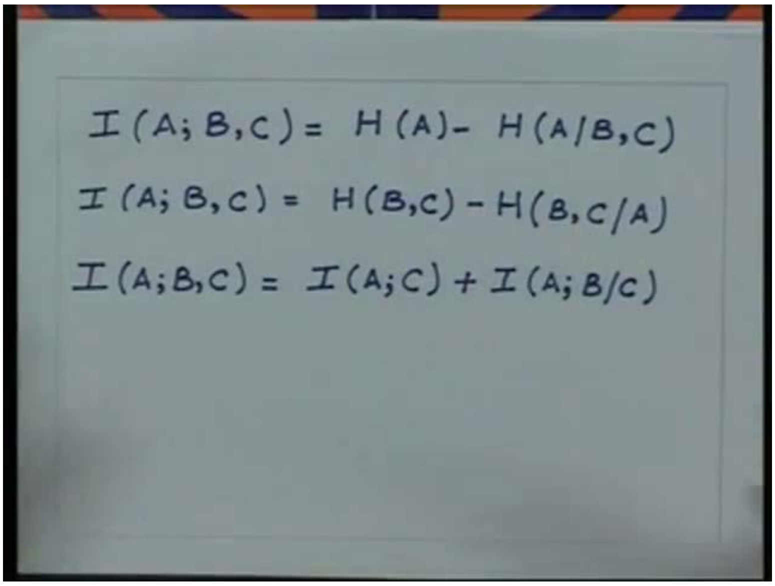
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of technology, Bombay

Lecture - 24
Properties of Mutual Information and Introduction to Channel Capacity

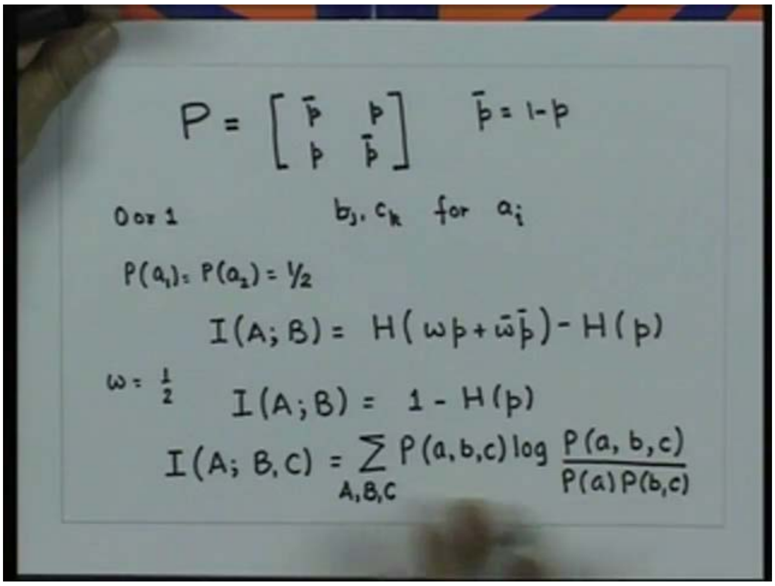
In the last class, we studied the additive property of mutual information and we derived some of the important results. So, these results are depicted here.

(Refer Slide Time: 00:54)



Mutual information of A and B and C is equal to uncertainty of alphabet A less uncertainty of alphabet A, after we have observed B and C. We can also interpret mutual information of A and B and C as uncertainty of alphabet B, C. Less uncertainty of alphabet B, C after we have observed alphabet A. The additive property of mutual information is given by this relationship which says that mutual information of A and B, C is equal to mutual information of A, C plus mutual information of A, B given, we have observed alphabet C. Now, let us illustrate the additivity property of mutual information by examining the binary symmetric channel.

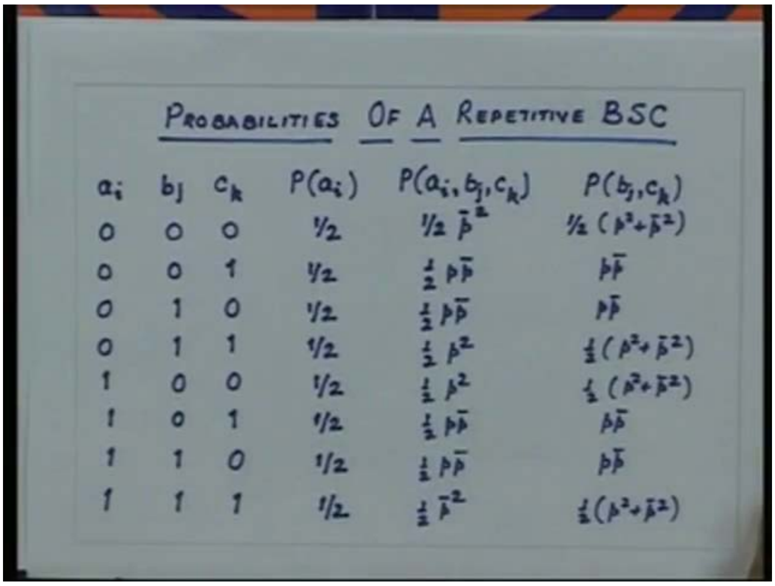
(Refer Slide Time: 02:14)



The binary symmetric channel matrix is given as follows, p bar, p, p, p bar, whereas usual p bar is equal to 1 minus p. Now, this time we assume that the input symbol either 0 or 1 is repeated, so that the output of the channel consist of 2 binary symbols b j and c k, for each input symbol a i. Furthermore for simplicity, we assume that the 2inputs a1 and a2 are chosen with equal probabilities.

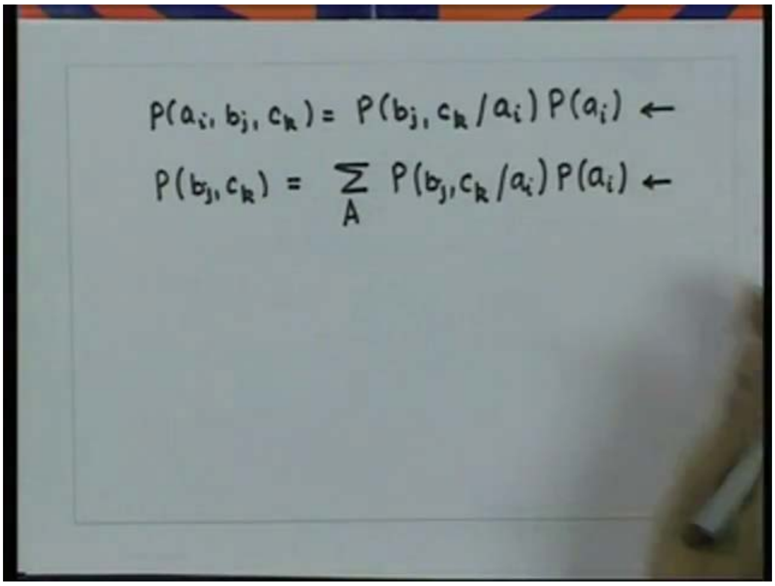
Therefore, we set probability of a1 equal to probability of a 2, is equal to half. We have also look at the relationship of mutual information between two alphabets A and B, which is equal to Hof w p plus w bar, p bar minus H p. Now, in this expression we substitute omega equal to half, we get I of A; B equal to 1 minus H of p. Now, to find mutual information of A semicolon B, C, we use the relationship which we derived in the last class and that is given by probability of a, b and c log of probability of a, b, c divided by probability a by probability b, c, the summation is over A, B and C alphabet.

(Refer Slide Time: 05:12)



Now, the table here gives the necessary probabilities associated, for calculation of mutual information. So, a i, b j, c k have been listed probability of a i, we assume equal to half, calculation of probability of a i, b j, c k is based on the fact that.

(Refer Slide Time: 05:42)

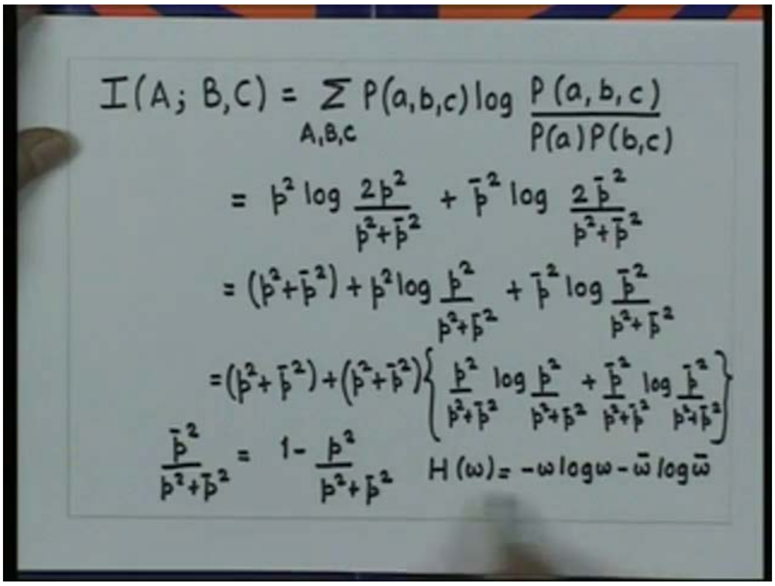


Probability of a i, b j, c k is equal to probability of b j, c k given a i, multiplied by probability of a i. If we use this relationship, we get this column as shown here. Similarly, to calculate the probability of b j, c k, we use the fact that the probability of b j, c k is equal to summation of probability of b j, c k given a i multiplied by probability of a i.

This is summed over alphabet A, so using this relationship, we get this column in this table. If we take this case the probability of b j, c k given a i that is 0and probability b j, c k that is 0, 0 given a i equal to 1. We add these two probabilities and we get this result. Similarly, to get this term probability of 0, 1 given 0 and probability of 0, 1 given 1, when we add these 2 terms we get pp bar.

Similarly, we get all the terms in this column. Once we have these probabilities we can calculate the mutual information of A semicolon B, C. Now, when we calculate the mutual information the this terms will not contribute to the calculation, because if we divide this half pp bar upon multiply half pp bar, this will cancel and we will get the ratio equal to 1.

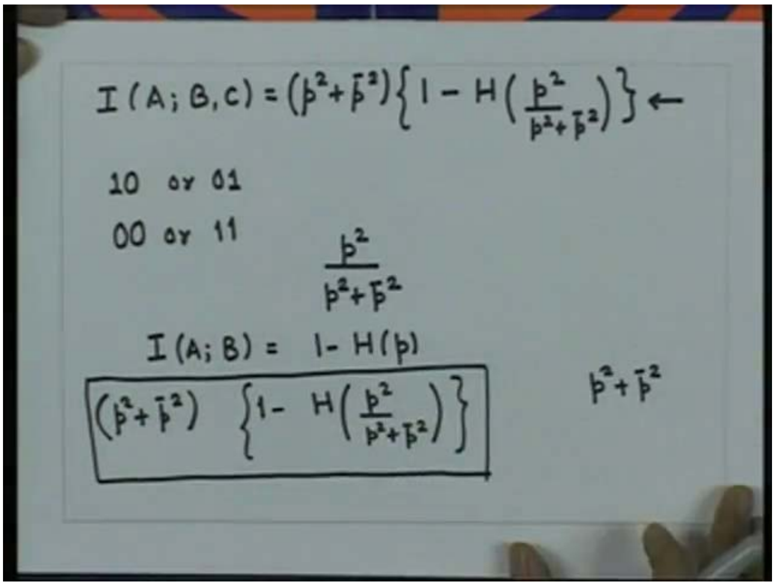
(Refer Slide Time: 08:24)



So, let us calculate I of A; B, C equal to probability of a, b, c log of probability of a, b, c, divided by probability of a probability b, c summed over A, B, C. When we substitute the terms from the table we get the relationship p squared log 2, p squared upon p squared plus p bar squared plus p bar squared log of 2 p bar squared p square plus p bar squared. So, in this expression only 1, 2, 3, 4 terms come into picture, where as the contravention from other terms are 0. Now, this can be simplified on expanding this on each of this term and we will get the result as follows plus p bar squared log of p bar squared upon p squared plus p bar squared.

Now, this quantity can be simplified as p squared plus p bar squared plus p squared plus p bar squared p squared upon p squared plus p bar squared log of p squared upon p squared plus p bar squared plus p bar squared upon p squared plus p bar squared log of p squared upon p squared plus p bar squared. Now, using the relationship that p bar squared upon p squared plus p bar squared is equal to 1 minus p squared upon p squared plus p bar squared and entropy function, which is defined as equal to H omega minus omega log w minus omega bar log omega bar, using these two relationships.

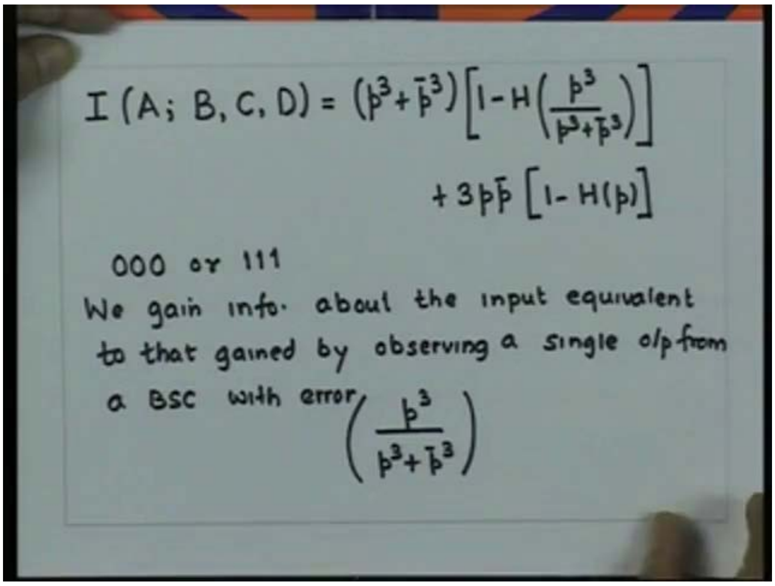
(Refer Slide Time: 12:10)



We can show that mutual relationship of A semicolon B, C is equal to p squared plus p bar squared, 1 minus H of p squared upon p squared plus p bar squared. Now, the interpretation of the equation is very clear, if we observe the output 1, 0 or 0, 1. From such a channel it is meaning is entirely ambiguous, the two possible inputs will still be equally probable and we gain no information from our observation. If, on the other hand we observe 0, 0 or 1, 1 we gain information about the input equivalent to that gain by observing a single output from binary symmetric channel with error probability equal to p squared upon p squared plus p bar squared.

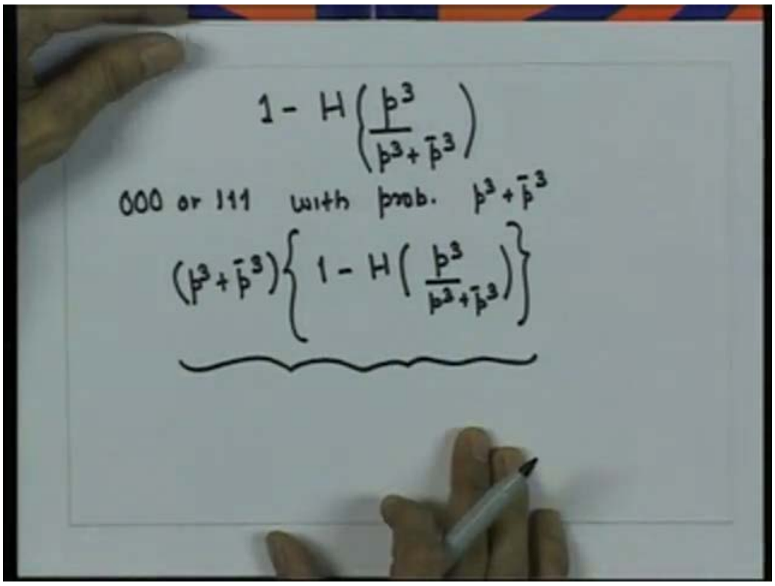
Now, we know that information I A; B is equal to 1 minus Hof p, if the two inputs probable. So, using this relationship the information we get on observing 0, 0 or 1, 1 is equal to 1 minus H of p squared upon p squared plus p bar squared. Now, we observe either 0, 0 or 1, 1 with probability p squared plus p bar squared. Therefore, it implies that the information which I get when I observe 0, 0 or 1, 1, gets this term gets multiplied by p squared plus p bar squared. So, this is the information mutual information of A semicolon B, C, which we have derived based on the arguments. Now, we can extend this case of a binary symmetric channel, when we have more than single repetitions. For example, each input produces three binary outputs then we can show that.

(Refer Slide Time: 15:22)



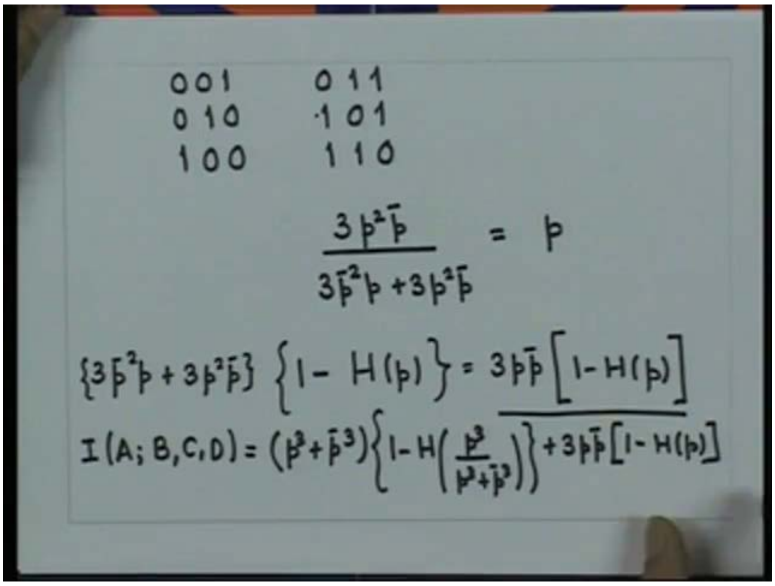
Mutual information of A semicolon B, C, D is equal to p cube plus p bar cube multiplied by 1 minus H of p cube over p cube plus p bar cube plus 3 pp bar plus1 minus H of p. Let us look at the arguments which go in the derivation of this expression. Now, when we observe 0, 0, 0 or 1, 1, 1 we gain information about the input equivalent to that gained by observing a single output from a binary symmetric channel with error equivalent to p cube upon p cube plus p bar cube.

(Refer Slide Time: 17:29)



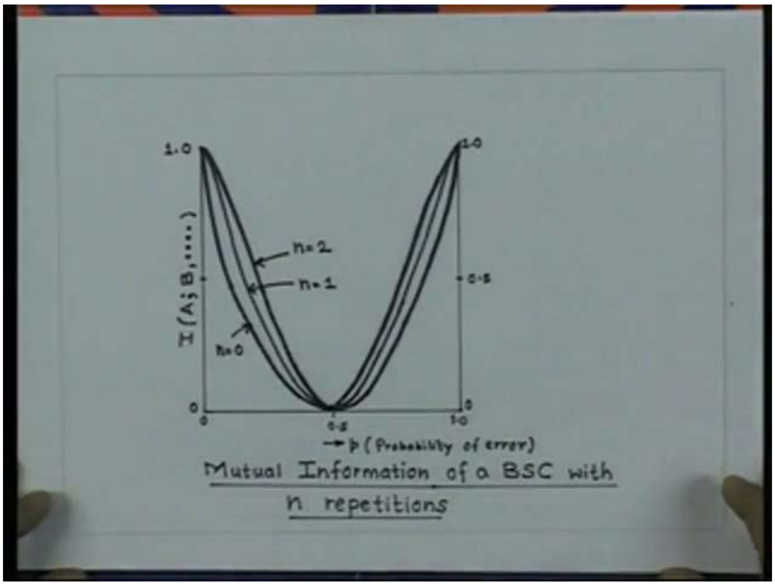
Therefore, information from such an observation is equivalent to1 minus H p cube over p cube plus p bar cube. We observe either 0, 0, 0 or 1, 1, 1 with probability p cube plus p bar cube. Therefore, the information which we gain when we observe 0, 0, 0 or 1, 1, 1 will be given by p 3 plus p bar 3 multiplied by this expression. So, this is of one term, which we get for mutual information of A, B, C, D.

(Refer Slide Time: 18:42)



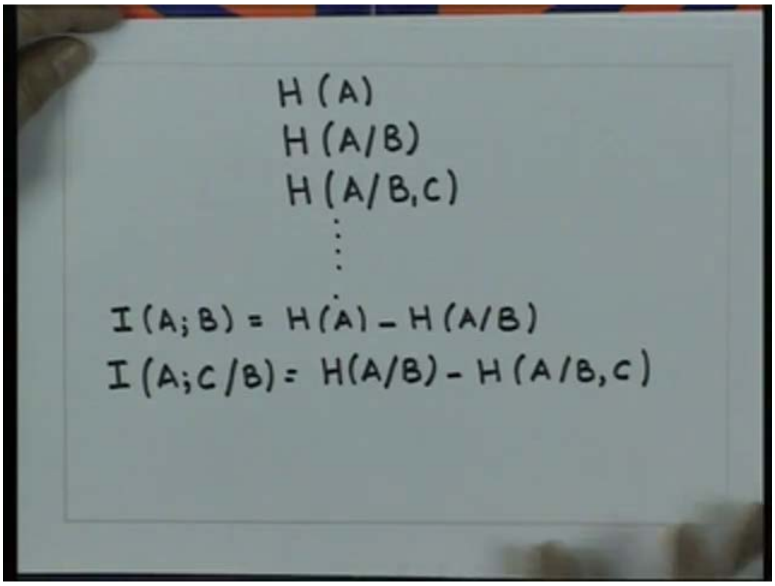
So, when we observe 0, 0, 1 0, 1, 0 1, 0, 0 or 0, 1, 1 1, 0, 1 1, 1, 0 then in this case, we gain information about the input equivalent to that gain by observing the output from a binary symmetric channel with error probability given by 3 p squared p bar upon 3 p bar squared p plus 3 p squared p bar. And this can be simplified as p therefore, when you observe either quantity, we will get information given by H of 1 minus H of p multiplied by the probability of observing this event, probability of observing this event is given by 3 p bar squared p plus 3 p squared p bar and this can be simplified as 3 pp bar multiplied by 1 minus H of p. Therefore, total information which we get between A and B, C, D is given by p cube plus p bar cube multiplied by 1 minus H of p cube upon p cube plus p bar cube plus 3 pp bar 1 minus Hof p. We can plot mutual information versus probability of error for different cases and if you do this, we get this result.

(Refer Slide Time: 21:22)



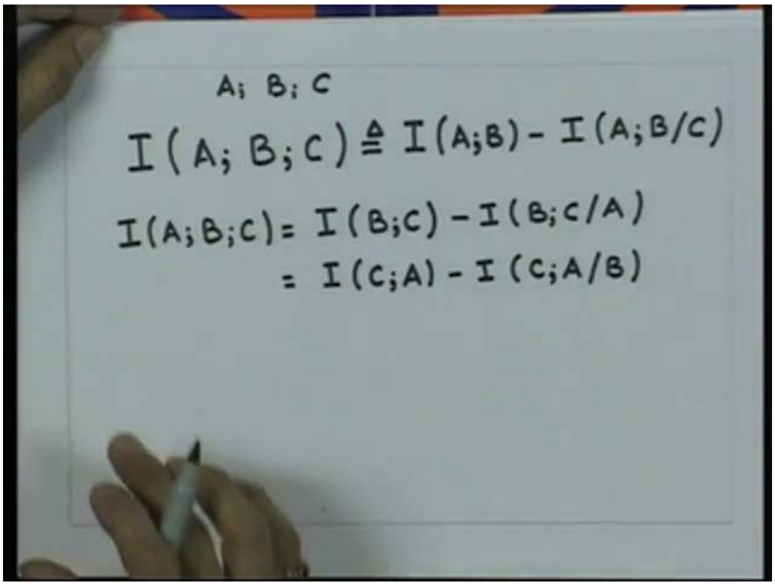
In this figure, X axis is depicting p that is probability of error on the Y axis, we have the mutual information and this is the graph we have for n equal to 0, this is for n equal to 1 and n equal to 2. So, from this graph it is clear that for given probability of error the mutual information of a binary symmetric channel increases with n repetitions. Now, in our investigation of additivity of mutual information, we were let to consideration of sequence of entropy quantinance of the form.

(Refer Slide Time: 22:17)



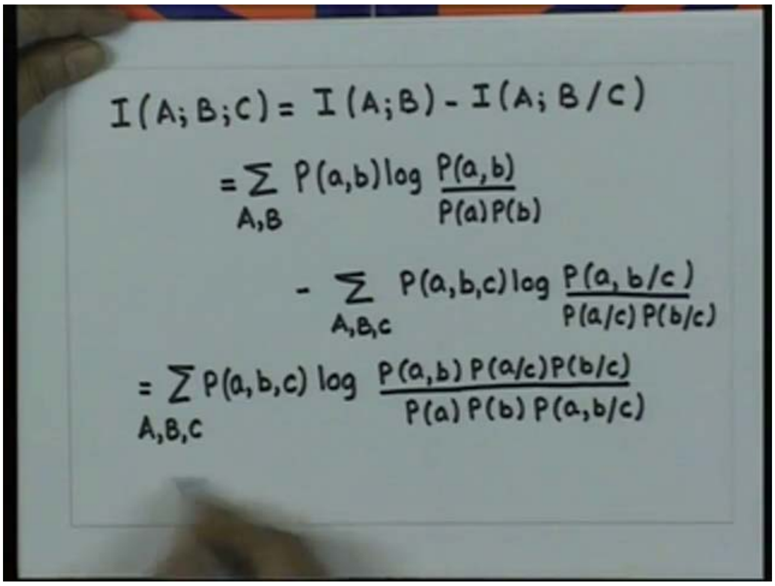
H of A, H of a given B, H of A given B, C and so forth. Now, each member of the sequence is no greater than the preceding member. Now, we also saw that the difference between 2 successive members could be interpreted as the average information about a provided new observation. So, I of A; B is equal to H of A minus H of A given B similarly, I of A; C given B is equal to H of A given B minus H of A given B, C. I A; B is the mutual information of A and B, I A semicolon C given B is the mutual information of A and C after we have given B. Now, both these quantities however involve the mutual information of just two alphabets; it is also possible to define mutual information of more than two alphabets.

(Refer Slide Time: 25:05)



So, let us define the mutual information of A; B and C alphabet, so by definition it is given by A semicolon B semicolon C is equal to I A semicolon B minus I A semicolon B given C. Now, our definition of the mutual information of A; B; C implies that this quantity is symmetric in A; B and C. Now, we will try to prove this, so let us if this is true we can prove I of A semicolon B semicolon C is equal to I of B semicolon C minus I of B semicolon C given A, or we can also write it as I of C semicolon A minus I of C semicolon A given B. So, to prove the symmetry of this quantity, let us write down.

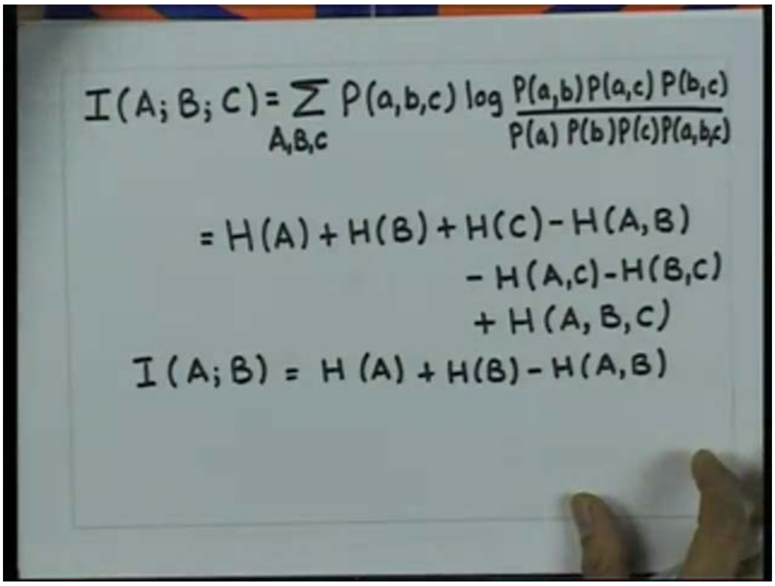
(Refer Slide Time: 25:41)



The expression for I A semicolon B semicolon C, which is given as I A semicolon B minus I A semicolon B given C. This we can rewrite as summation over the alphabets A, B probability of a, b log of probability a, b divided by probability of a multiplied by probability b. This is the expression for I A semicolon B. Similarly, we can write down the expression for this term and that is as follows summation over alphabet A, B, C probability of a, b, c log of probability a, b given c divided by probability of a given c, multiplied by probability of b given c.

Now, this can be simplified as follows we can combine these two term by rewriting probability of a, b as probability of a, b, c that with the summation over a, b, c log of probability a, b probability of a given c probability of b given c divided by probability of a multiplied by probability of b multiplied by probability of a, b given c. Now, if you multiply the term in the argument of log both the numerator and denominator by probability of c into probability c.

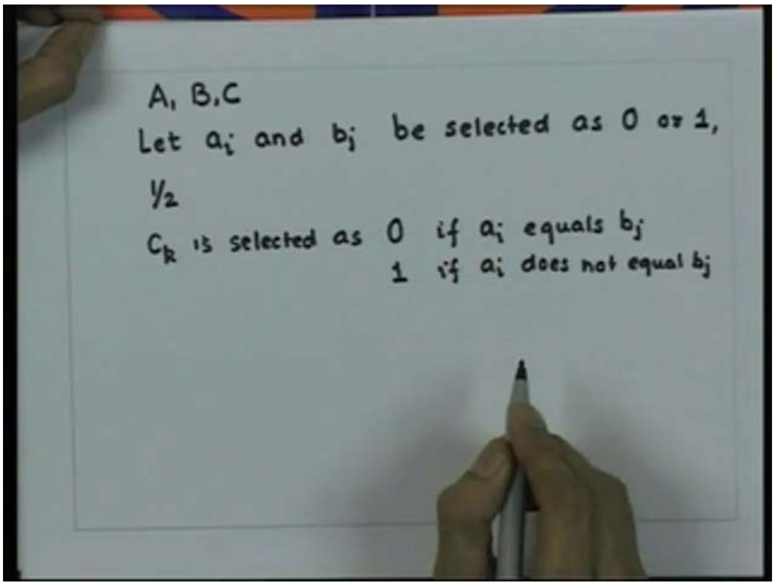
(Refer Slide Time: 28:24)



Then we can simplify the expression as I of A semicolon B semicolon C, equal to summation over alphabets A, B, C log of probability of a, b probability of a, c probability of b, c divided by probability a probability b probability c and probability of a, b, c. Now, by definition of entropy, we can simplify the expression as, entropy of alphabet A plus entropy of alphabet B plus entropy of alphabet C minus joint entropy of alphabet A, B minus joint entropy of alphabet A, C minus H of B, C plus H of A, B, C.

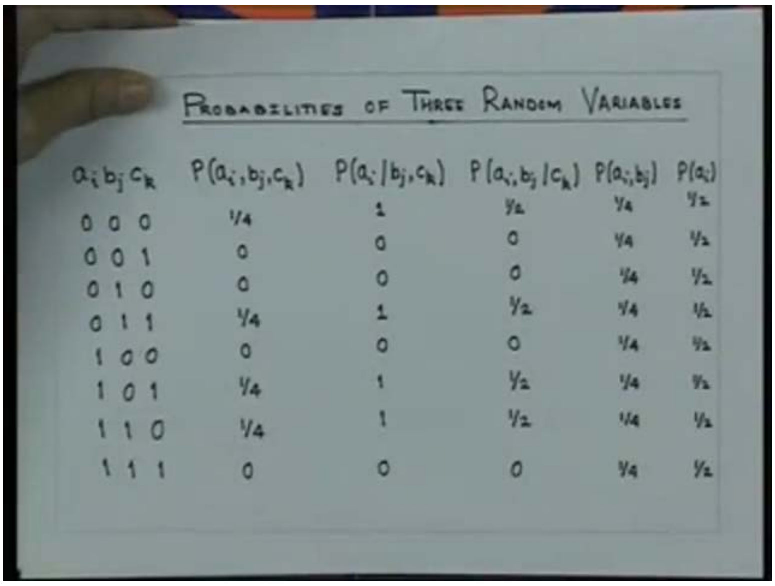
Now, this expression which we have derived is suggestive of the expression for mutual information of two alphabets which is given by I A; B is equal to H A plus H B minus H of A, B. Now, it is easy to generalise both these expression to more than three alphabets, but we will restrain ourselves from doing that, because that is not required for our course. Now, it is also important to note that the mutual information of A semicolon B is always non negative, however the mutual information of A semicolon B semicolon C can be negative. So, to show this let us take an example.

(Refer Slide Time: 31:02)



So, consider the three binary alphabets A, B, C. Let a i and b j be selected as 0 or 1 each with the probability of half, and each independently of the other, finally we assume that c k is selected as 0, if a i equals b j and ck is selected as 1, if a i does not equal b j. So, some of the probabilities of these random variables are given in the table as shown here.

(Refer Slide Time: 32:24)



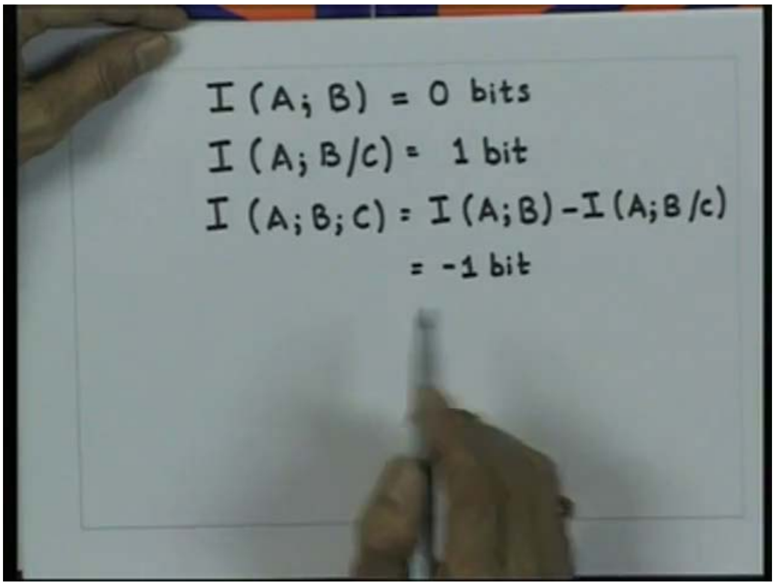
$a_i\ b_j\ c_k$	$P(a_i, b_j, c_k)$	$P(a_i / b_j, c_k)$	$P(a_i, b_j / c_k)$	$P(a_i, b_j)$	$P(a_i)$
0 0 0	1/4	1	1/2	1/4	1/2
0 0 1	0	0	0	1/4	1/2
0 1 0	0	0	0	1/4	1/2
0 1 1	1/4	1	1/2	1/4	1/2
1 0 0	0	0	0	1/4	1/2
1 0 1	1/4	1	1/2	1/4	1/2
1 1 0	1/4	1	1/2	1/4	1/2
1 1 1	0	0	0	1/4	1/2

So, we have listed all the different combinations of a i, b j, c k. With each of these combination we have provided the probabilities a i, b j, c k, probability of this quantity occurring is probability of getting a i equal to 0 and probability of b j equal to 0. Assuming that each of this is half and independent, probability of 0, 0, 0 is one-fourth probability of occurrence of 0, 0, 1 will be 0.

Similarly, probability of 0, 1, 0 will be 0 and probability of 1, 1, 0 is one-fourth based on the same arguments. Probability of a i given b j comma c k is indicated in the column here. Now, once we have given b j and c k we know with probability 1 what a i we have received, so if we know that b j and c k are 0, 0, we know that a i has to be 0 and therefore, this probability of 0 given is equal to 0, 0, 1. Similarly, the probability of 0 given of 0, 1 is always 0, because this possibility can never occur.

So, based on this argument we have the listings for the probabilities of the terms. Similarly, we can find out the probability of a i, b j given c k. Once c k has been given if it is 0 then the probability of 0, 0 is again decided by the probability of the occurrence of a i and that is equal to half. Similarly, probability of 0, 0 given 1 is 0 and probability of a i, b j assuming they are independent are each equal to one-fourth and probability of a i is equal to half.

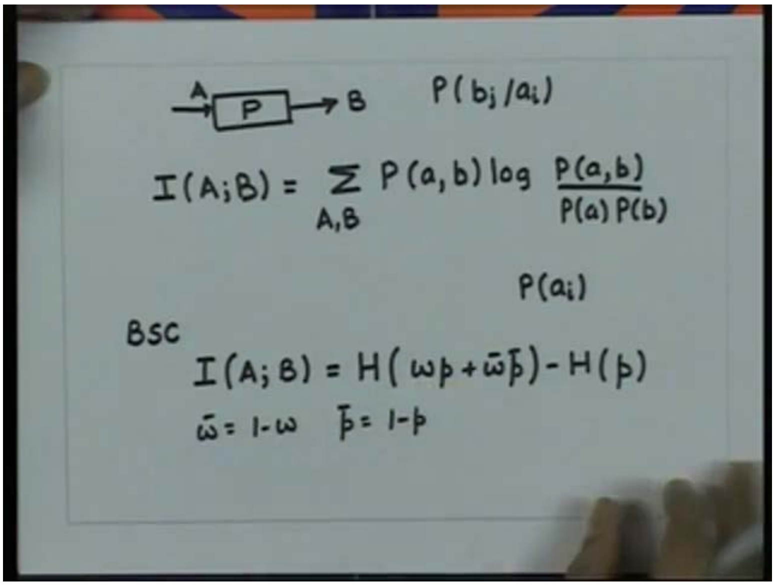
(Refer Slide Time: 35:10)



Now, using this table and using the expression, which we have derived earlier, we can calculate the information of... A semicolon B is equal to 0 bits information that is mutual information. That is A semicolon B given C is equal to 1 bit and mutual information A semicolon B semicolon C is equal to I of A semicolon B minus I of A semicolon B given C and that is equal to minus 1 bit. It is clear, why we get such an answer since A and B are statistically independent I of A semicolon B is equal to 0 bits that is b provides no information of A.

Now, if you already know C, however learning B tells us which a was chosen and therefore, probability this provides us with one bit of information. Now, after studying mutual information its property. Let us study very important property of information channel and that is channel capacity. So, let us consider information channel with input alphabet.

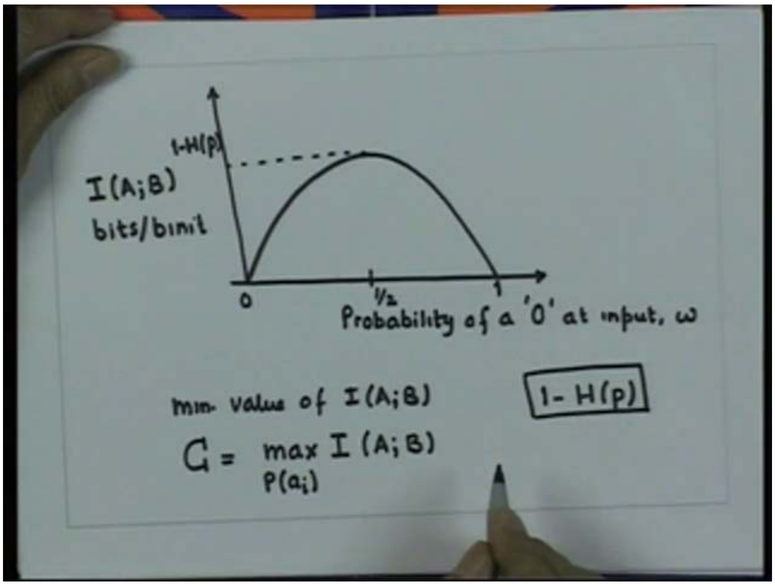
(Refer Slide Time: 36:38)



A output alphabet B with the channel matrix given by P, so the conditional probabilities that is probability b j given a i is specified. Now, in order to calculate the mutual information which is given by I A; B is equal to summation of probability a, b log of probability of a, b divided by probability of a multiplied by probability of b, summation over alphabet a, b. In order to calculate this mutual information, it is necessary to know the input symbol probabilities P a i, that means I should know the source model the mutual information. Therefore, depends not only upon the channel matrix P, but also upon how we use the channel that is the probability which we choose the channel inputs.

Now, it is of some interest to examine the variation of I A; B. As we change the input probabilities. So, let us reconsider binary symmetric channel with probability of error P. Now, if you have a binary symmetric channel with probability of P, we have derived that I A semicolon B is given by the expression H of w p plus w bar p bar minus H of p, where w is the probability of selecting 0 at the input and w bar is equal to 1 minus w p bar is equal to 1 minus p. Now, we may plot I A; B as the function of w for a fix p.

(Refer Slide Time: 39:06)

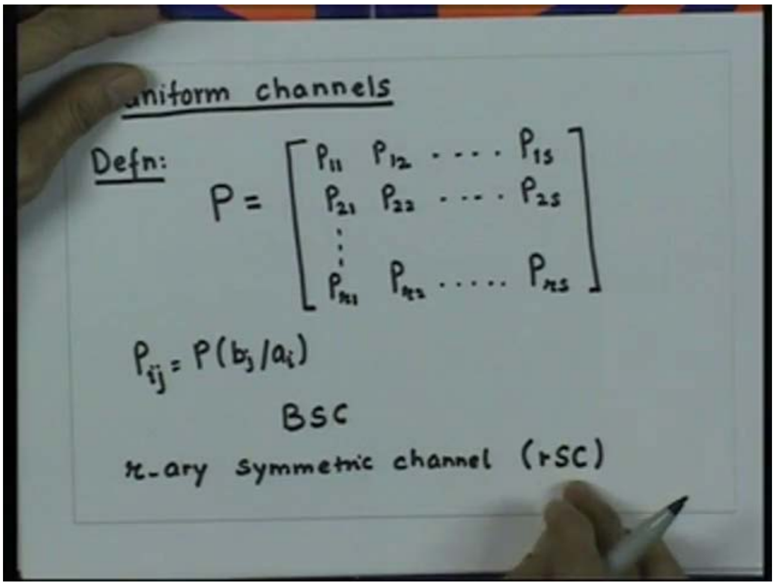


If we do that, we get this figure on the X axis we have the probability of 0 at the input which is indicated by w and on the Y axis, we have mutual information I A semicolon Bin bits per binit. Now, from this figure it is clear that mutual information of a binary symmetric channel where is from 0 to 1 minus H p. The minimum value of 0 is achieved when w is equal to 0 or 1. In this cases the input is known with probability 1 at the output, even before an input value is received maximum value of 1 minus H p is achieved, when omega is equal to half that is when both inputs are equally probable.

Now, for general information channel, we see that the mutual information can always be made 0 by choosing one of the input symbols with probability 1. Since, the mutual information is not negative, this is an easier answer to the question of what is the minimum of I A; B, the question of the maximum value of I A; B for general channel is not so easily answered. The maximum value of I A; B as we vary the input symbol probability is called the channel capacity denoted by C, and is equal to by definition maximum of I A semicolon B over input probability distribution that is P a i.

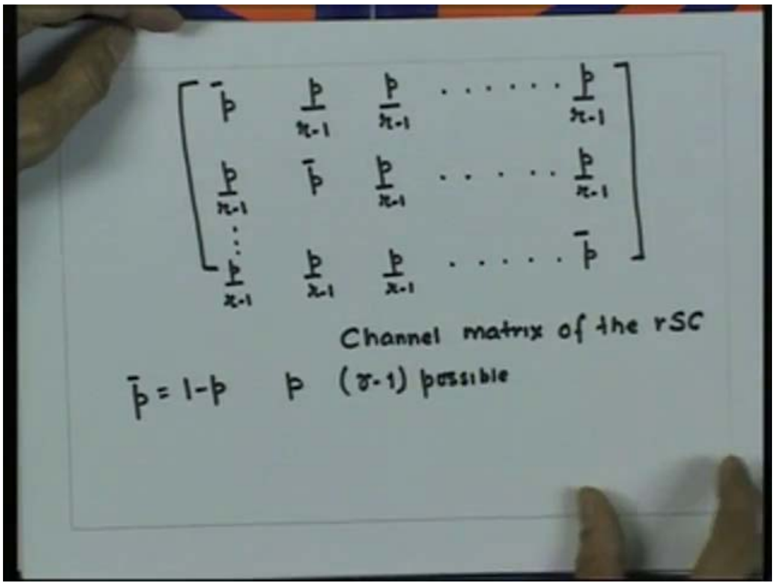
So, it is also important to know that the capacity of an information channel is a function only of the conditional probabilities defining the channel. It does not depend on the input probabilities, that is on how we use the channel. Now, from this figure we see the capacity of the binary symmetric channel with error probability P equal to 1minus H p. Now, the calculation of the capacity of information channel in general is quite involved, but in certain cases however the calculation can be simplified. The most important class of channels for which the channel simplifies.

(Refer Slide Time: 42:55)



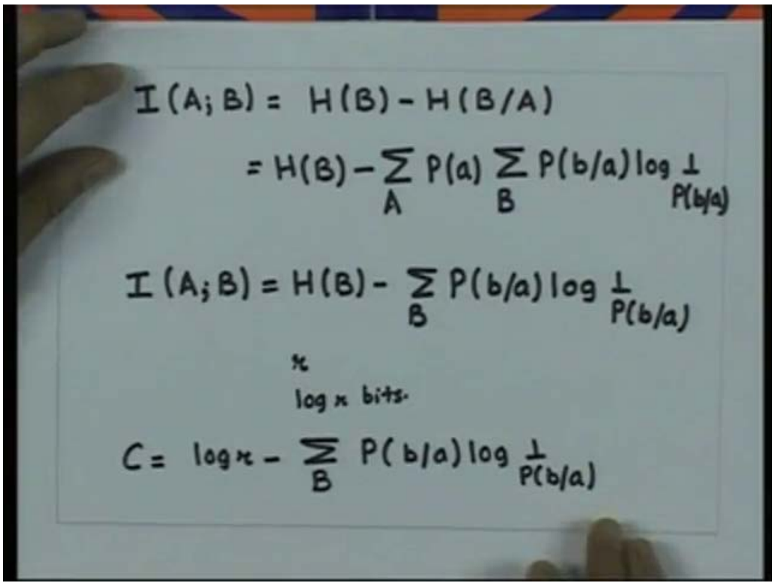
Is the class of uniform channels, so let us define what is a uniform channel consider channel defined by the channel by the matrix P is equal to P 11, P 12 up to P 1s. P 21, P 22 up to P 2s. P r1, P r2, P rs. As before, P ij is equal to probability b j given a i. Now, this channel is said to be in uniform if the terns in every row and every column of the channel matrix consists of arbitrary permutations of the terms in the first row. We have already dealt with example of uniform channel and that is the binary symmetric channel, now the natural generalisation of binary symmetric channel is r-ary symmetric channel, that is known as r SC. Now, r SC is uniform channel with r input and r output symbols. The channel matrix of r SC channel is shown as...

(Refer Slide Time: 45:09)



So, this matrix is of the size r by r, this is a channel matrix of r SC channel. As usual, p bar is equal to 1 minus p. The overall probability of error for this channel is p, but there r minus 1 possible incorrect output symbols for each input symbol. Now, we will calculate the capacity of general uniform channel is the capacity of maximum.

(Refer Slide Time: 46:45)

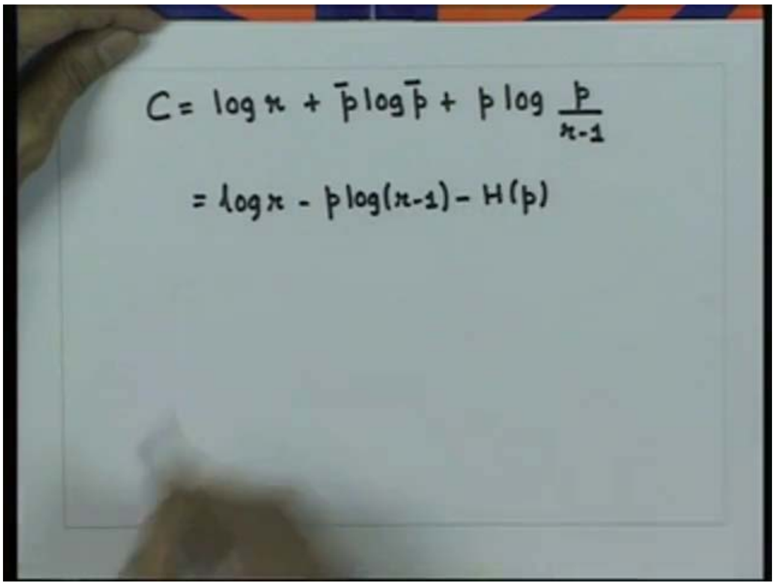


I A semicolon B as we vary the input distribution. I A; B is equal to H of B minus H of B given A, this is equal to H of b minus summation of probability a summed over alphabet A summation over alphabet B probability of b given a log of 1 by probability of b given

a. The summation over B in the last term of this equation is the summation for each a i terms in the higher row of channel matrix for a uniform channel. However, the summation is independent over I and therefore, we can write I of A semicolon B is equal to H of B minus summation of probability b given a log of 1 by probability of b given a of summation b.

Now, last term of the equation is independent of the input symbol distribution, so to find the maximum of the right side of this equation. We need only to find the maximum of H b. Since, the output alphabet consists r symbols, we cannot exceed the log r bits, so H b will equal log r bits, if and only if all the output symbols occur with equal probabilities. Now, in general it is not true that there exists a distribution over the input symbols such that the output symbols are equi-probable, but for the case of a uniform channel however it is easy to check that equi-probable symbols at the inputs produce equi-probable symbols at the output. Therefore, the maximum value of this equation is which is the capacity of the uniform channel, is given by C is equal to log of r minus summation over b probability of b given a log of 1 by probability of b given a.

(Refer Slide Time: 50:02)



Now, using this expression, we can calculate the capacity of the r SC channel as, C is equal to log of r plus p bar log p bar plus p log of p upon r minus 1, which can be simplified as log r minus p times log r minus 1 minus H of p. As said earlier that the calculation of the capacity of an information channel is in general quite involved. In the next class, we will carry out the calculation of capacity for few more channels, one of the important channel being binary eraser channel, we will look at the properties of binary eraser channel and calculate its capacity.