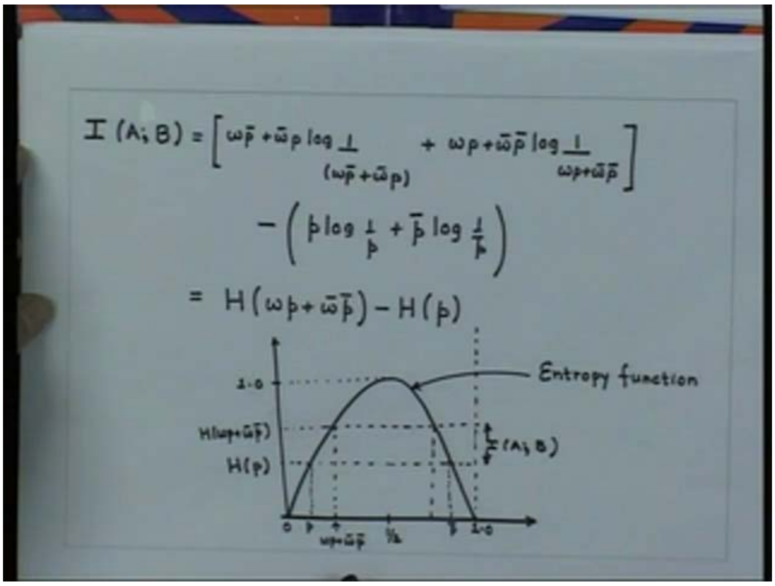
**Information Theory and Coding**
**Prof. S. N. Merchant**
**Department Electrical Engineering**
**Indian institute of Technology, Bombay**

**Lecture - 22**
**Properties of Different Information Channel**

In the last class we defined mutual information. Mutual information is equal to average number of minutes necessary to specify an input signal before receiving an output signal, less the average number of minutes necessary to specify an input signal after receiving an output symbol. This definition of mutual information can also be interpreted as the average uncertainty of an input signal before receiving an output signal, less the average uncertainty of an input signal after receiving an output symbol. This in turn could be interpreted as the uncertainty resolved information received. In the last class, we also took an example of a binary symmetric channel, and we calculated mutual information for this binary symmetric channel, the expression which we had derived is given as shown here.

(Refer Slide Time: 01:58)



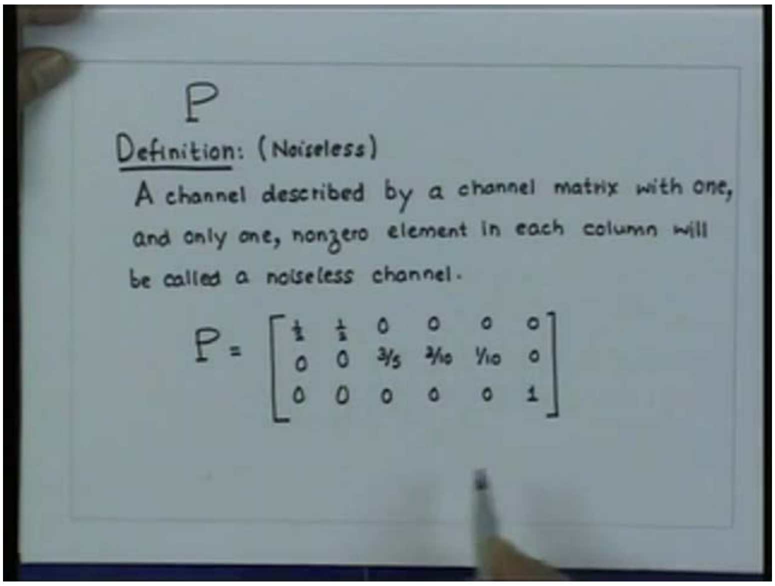

We show that the mutual information I A B is equal to entropy function, where the argument is w p plus w bar p bar minus the entropy function where the argument is p. Now, let us try to provide a geometric interpretation of this result entropy function is depicted by this curve shown here. So, if you choose some value of p where p is probability of error on the binary symmetric channel than corresponding to this pi have p bar which is indicated here corresponding to this p. We have H p as shown on the y-axis here, now omega p plus omega bar p bar is a point which lies between p, and p bar because omega is the probability of occurrence of input symbol 0 which lies between 0 and 1.

So, the point indicated by w p plus w bar p bar is a linear interpolated point between p and p bar. Therefore, from this figure it implies that H of w p plus w bar p bar will be always greater than H p. So, what it implies that this quantity will be always non-negative. So, this is a geometric proof of the nonnegative negativity of mutual information we had proved in the last class that the mutual information is always non negative.

So, if you look on this figure the difference between these two values is the mutual information, now certain limiting conditions of interest may also be seen from this figure. For example, for a fixed value of p, we may vary omega and examine the behavior of I A B, we see that I A B achieves it maximum when omega is equal to half and the value of this maximum will be equal to 1.

And this case the mutual information will be 1 minus H p, this will be the value of mutual information for omega equal to 0 or 1, on the other hand the mutual information is 0. Now, after having looked at definition of mutual information, let us look at some of the information channels. We define two special types of channels and obtain simplified expressions for the mutual information of these channels.

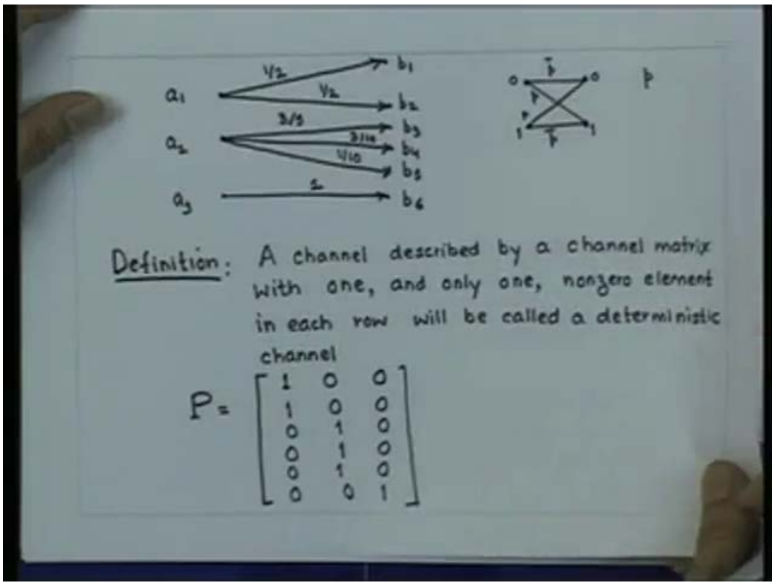
(Refer Slide Time: 05:23)




Now, in our discussion that follows we will assume that each column of the channel matrix, that is P has at least one non zero element. So, what this means that an output symbol corresponding to a column of zeros will occur with probability 0, for any distribution over the input symbols. It is therefore of no interest and maybe ignored, so let us first define a channel which is called as a noiseless channel.

A channel described by a channel metrics P with one and only one, non zero element in each column will be called a noiseless channel. An example of the channel metrics of a noiseless channel is given as P is equal to so each column of this metrics has only one and only one non zero elements. Now, the channel diagram for this noiseless channel is indicated as shown here.

(Refer Slide Time: 08:26)

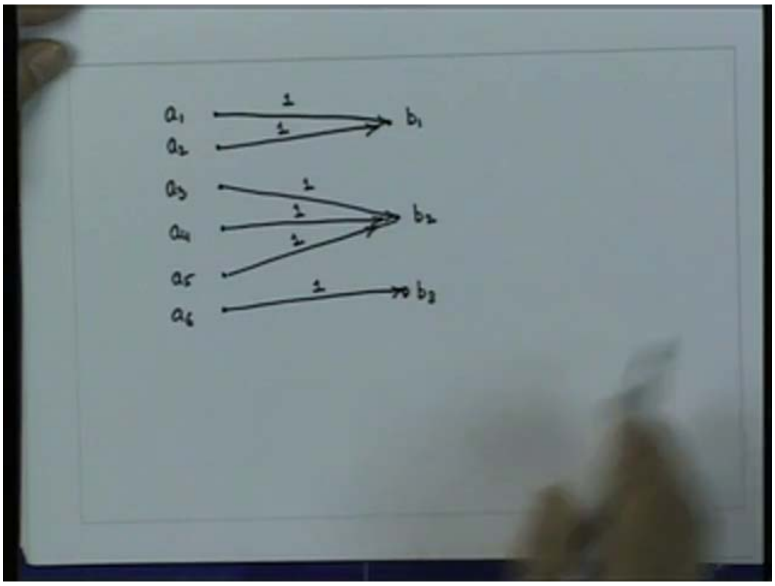


The size of the channel metrics is 3 by 6, so we have three inputs. Let us denote these inputs a 1 a 2 and a 3 and we have six outputs indicated by b 1, b 2, b 3, b 4, b 5 and finally, b 6. So, probability of b 1 given a 1 is half, probability of b 2 given a 1 is also half, probability of b 3 given a 2 is 3 by 5, probability of b 4 given a 2 is 3 by 10 and probability of b 5 given a 2 is 1 by 10 and finally, the probability of b 6 given a 3 is equal to 1.

Now, a binary symmetric channel diagram is as shown here, so binary symmetric channel with the probability of error that is equal to P equal to 0 is an example of a noiseless channel. Note however that a binary symmetric channel with probability of error equal to 1 is also a noiseless channel. So, what it means that a channel which is consistently in error can be as useful as the channel which is consistently correct, so after having define a noiseless channel.

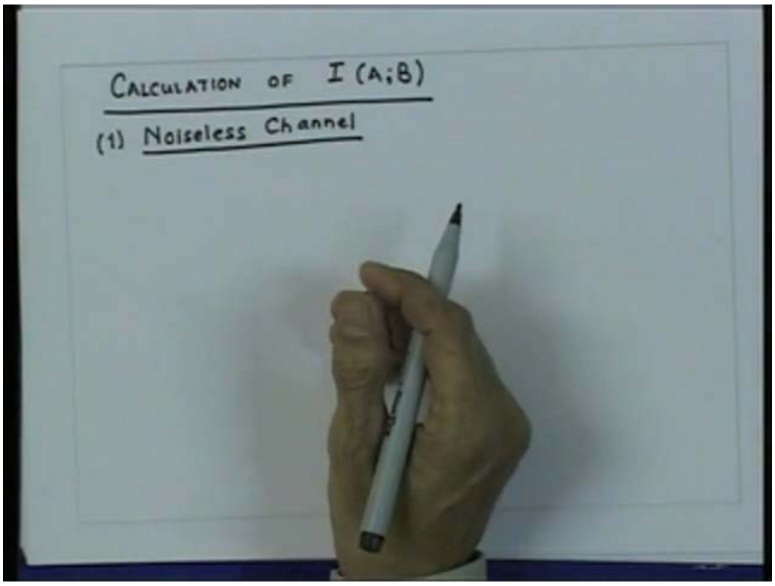
Let us look at the definition of another specific channel and the definition for that is as follows a channel described by a channel metrics is capital P with one and only one non zero element in each row will be called a deterministic channel. An example of a channel matrix for a deterministic channel will be as indicated here. Now, since in this channel matrix there is only one non zero element in each row and since the sum of the elements in each row must be one for a channel matrix what it implies the elements of a deterministic channel matrix are all either 0 or 1.

(Refer Slide Time: 13:36)



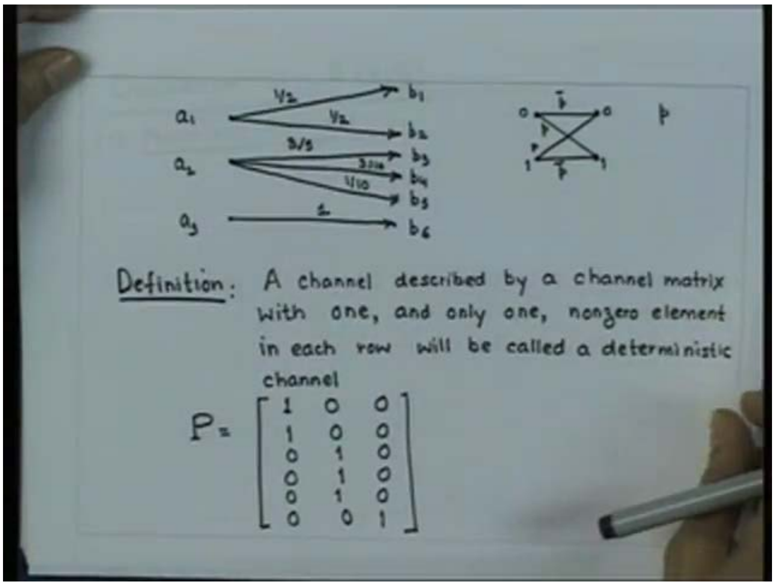

The channel diagram for a deterministic channel will look as follows since the channel matrix sizes 6 by 3 what it implies that we have six inputs which are indicated by a 1, a 2, a 3, a 4, a 5, a 6 and we have three outputs indicated by b 1, b 2, and b 3. Probability of b 1 given a 1 is 1, probability of b 1 given a 2 is also 1, probability of b 2 given a 3 is 1, probability of b 2 given a 4 is 1 this is also 1 and finally, probability of b 3 given a 6 is equal to 1, so this is a channel diagram for a deterministic channel. Now, let us carry out the calculation of mutual information for these two types of channels which we have just defined.

(Refer Slide Time: 14:58)



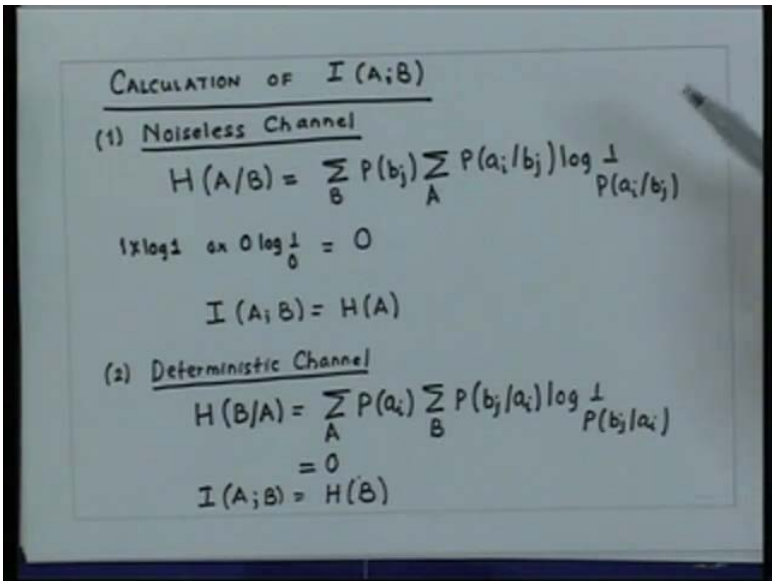
So, let us first consider a noiseless channel, now in a noiseless channel an example of which we saw earlier of this form.

(Refer Slide Time: 15:42)




When we observe the output b j, we know with probability 1 which a i was transmitted that is the conditional probabilities of P a i given b j are all either 0 or 1.

(Refer Slide Time: 16:12)



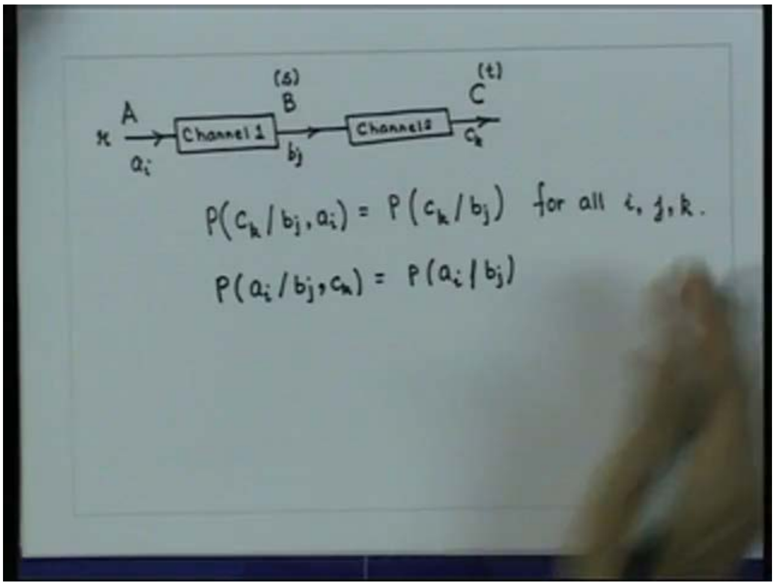

So, using this fact we can write the equivocation that is defined as H of A given B, which is equal to probability of b j summation probability a i given b j log of 1 of a i b j. Now, since each of this term in the inner summation being of the form 1 multiplied by log 1, if this is equal to one 0 multiplied by log 1 by 0, this quantity 0 what it implies that H of A given will be equal to 0.

Now, this conclusion is also evident in view of the generation of Shannon’s first theorem, the outputs of a noiseless channel are sufficient by themselves to specify the inputs to the channel. Hence, the average number of minutes necessary to specify the inputs when we know the output is obviously 0 and therefore, I A B which is equal to H A minus H of A given B will be equal to H of A.

So, the amount of information transmitted through such a noiseless channel is equal to the total uncertainty of the input alphabet, so we get complete information. Now, let us look at the calculation of I A B for deterministic channel, now example of the deterministic channel is shown here. So, for deterministic channel the input symbol a i is sufficient to determine the output symbol b j with probability 1, so what it implies that all the probabilities probability b j given a i are all either 0 or 1. Therefore, for this case H of B given A which is of the form summation p a i, A B summation over b probability b j given a i log of b j given a i.

Now, since each of this is of form 0 and 1, what it implies again H of B given A is equal to 0 and for deterministic channel I get mutual information equal to H of B that is the entropy of the output alphabet. Now, having look at the definition of mutual information and some of the specific information channel, it would be interesting to see what happens when we cascade two information channels. So, let us look into some of the interesting properties of entropy and mutual information by considering the cascades of two channels, so will assume that we have two channels which have been cascaded as shown in this figure.

(Refer Slide Time: 20:39)

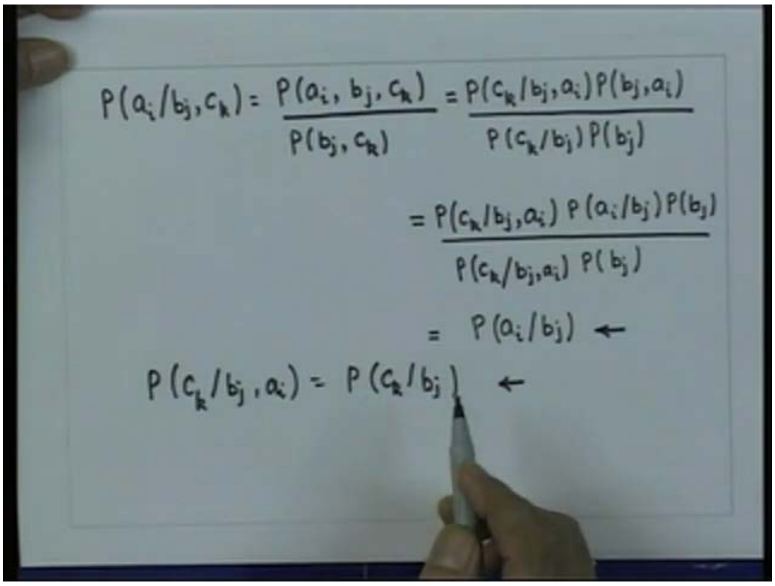



I have channel 1 which is cascaded with channel 2, the input alphabet for the channel 1 is denoted by A and the size of this alphabet is r, the output alphabet of channel 1 is denoted by B and the size of which is s. Now, the input alphabet of the second channel is identified with the output alphabet of channel 1, so the input alphabet for channel 2 would be equal to B and the output alphabet for channel 2 is denoted by C with the size of the alphabet being t. Now, from this figure, it is clear that alphabets are connected by the cascades as shown in this figure, so what it implies that certain relationship among the symbol probabilities will exist.

So, let us look into those relationship when a i assembled from A is transmitted along the channel 1 the output of the first channel is say b j in turn b j produce an output c k from the second channel. The symbol c k depends on the original input a i only through b j what it means that, if you know the intermediate symbol b j the probability of obtaining the terminal symbol c k depends only upon b j and not upon the initial symbol a i which produced b j.

So, this property of cascaded channels may be written as probability of c k given b j and a i is equal to probability of c k given b j for all i j k, in fact this equation may be taken as a definition of what we mean by the cascade of two channels. Now, a direct application of Bayes rule which we studied in the earlier class yields a similar equation in the reverse direction, so we can show that probability of a i given b j, c k is equal to probability of a i given b j. Let us try to prove this relationship so it is not very difficult to prove that.

(Refer Slide Time: 24:24)

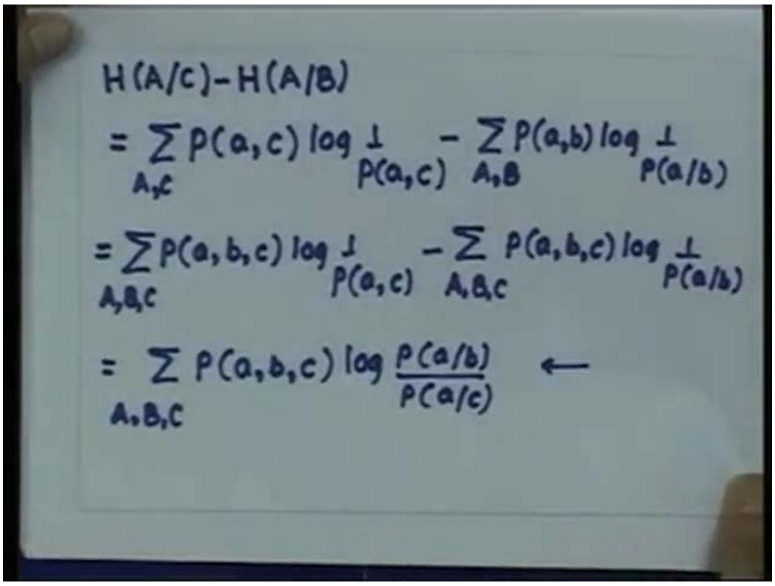



Probability of a i given b j c k can be written as by Bayes rule probability of joint event, a I b j c k divided by probability of b j c k provided probability of b j c k is not equal to 0. Now, this can be written as probability of c k given b j a i multiplied by probability of b j a i probability of b j c k is probability of c k given b j multiplied by probability of b j. Now, this can be further simplified as probability of c k given b j a i, this is equal to probability of a i given b j multiplied by probability of b j.

Now, probability of c k given b j is equal to probability of c k given b j a i because a i, b j, c k is are the inputs and the outputs of the two cascaded channels. So, we have just seen the definition of cascaded channel is that probability of c k given b j a i is equal to probability of c k given b j. So, this quantity is substituted by this and we write this as probability b j and these terms cancel, so what we get finally is probability a i given b j, so we have shown that for cascaded channel this condition is true.

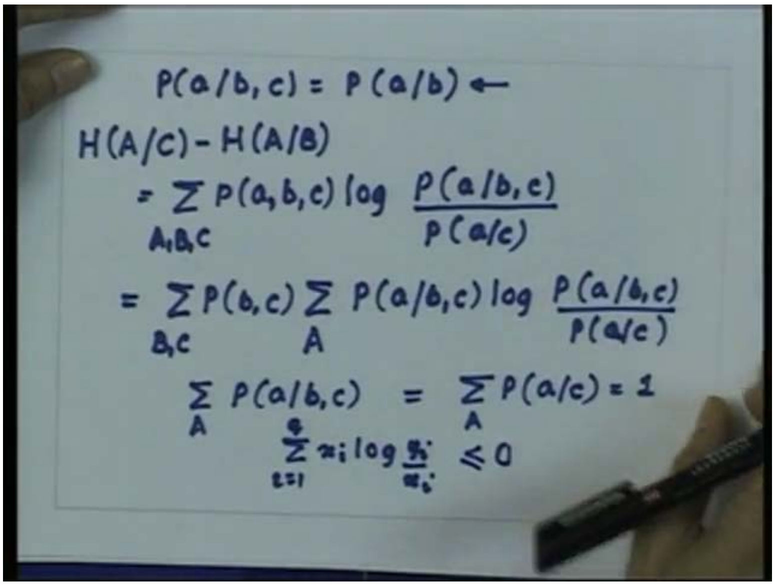
Now, it should be emphasized that probability of a i given b j c k is equal to this and probability of c k given b j a i is equal to probability of c k given b j. These two relationships are only valid for the case where a b c are the alphabets of cascaded channels. Now, as we transmit information through cascaded channels from A to B to C, it is expected that the equivocation that is H of A given C should be greater than the equivocation H of A given B.

(Refer Slide Time: 28:03)




Let us investigate this question, so H of A given C minus H of A given B is equal to probability of a, c log of 1 by probability a, c minus probability of a, b log of 1 by probability a given b this summations are double summation over the alphabet A C this is a double summation over alphabet A B. Now, for convenience we will draw the subscripts a i, b j, c k whenever we are summing over the symbols of the alphabet A B and C. Now, this can be rewritten as probability of a b c over input alphabet a b c log of p a c minus a b c probability of a b c log of 1 by probability of a given b. This can be simplified as probability of a b c summed over the alphabets A B C log of P a given b divided by P a given c.

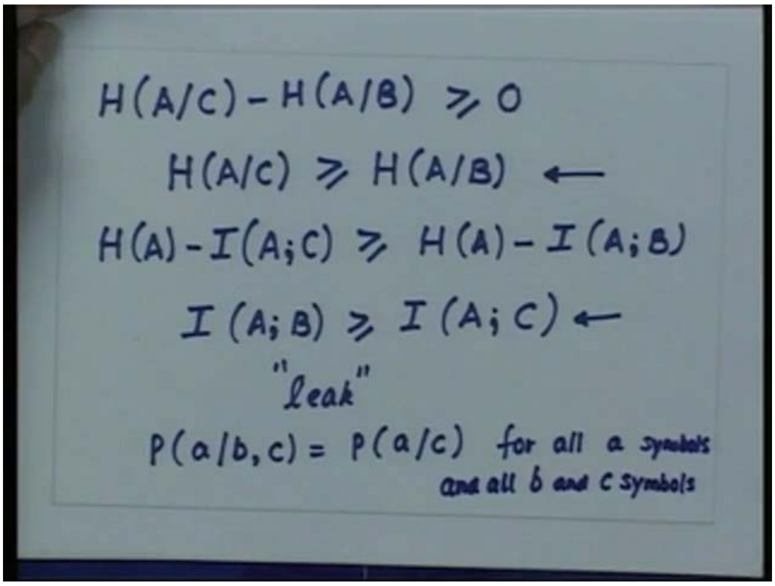
(Refer Slide Time: 30:56)



Now, in this expression we use the relationship which we just derived that is probability of a given b, c is equal to probability of a given b. If we use this relationship in the last equation then we can write as H of A given C minus H of A given B equal to probability of a b c log of probability a given b, c divided by probability a c and we get the relationship finally, as probability of a given c summed over A.

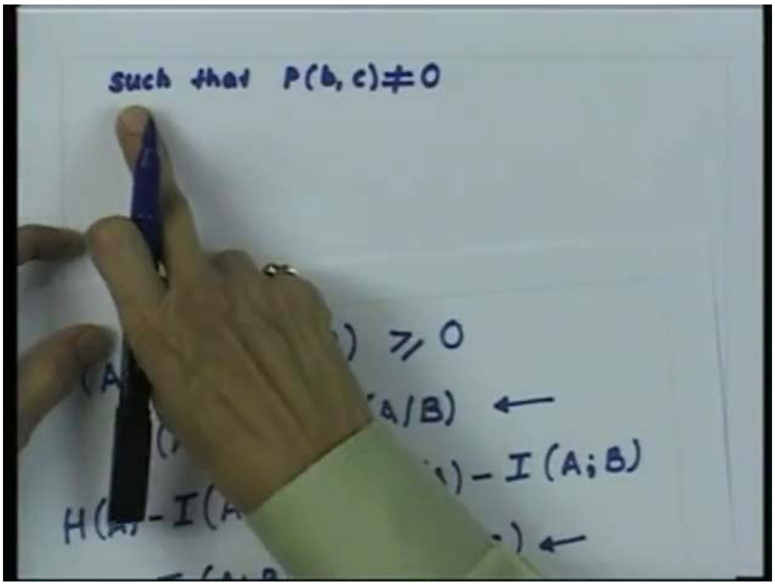
Now, summation of probability a given b, c over alphabet A and the summation of probability a given c over alphabet A both is equal to 1. So, for this term out here we can use the earlier relationship which we had derived that is x i log of y i x i summed over i equal to 1 to q is less than or equal to 0. So, using this relationship this quantity out here is always greater than equal to 0.

(Refer Slide Time: 33:44)




So, what we get finally is H of A given C minus H of A given B is greater than equal to 0, which implies that H of A given C is greater than equal to H of A given B. This quantity can be written as H A minus I of A C greater than equal to H A minus I of A B this implies that I of A B is always greater than equal to I of A C. So, this shows that the information channels tend to leak information, what it implies that the information that finally comes out of cascaded channels can be no greater than information, which would emerge from an intermediate point in the cascade, if you could trap such a point. The condition for equality to hold in this equation is of some interest. So, retracing the steps in our proof to arrive at this equation; we see that the equality holds if and only if probability of a given b, c is equal to probability of a given c.

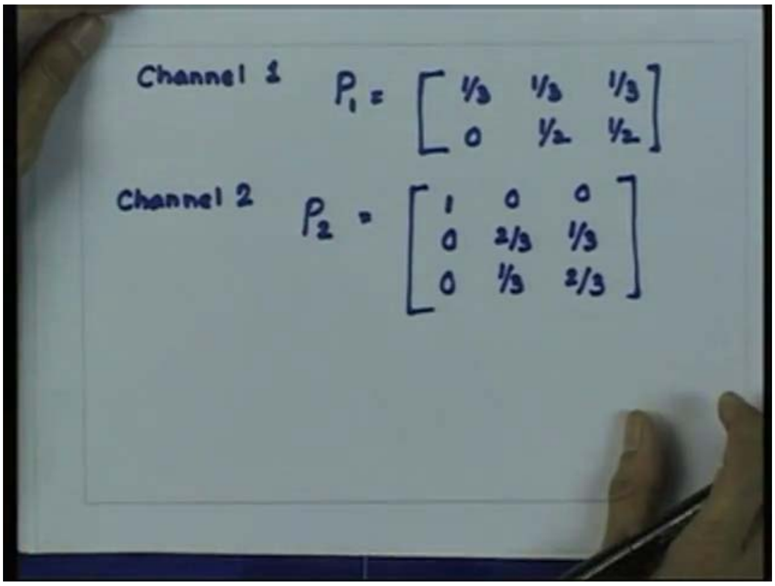
(Refer Slide Time: 36:36)




So, this is should be true for all a symbols and all b and c symbols such that probability of b, c is not equal to 0. So, b c not equal to 0, now the condition for equality deserve some comment at first glance, it might appear that equality would apply if and only if the second channel in the cascade is noiseless. Now, it is not very difficult to show that for noiseless channel this condition is valid along with this condition.

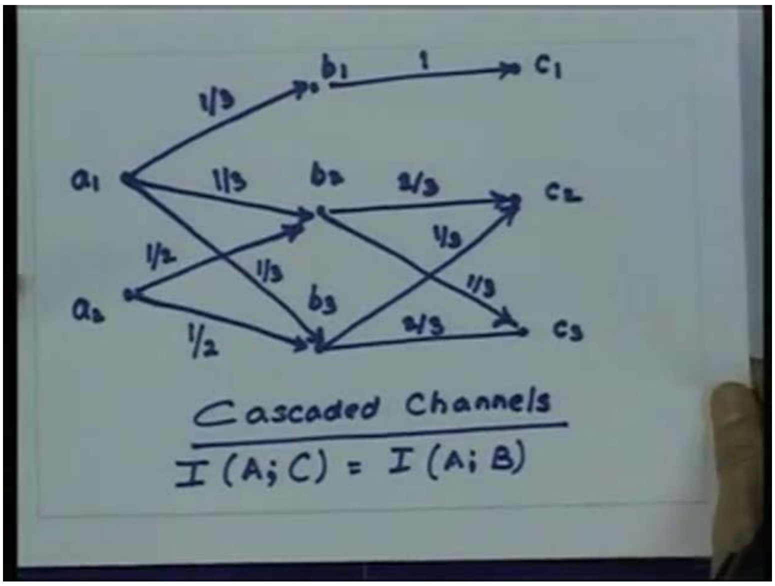
So, for noiseless channel you can ensure that this condition is always valid in which case you get the equality, but other than noiseless channel also it is possible that this equality is satisfied in which case H of A given C would be equal to H of A given B, which in turn would imply that mutual information that is I of A B will be equal to mutual information I A C. So, let us look at one example which illustrates this point.

(Refer Slide Time: 38:12)



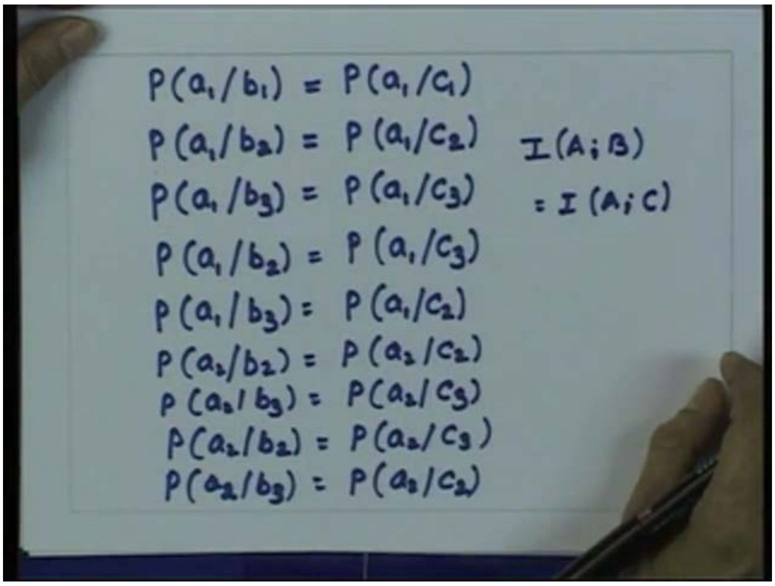

So, let us take a channel 1 with the channel matrix given by P 1 equal to one-third, one-third, one-third, 0, half, half. So, this channel has two inputs and three outputs and this channel is cascaded with another channel which is indicated by channel 2 with the channel matrix given by P 2 equal to 1, 0, 0, 0, two-third, one-third, 0, one-third, two-third. So, when we cascade channel 1 with channel 2, we can get the channel diagram for the cascaded channels as indicated here.

(Refer Slide Time: 39:22)



For the channel 1 we have, so this is a 1, a 2, b 1, b 2, b 3 this is one-third probability of b 1 given a 1, probability of b 2 given a 1 is also one-third and probability of b 3 given a 1 is also one-third, probability of b 2 given a 2 is equal to half and probability of b 3 given a 2 is half. Now, this channel is cascaded with another channel, so the channel diagram follows this is c 1, c 2, c 3 this is equal to 1, probability of c 2 given b 2 is two-third, this is one-third, this is one-third and finally, this is two-third. So, this is a channel diagram for the cascaded channels, now for this channel if we wish that I A C is equal to I A B then the recently derived condition should be valid.

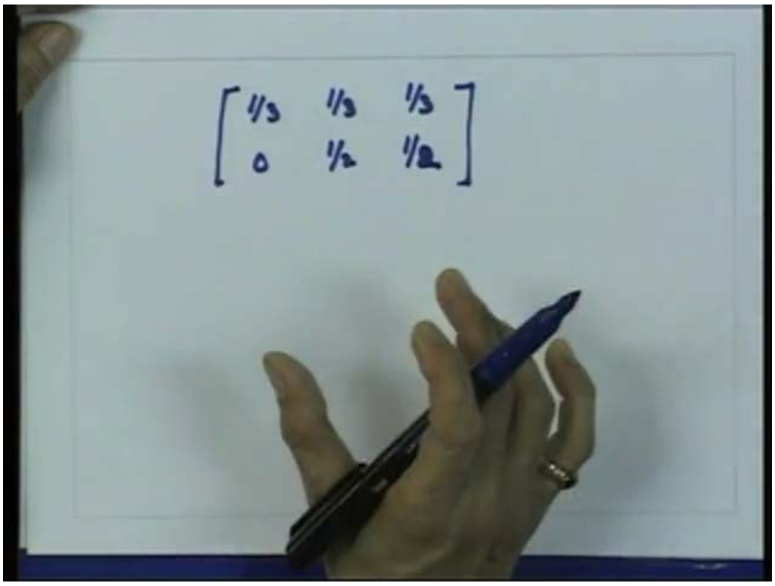
(Refer Slide Time: 42:01)




For this example what we are supposed to show is that probability of a 1 given b 1 is equal to probability of a 1 given c 1. Probability of a 1 given b 2 is equal to probability of a 1 given c 2 probability of a 1 given b 3 is equal to probability of a 1 given c 3. Probability of a 1 given b 2 is equal to probability of a 1 given c 3, probability of a 1 given b 3 is equal to probability of a 1 given c 2. And probability of a 2 given b 2 is equal to probability of a 2 given c 2 probability of a 2 given b 3 is equal to probability of a 2 given c 3. Probability of a 2 given b 2 is equal to probability of a 2 given c 3 and finally, probability of a 2 given b 3 is equal to probability of a 2 given c 2.

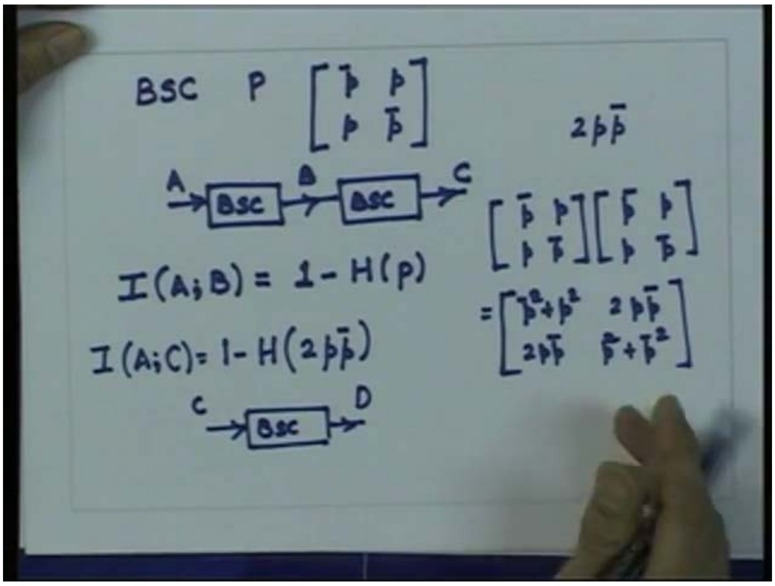
So, if we want that a mutual information of the cascaded channel I A C should be equal to I A B, then all this conditional probabilities should be equal. We have taken all those cases for which probability b c is not equal to 0. Now, for this example it is easy to show that this is true in which case we get I of A B is equal to I A C. Now, to calculate probability of a given c, we would require to calculate the backward forward probabilities b of c given a, and also probability c.

(Refer Slide Time: 45:07)




Now, this can be easily done for this example and it can be shown that the equivalent channel matrix between A and C is of the form, so probability of c 1 given a 1 is this, probability of c 2 given a 1 is one-third. Probability of c 3 given a 1 is again one-third and probability of c 1 given a 2 is 0. Probability of c 2 given a 2 is half and probability of c 3 given a 2 is half. So, we can use this relationship this equivalent channel matrix to find out the backward probabilities and probabilities of c and we can show that this relationship is valid. Now let us look take an example of a 2 binary symmetric channels and cascade them.

(Refer Slide Time: 46:15)



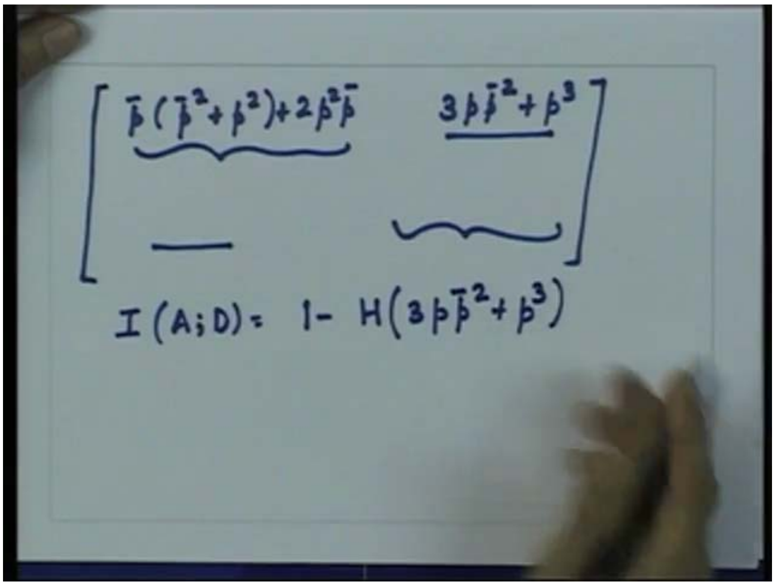

So, let us assume that the binary symmetric channel is given with the channel matrix as follows, we take two identical binary symmetric channels and cascade as indicated here. Let us try to evaluate the mutual information between A and C and also between A and B. We will assume that the two possible inputs of the first binary symmetric channel are chosen with equal probability

So, we have seen that I of A B is equal to 1 minus H p because if omega is equal to half, then H of omega p plus omega bar p bar is equal to 1 in which case we get this relationship. Now, if we cascade the two channels than the equivalent channel matrix between A and C can be obtained. It is not difficult to show that for the cascade of these two binary symmetric channel, it is equivalent to a single binary symmetric channel with probability of error as 2 p p bar.

So, if we cascade we will get the equivalent binary symmetric channel as follows p square plus p bar square, so this is the equivalent channel matrix between A and C. So, for this based on a single binary symmetric channel results we can find out I of A C is equal to 1 minus H of probability of error in this case is 2 p p bar. Now, we can extend this result to three binary symmetric channels.

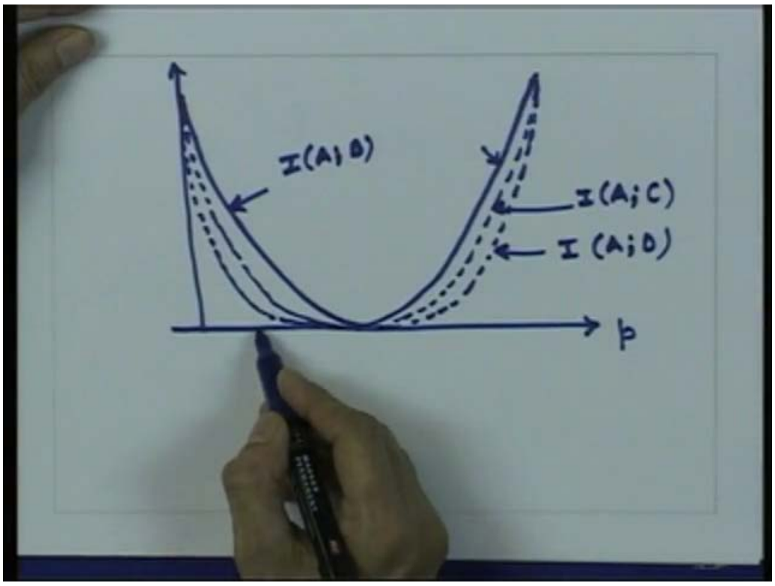
So, if we take this and cascaded it with another binary symmetric channel, where the output of the second binary channel that is C is fed to and we get an output final output is D. Now, for this case we can obtain the equivalent binary symmetric channel between A and D by taking this matrix and multiplying by this matrix to get the final result which is indicated here.

(Refer Slide Time: 50:17)




P bar p bar square plus p square plus 2 p square p bar, 3 p p bar square plus this quantity is repeated here and this quantity will be repeated here. So, for this equivalent channel matrix the probability of error now is given by the term here. So, for this between A to D, we can find out the mutual information as 1 minus H of 3 p p bar square plus p cube.

(Refer Slide Time: 51:22)



Now, if we plot the mutual information for all this three cases, we will get a result something of this form this is I A B, this is I A C and finally, we have I A D indicated here. So, on the x axis we have p and on the y axis we have the mutual information, this is I A B, this is I A C and finally, this is I A D. So, from this it is clear that for given probability of error p, I A D is less than I A B in many types of information channels encountered in real life situation, the set of channel output is far larger than the user would.

So, the question is, is it possible to reduce the number of outputs by cascading the output of this channel to with another channel, where the output size of this new channel is less than the input alphabet size? If you are able to achieve this then is it possible to simplify the channel without paying a penalty in reducing mutual information. This question we will examine in the next class.