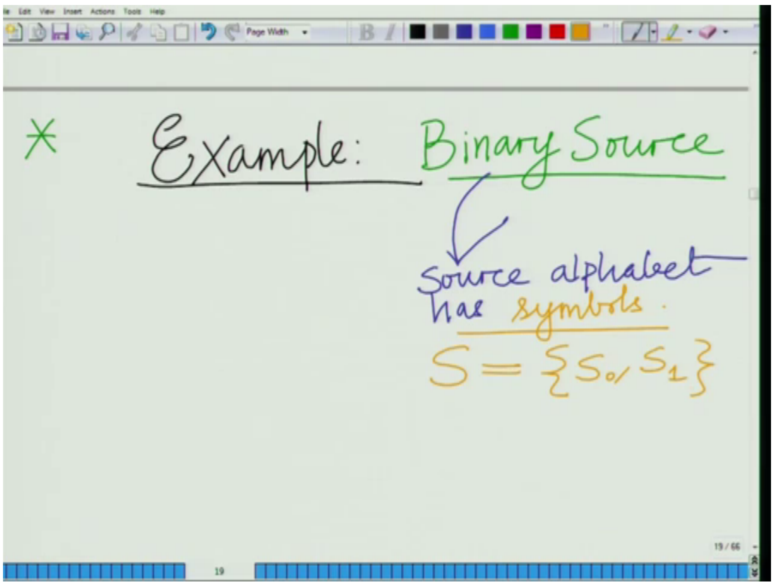
Principles of Communication Systems – Part II
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 28
Entropy Example – Binary Source, Maximum and Minimum Entropy of Binary Source

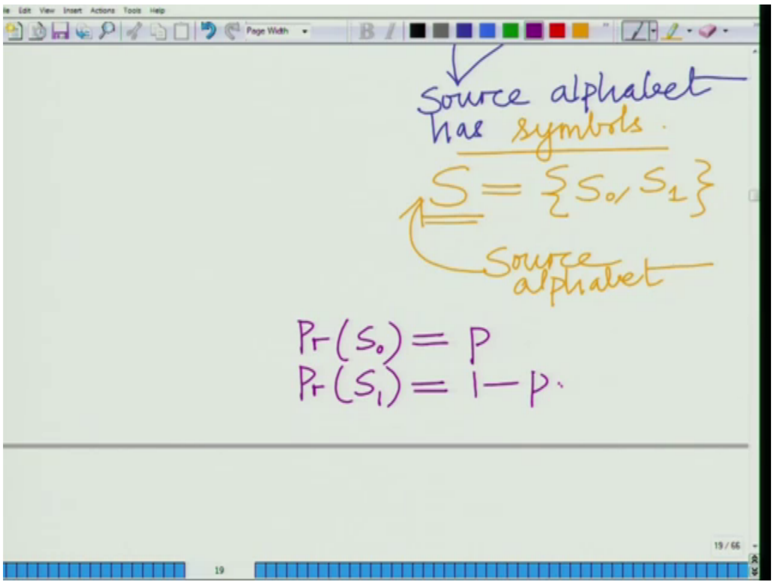
Hello, welcome to another module in this massive open online course. So, we are looking at various aspects of information theory and we have also defined the concept of entropy correct and we have seen various properties of the entropy of the source which based characterizes the average information or also the average uncertainty associated with this source. Average information per symbol of a source or the average uncertainty that is associated with this (Refer Time: 00:41). So, let us now look at an example to understand this better.

(Refer Slide Time: 00:48)



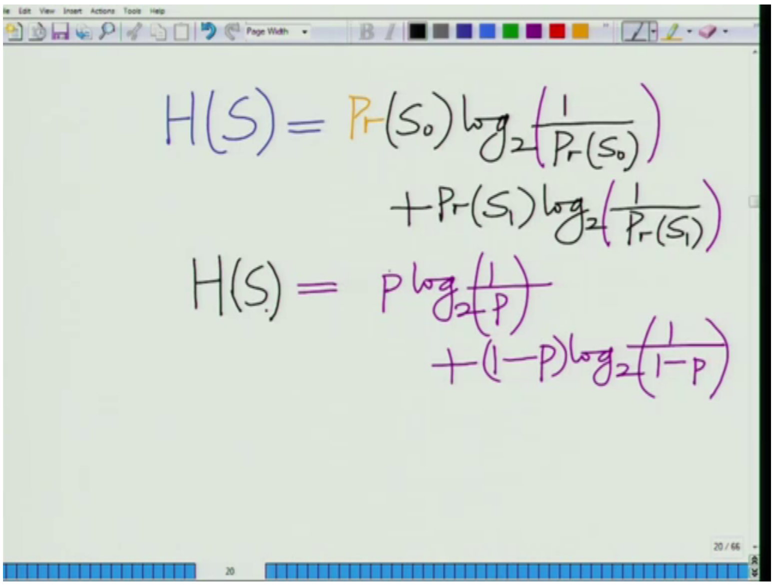
So, I want to look at an example, an example for entropies, correct. So, let us consider a binary source, binary source implies the source alphabet has 2 symbols - similar to 0 and 1 a digital. Source alphabet has 2 symbols. So, I can say S equals S naught and S 1 these are the 2 symbols of the source, correct.

(Refer Slide Time: 01:51)



So, remember this is the source alphabets of the 2 symbols which you call the source alphabet, this set is known as the source is not a source alphabet. Let probability of the symbol S naught be P the naturally probability of symbol S 1 equals 1 minus P, alright.

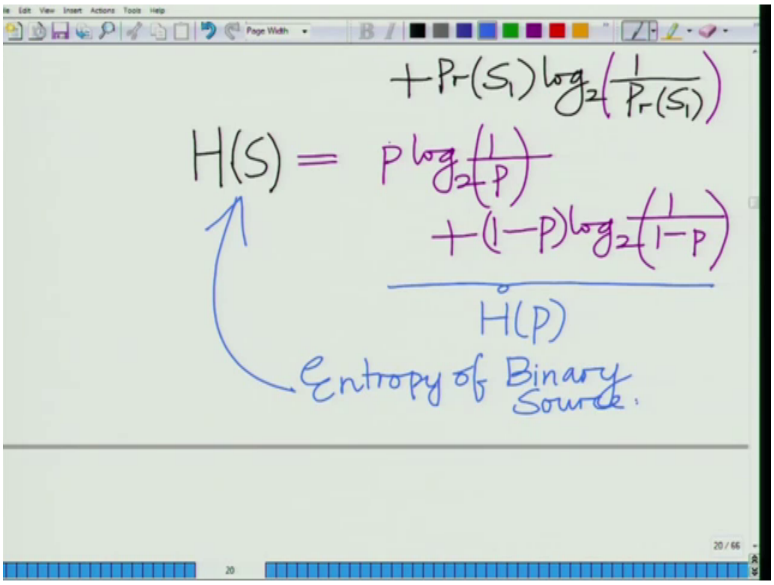
(Refer Slide Time: 02:20)



So, therefore, the entropy H of S is well P naught or basically let me write it, let me elaborate this that is probability of S naught log 2 the base 2, 1 over probability of S naught plus probability of S 1 log to the base 2 1 over the probability of S 1 which is basically nothing, but P times log to the base 2 1 over P plus 1 minus P log to the base 2

1 over 1 minus P. This is your H of S which is the entropy of the binary source and this quantity, this quantity is also known as H of P this quantity P times log to the base 2 1 over P plus 1 minus P times log to the base 2 1 over 1 minus P this is also denoted by H of P.

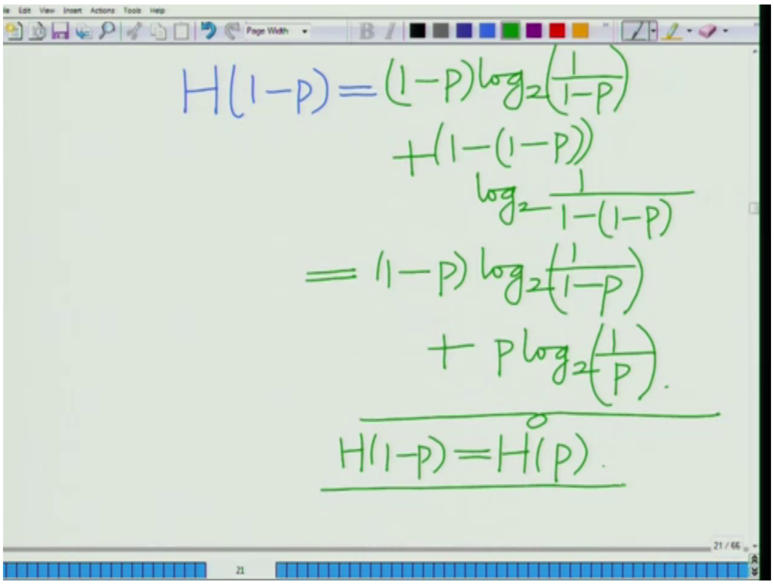
(Refer Slide Time: 03:40)



So, this is basically the entropy of the source or rather we can say entropy of a binary source, entropy of the binary source.

And now you will also see the first thing that the first property. So, we have derived the entropy of the binary source as P times log to the base 2 1 over P plus 1 minus P log to the base 2 1 over 1 minus P. Now you will also realize that for the first property is that you will see that H of P we have denoted this by H of P H of P is equal to H of 1 minus P.

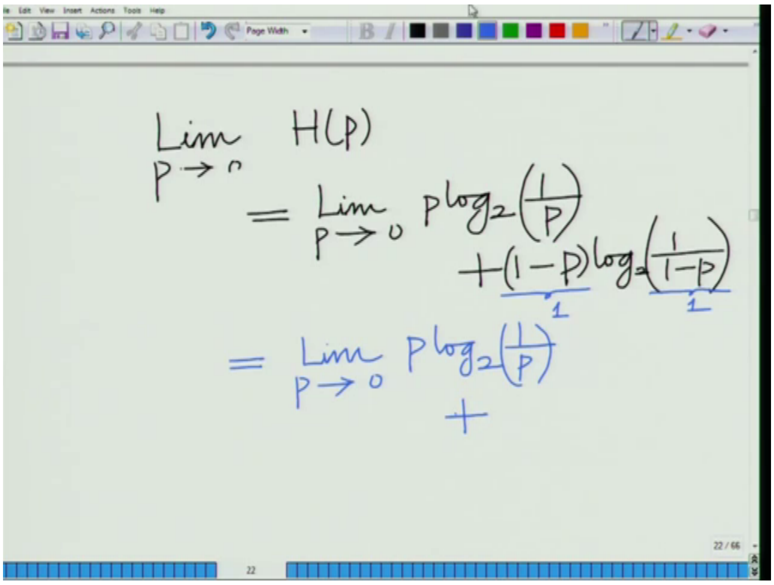
(Refer Slide Time: 04:39)



So, you have H of 1 minus P naturally you can see H of 1 minus P is 1 minus P log to the base 2 1 over 1 minus P plus 1 minus 1 minus P log to the base 2 1 over 1 minus 1 minus P, which is again you can see 1 minus P log to the base 2 1 over 1 minus P plus 1 minus 1 minus P is P log to the base 2 1 over.

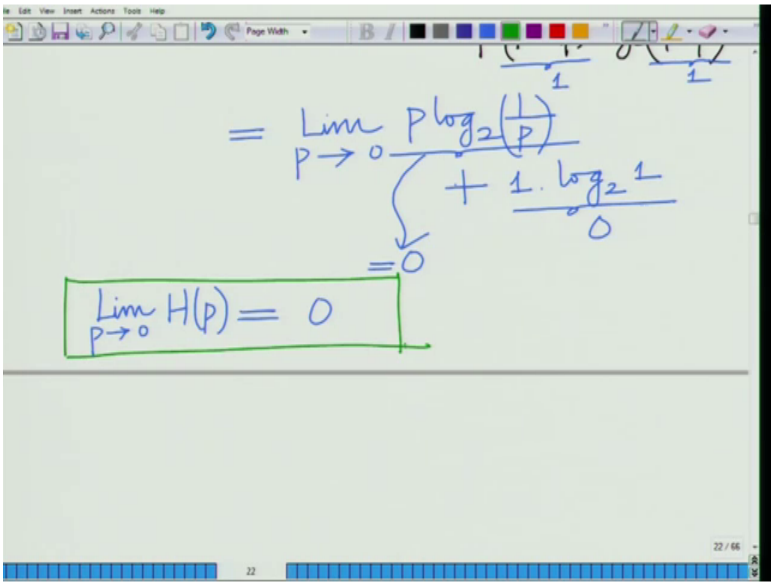
So, this is your, this is your quantity H of P. So, H of 1 minus P, so we have H of 1 minus P equals H of P that is the property and this is natural because we have seen that it only depends on the probabilities right it does not depend on the actual symbol. So, if we have binary source with the probabilities given by P that is 1 of the symbols as probability P the other as probability 1 minus P has a same information as were the first symbol as probability 1 minus P and other symbol as probability P.

(Refer Slide Time: 06:32)



Because we said it only depends on the probability the combination of probabilities, it does not depend on the symbols the information does not depend on the symbols themselves. Depends only on the probabilities with which they occur.

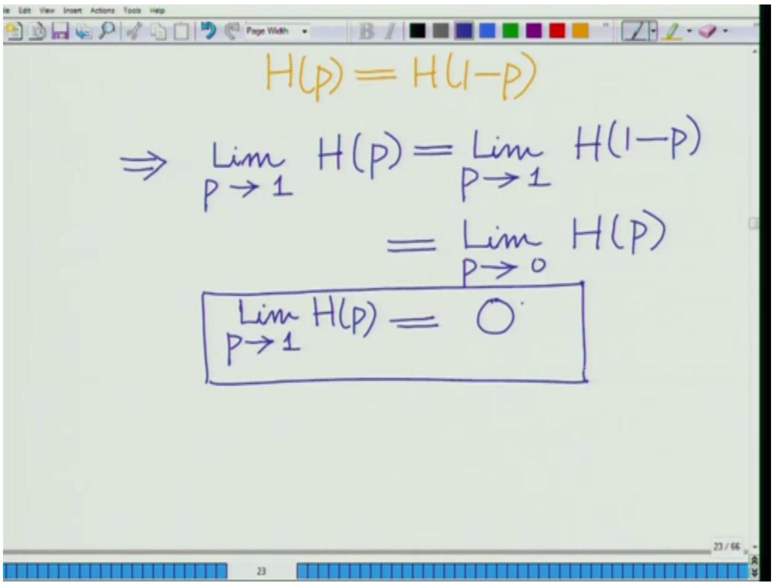
(Refer Slide Time: 07:40)



Now, also observe that limit P tending to 0 of H of P this is equal to well limit P tending 2 0 P log to the base 2 1 over P plus 1 minus P log to the base 2 1 over 1 minus P. Now as P tends to 0 1 minus P this becomes 1, 1 over 1 minus P this becomes 1. So, this tends to this is equal to well limit P tends to 0, of P log to the base 2 1 over P plus I can substitute

1. So, 1 log to the base 2 of 1 this quantity 0, we know we have previously derived that this quantity limit P tends to 0 P log to the base 2 1 over P this is equal to 0. So, overall this quantity equal to 0. So, limit P tending to 0 H of P is 0 we know. So, at P equal to 0. So, limit P tends, as P tends to 0 H of P tends to 0, correct. So, we have limit P tends to (Refer Time: 08:14).

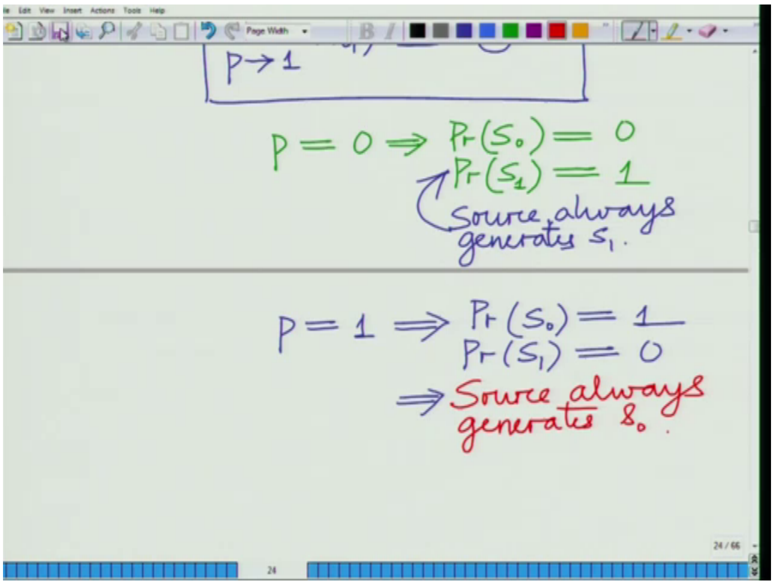
(Refer Slide Time: 08:16)



Similarly we have H of P equals H of 1 minus P we use the property H of P equals H of 1 minus P which now from which it now follows that limit P tend to 1 correct. So, now, it naturally follows that it naturally follows that limit P tending to 1 of H of P is also equal to well this is equal to since H of P is H of 1 minus P equals limit P tending to 1 of H of 1 minus P equals limit P tending to 0 of H of P this is equal to 0.

So, because if P tends to if P tends to 1 correct if P tends to 1 then 1 minus P tends to 0. So, which means that limit P tends to 1 H of P is also equal to 0. So, what you see is something very interesting when either P equal to 0 or P equal to 1 that is if the probability of 0 is equal to 0 or the probability of 0 is equal to 1 let us look at these 2 scenarios.

(Refer Slide Time: 10:07)

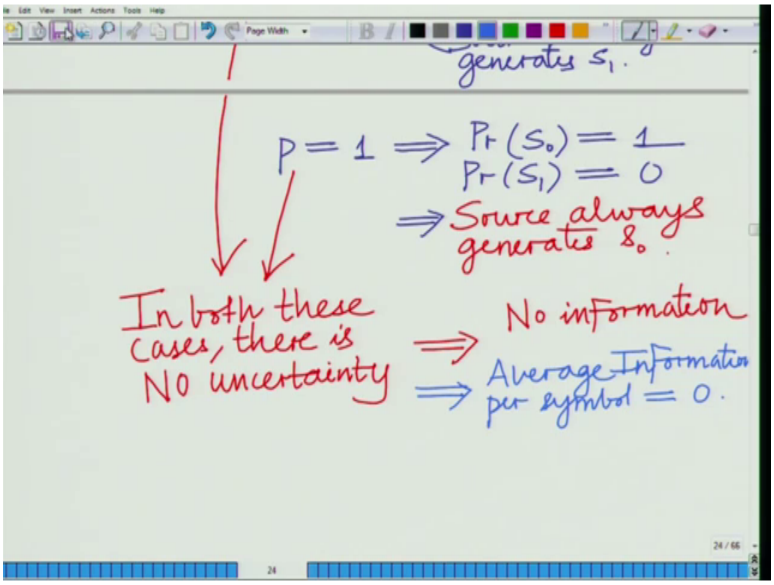


So, H P equal to 0 implies probability of S of 0, probability of S or 0 equal to 0, probability of S of 1 is equal to 1 which implies that source always generate symbol S 1. Because probability of S of 0 equals 0, source always generates S 1. On the other hand if P equal to 1 this implies probability of S of 0 equal to 1 probability S of 1 equal to 0 this implies that source always generates, source always generates the symbol S 0.

So, in either of these 2 cases when the source always generates S 0 or source always generates S 1 that is P equal to 0 or let us P equal to 0 or P equal to 1, the uncertainty is 0 because the source is always generating either S 0 or S 1. So, there is no information in the source because we know the next symbol the source is always going to; is going to generate is going to be either S 0 right, if P equal to 1 because with probability 1 it generates S 0 if equal to 0 with probability 1 it generates S 1. So, there is no uncertainty.

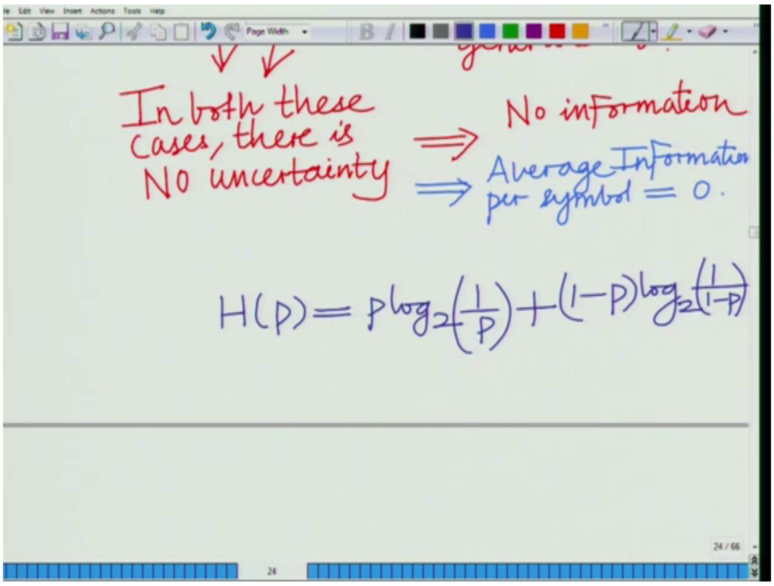
The source is always generating either 1 symbol or the other and hence the information right, since there is no since there is no uncertainty right, the information associated with the source average the entropy the average information per symbol associated with the source is 0. Since we know with certainty that the source is going to if P equal to 0 with certainty the source is generating S 1 if P equal to 1 with certainty the source is generating S c.

(Refer Slide Time: 12:40)



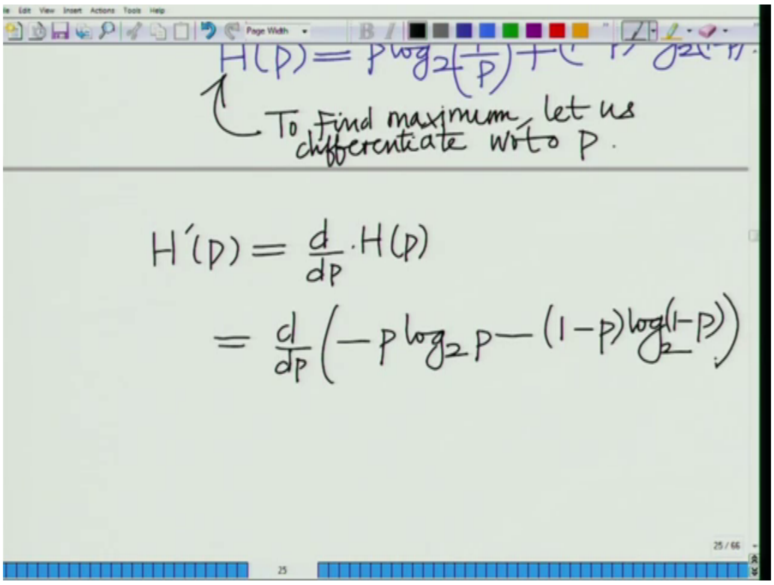
Now, let us look at when does the maximum entropy occur, when does the maximum average information per symbol occur for the source. So, in both these cases P equal to 0, in both these cases there is no uncertainty, this implies no information or that information content information content is 0. Implies, the average information per symbol is equal to 0.

(Refer Slide Time: 13:07)



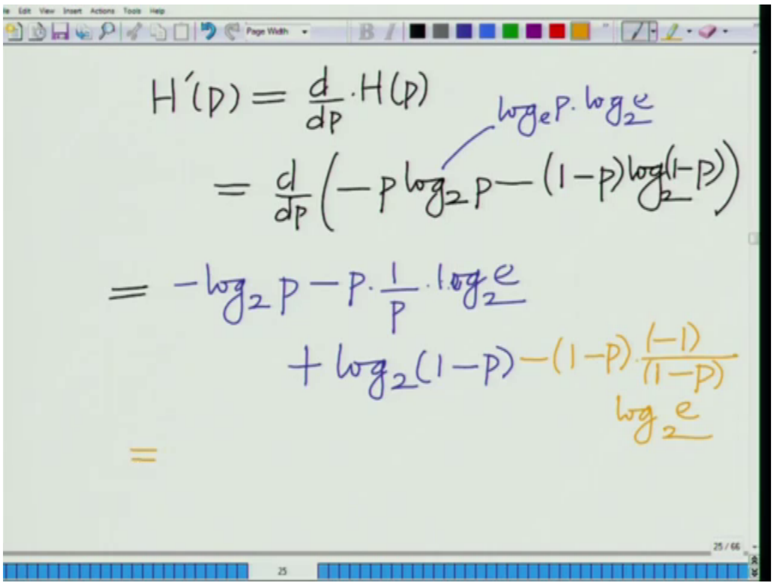
Now consider again H of P equals P log to the base 2 1 over P plus 1 minus P log to the base 2 1 over 1 minus P.

(Refer Slide Time: 14:12)



Now let us to find the maximum let us differentiate this with respect to P; to find maximum let us, to find the maximum let us differentiate with respect to P. So, H prime of P the derivative that is d over d P of H P well that is P log 2 to the base 1 minus that is basically you can write it as d or d P of minus P log 2 log 2 the base 2 P minus 1 over P minus 1 minus P log to the base 2 1 minus P.

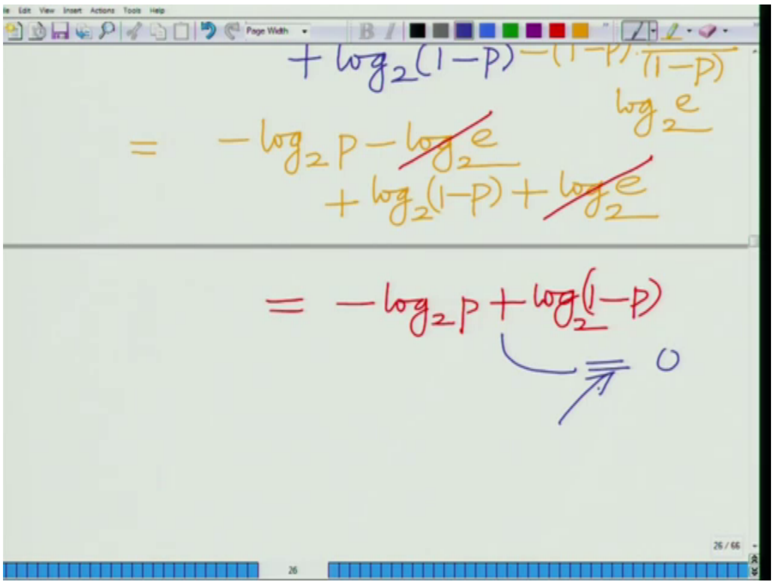
(Refer Slide Time: 15:21)



When you differentiate it of course, derivative with respect to P is 1. So, this is minus log to the base 2 P plus when you differentiate it with respect to the log 2 p. So, I am using the chain rule minus P log to the base 2 of P that can be written as well I can write this as log P to the base e that is the natural logarithm times log e to the base 2. So, that is minus P times well log P to the base e if you differentiate it that becomes 1 over P times log e to the base 2, correct.

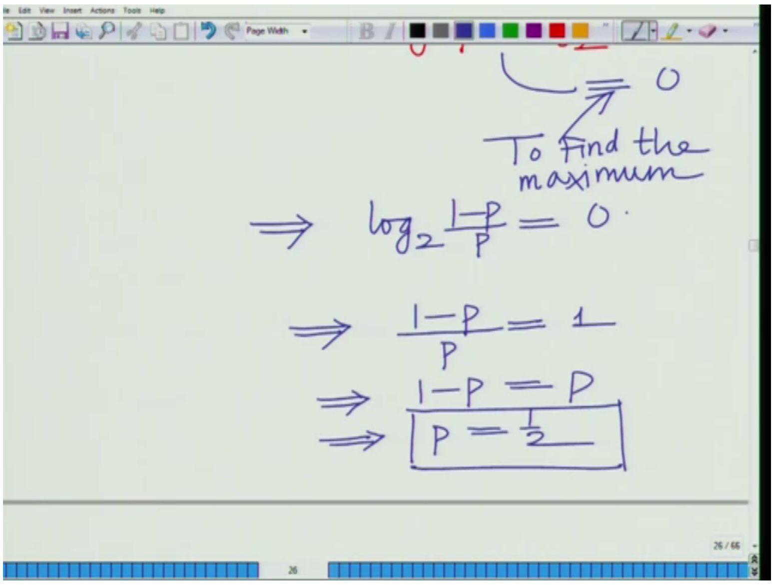
Similarly this is minus derivative of 1 minus P is 1. So, 1 minus P is minus 1, so this becomes plus log 2 to the base 1 minus p. And what we finally, left with is minus 1 minus P derivative of log to the base 2, 1 minus P is minus 1 over 1 minus P times log e to the base 2 which is equal to well minus well that will become minus log 2 to the base P minus log e to the base 2 plus log to the base 2, 1 minus P plus log to the base.

(Refer Slide Time: 16:59)



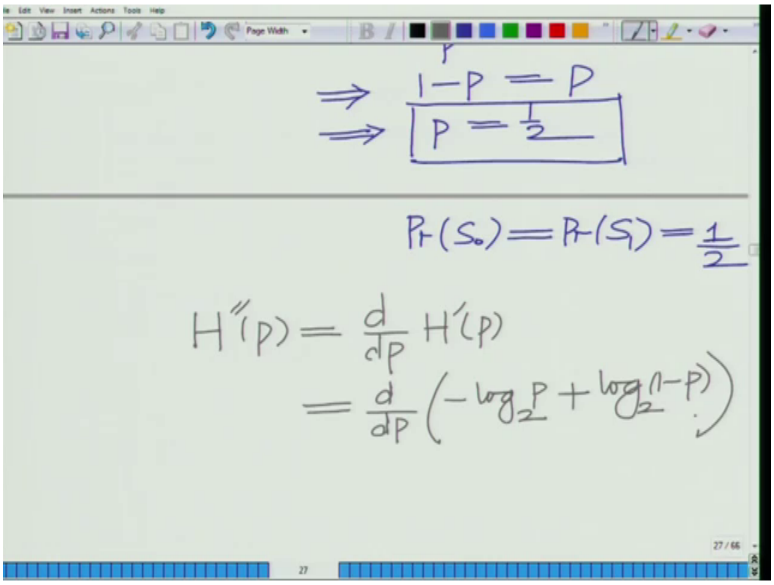
So, you cancel these and what we get is minus log 2 the base, minus log P to the base 2 plus log 1 minus P to the base 2. Now if you equate this to 0, equate to 0 to find, so differentiate.

(Refer Slide Time: 17:53)



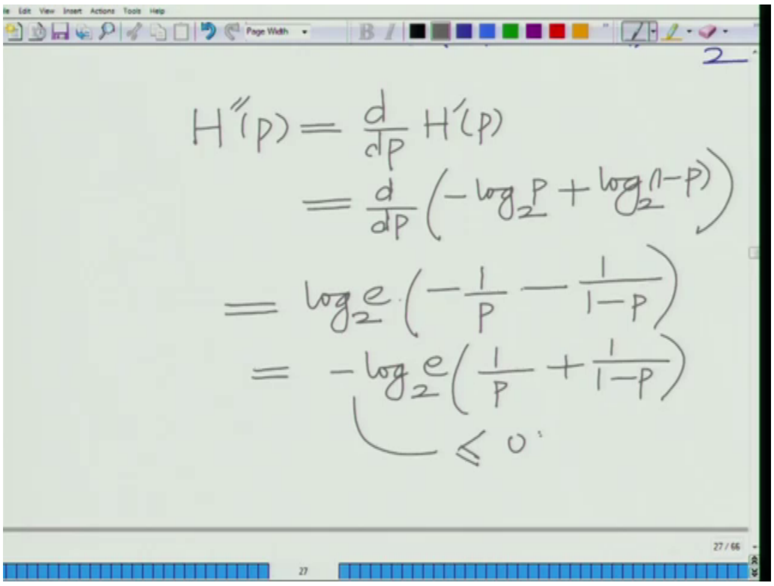
So, differentiate and equate to 0 to find maximum and this implies that basically log 2 to the base 1 minus P over P equal to 0 which implies 1 minus P over P equal to 1 which implies 1 minus P equals P which naturally implies P equal to half. So, we get that entropy is maximum for P equal to half. That is when both the symbols that is S naught both the symbols S naught in S 1 of this binary source an equal probability then the entropy is maximum.

(Refer Slide Time: 18:50)



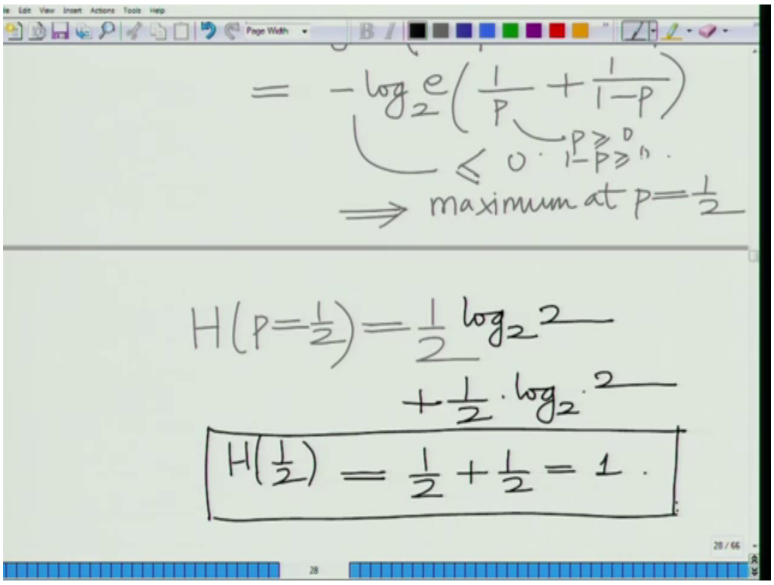
So, we have entropy is maximum for probability S naught equals probability S 1 equals 1 half and you can also see that this is the maximum because if you differentiate it again H double prime of P that will be well, H prime of P we have that is derivative of H prime of P that is the derivative of well minus log to the base 2, minus log P to the base 2 plus log 1 minus P to the base 2 which is equal to if you take log e to the base 2 common, this will be minus 1 over P minus 1 over 1 minus P.

(Refer Slide Time: 19:45)



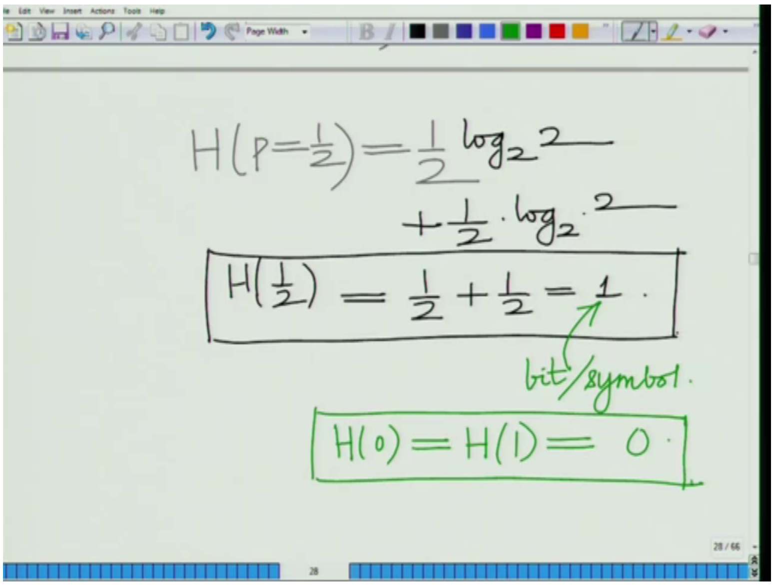
So, this is minus log e to the base 2 1 over P plus 1 minus P which is clearly less than or equal to 0, second derivative less than equal to 0 implies there is a maximum and this quantity remember P is greater than equal to 0. 1 minus P is greater than equal to 0, therefore, this quantity is less than equal because of the negative sign it is less than equal to 0 implying that indeed at P equal to half there is a maximum the entropy reaches its maximum.

(Refer Slide Time: 20:22)



And what is the entropy at P equal to half? H of P equal to half or rather H of half equals half log to the base 2 1 over half that is 2 plus 1 minus P that is half log to the base 2 1 over 1 minus P that is 1 over half that is 2 log to the base 2 is log to the base 2 of 2 is 1. So, this is simply half plus half equals 1; so H of half.

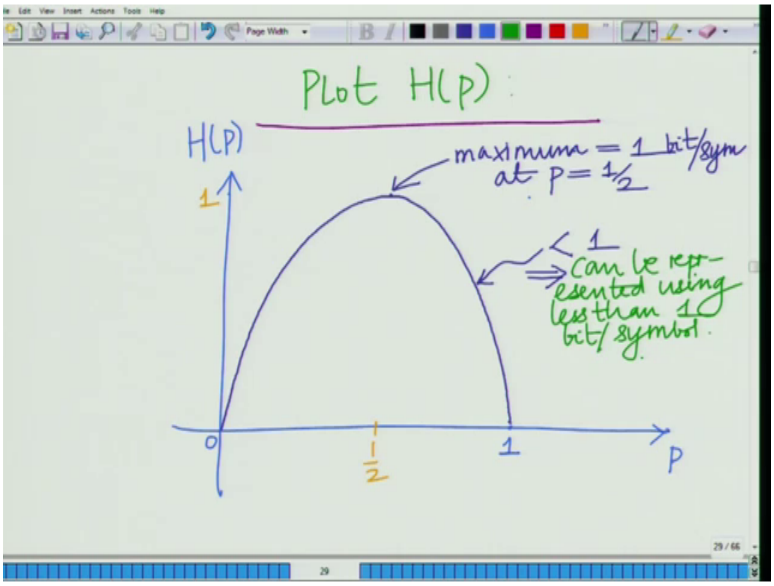
(Refer Slide Time: 21:33)



So if you observe this is something very interesting. So, H of half equal to 1 which means this is 1 bit, 1 bit per symbol, this is 1 bit which means H of half equal to 1 this is

1 bit per symbol and we have already seen that H of 0 that is limit P tends to 0, P tends to 1 if we denote them by H of 0 and H of 1 they are 0, and it reaches maximum at P equal.

(Refer Slide Time: 22:06)



So, if you plot this. So, if you plot this entropy you will realize that it looks something like; you will realize it looks something like this. So, this is on the x axis we have p, on the y axis we have H of p. So, this is P equal to 0 this is P equal to 1 and we have this add it starts at it reaches its maximum that is at half it is equal to 1. So, it starts from 0 goes all the way up to 1 at P equal to half. So, it is something like this. It is not pointed in fact, its derivative; its derivative at half is 0 which is what I am trying to show approximately. So, this is how it looks like. So, at half it reaches its maximum that is maximum equal to 1 at P equal to half.

This implies that here this is in fact 1 bit per symbol correct. So, at all these points, for instance if you look at this point this is less than 1 this implies it can be represented with lower number of. So, we have 2 symbols 0 and 1. What this means is the average entropy is less than 1 implies that it can be represented with less number of bits per symbol than 1. Of course, when P equal to half it implies that you need on an average 1 bit per symbol to represent this and when entropy is less than 1 for instance when P equal to 3 by 4 correct. Entropy is less than 1 you can see that it needs requires less than 1 bit per symbol on an average to represent this all right and that is what we are going to see as you progress to the rest of this course.

So, for instance less than 1 implies lower than 1 symbol can be represented using less than. So, implies can be represented and this is what the encoding process does right, the encoding process expected to do is take this source symbols and come up with an efficient code conduct with an efficiency stream write as information bits stream to represent this using the minimum number of bits right take to binary symbols S naught and S 1 correct, S naught and S 1 and come up with an information bit stream to represent this efficiently using the lowest the least number of possible bits that is the lowest possible bits in an average sense. That is the average number of bits used to represent each symbol must be the smallest and that is what we are going to this process termed as coding or encoding. In fact, that is the theory regarding this is what we are going to look at in detail as you go through the rest of the modules of this course.

So, basically that is what we have seen we have seen an interesting example of a binary source with 2 symbols S naught an S 1 probabilities P and 1 minus P we have looked at what happens at the extremes that is P equal to 0 P equal to 1 there is no uncertainty, the information is 0, maximum information or entropy occurs at P equal to half. And of course, if you plot it, it looks like a nice concave function which rises starts from 0 at P equal to 0 rises to 1 at P equal to half and falls down again to 0 at P equal to 1, all right.

So, will stop here and look at explorer other aspects, other properties study this behavior of entropy further in a subsequent modules.

Thank you very much.