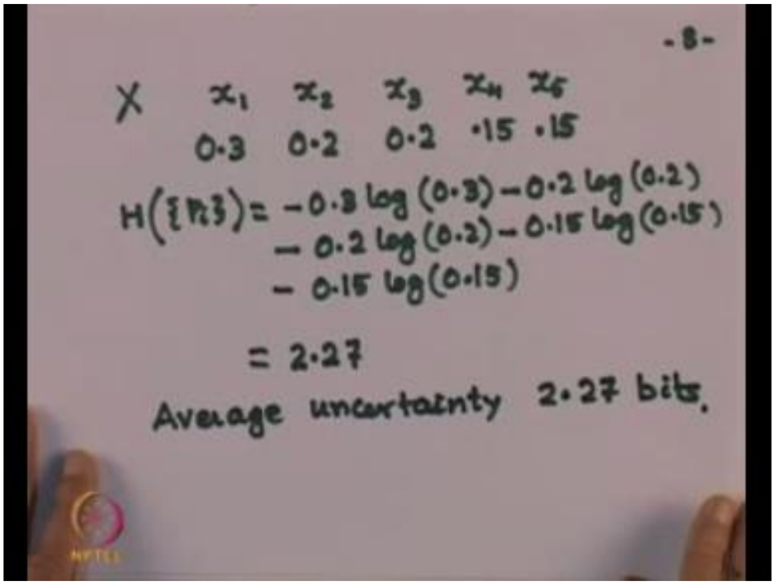
So coin toss is obviously given by H(1/2, ½) so which is equal to – ½ log ½ + (1-1/2) log (1-1/2) and you can immediately see since I have said that the base of logarithmic is 2, so this is minus goes away I get ½ log 2 which is 1 and sorry which is ½ and this is also another ½ log 2 so add it together I get 1. So we say associated uncertainly with a coin toss as 1 bit of uncertainty, 1 bit of uncertainty and thus it is clear because a single bit takes the value 0 as 1, so if you say head is 0 and tail is 1 then of course there is one but on uncertainty associated with it.

(Refer Slide Time: 23:16)



If you plot H(p, 1-p) against p two events the type of curve that you will get will be like this. Obviously once P exceeds ½ it must have a symmetric because the other event then has have probability. Just to tell you, what is the connection of this with the decision tree issue that we talking.

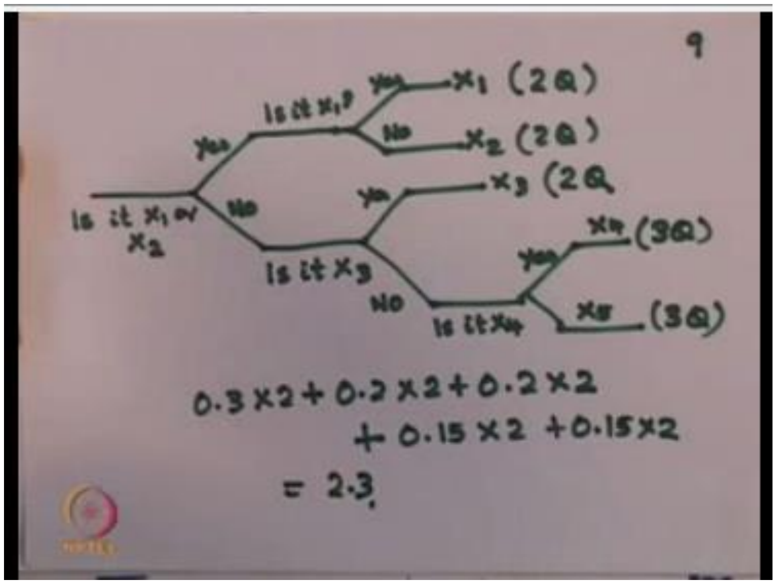
(Refer Slide Time: 23:49)



Let me talk about an event described by a random variable which has let us say following five relation possibility x1, x2, x3 and x4 and x5, this has probability 0.3, this has 0.2 this is 0.2, 0,15 and 0,15 total add is upto 1 of course. Now if you go by the definition of H P_i I can easily calculate it with a calculator, so I get this is -0.3 log of course to the base of 2(0.3) – 0.2 log (0.2) once more the same thing is appearing for the third event –(0.15) log (0.15) and –(0.15) log 0.15.

You could just use a calculator and work it is out that it gives you 2.27, so it tells you that the average uncertainty associated with this event is 2.27 bits average uncertainty is 2.27 bits, now let us examine the same thing on our question answer way that is now I have various events which I have been listed here x1, x2, x3, x4, x5 and I have given the probability. So let us try to do n decision tree.

(Refer Slide Time: 25:52)



The first question I ask is, is it x1 or x2? I can have an answer yes, I can have an answer no, if the answer is no it simply means the group belongs to x3, x4, x5 if the answer is yes then it means the answer is either x1 or x2. So again I ask a question is it x1? Now if the answer is yes I say the result is x1, if the answer is no I say the result is x2. Now look at this I had two questions to get at either x1 or x2 so let me write two questions.

If the answer is no I have x3, x4 and x5 I ask the question is it x3? Now if the answer is yes I get immediately x3 to the answer and obviously I have taken only 2 questions, if the answer is no I still have two automatics so I ask is it x4? If the answer is yes the answer is yes it is x4 so I need 3 question for x4, if the answer is no also I need 3 questions for x5, what is the average number of questions?

So this is 0.3 x 2 + 0.2 x 2 + 0.2 x 2 + these probabilities was 0.15, 0.15 x 2 + 0.15 x 2 and this works out to 2.3. So you notice the uncertainty here is greater than what is calculated using the uncertainty function or Shannon entropy as we call it. The Shannon entropy primarily gives me the minimum bits of uncertainties that any code that will have the optimal value of a communication code.

Now this is applicable to what is calling unique desirable code and that why it is called Shannon trophy we will see in the next lecture.

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)