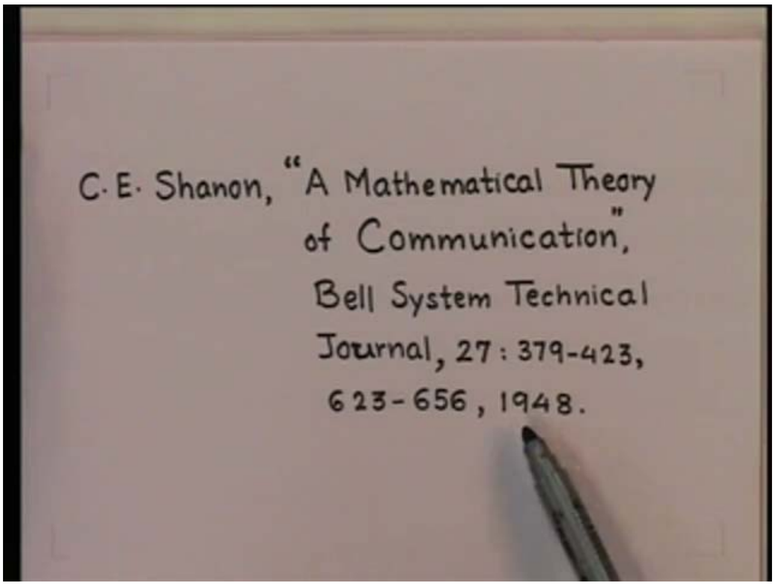
Information Theory and Coding

Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 33
Calculation of Rate - Distortion Functions

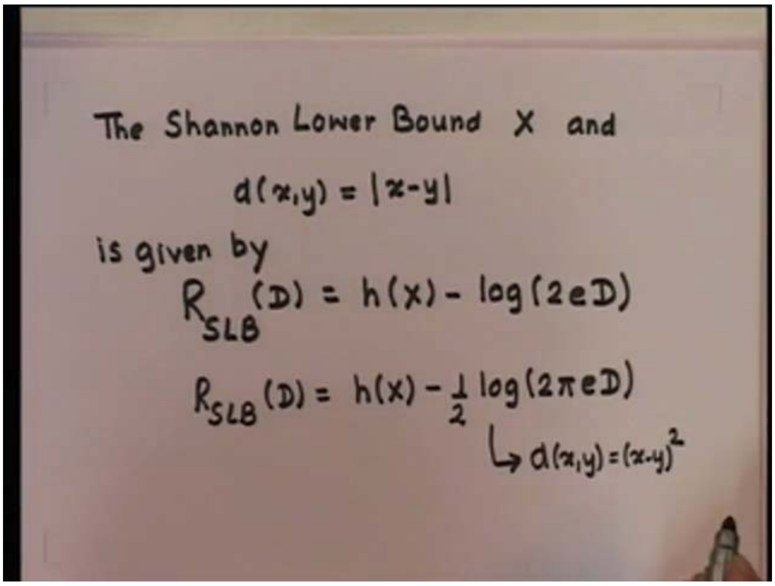
Based on the mathematical definition of the rate distortion function, we have shown that like the differential entropy for a Gaussian source. The rate distortion function for a Gaussian source also has a distinction of being larger than the rate distortion function, for any other source with continuous distribution and the same variance. Now, this is a very important result, because for many sources it can be very difficult to calculate the rate distortion functions. In such situations it is helpful to have an upper bound for the rate distortion function. Now, it will be very useful if you could also find the lower bound for the rate distortion function. Shannon, in his 1948 paper has described this lower bound and appropriately called Shannon's lower bound.

(Refer Slide Time: 02:28)



The references, C. E. Shannon, a Mathematical Theory of Communication, Bell system technical journal, volume 27. There were 2 parts and published in 1948. We will state the lower bound here without derivation, if 1 is interested in detailed derivation then it could be referred to this reference. T. Berger, rate distortion theory a mathematical basis for data compression, published by Prentice Hall in 1971.

(Refer Slide Time: 03:15)



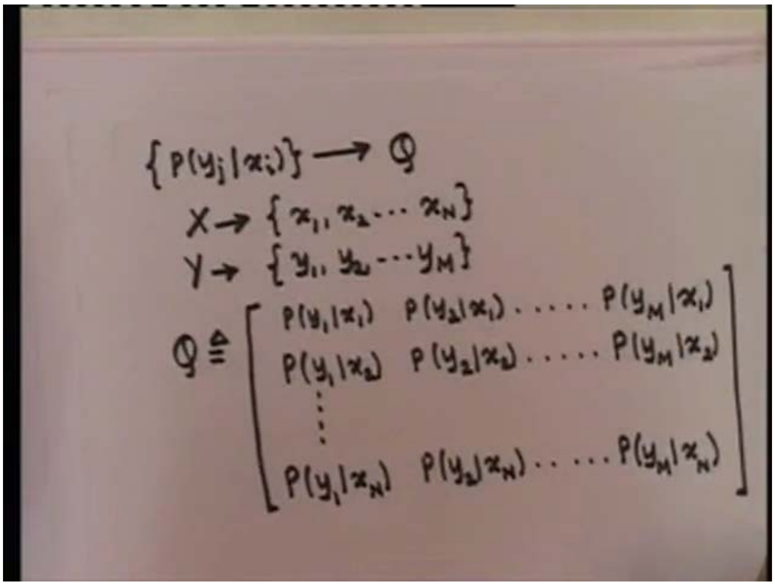
So, the Shannon's lower bound for a random variable X and the magnitude error criterion, that is D x y equal to the mod of the difference between x and y is given by Shannon lower bound is equal to differential entropy minus log of 2 e D, where D denotes the specified distortion level. Now, if we use this squared error criteria then Shannon's lower bound is given by differential entropy minus half log 2 pi e D.

So, this result is for squared error criterion that is D x y is equal to x minus y squared. So, to summarize we have provided a mathematical definition for the rate distortion function and also calculated the same for 2 important sources, 1 binary source and the other was the Gaussian source. We have also obtained the upper bounds and lower bounds for any arbitrary independent identically distributed source. Now, these functions and results are specially useful when we want to know if it is possible to design compression schemes to provide the specified rate, and distortion for a given source.

These results are also useful in determining the amount of performance improvement that we could obtain by designing better compression schemes. Therefore, the rate distortion function plays the same role for lossy compression scheme that entropy plays for lossless compression. Now, in the last class we had discussed 2 approaches to calculate the rate distortion function for a source.

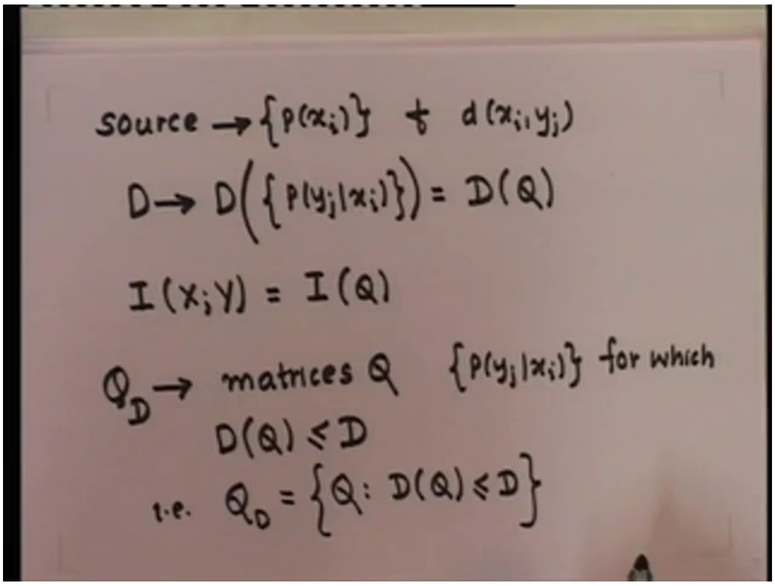
One approach which we utilize to calculate the ray distortion function for the binary source and the Gaussian source was to calculate the lower bound, for the average mutual information and then show that this bound is achievable. The other approach is based on a computational approach. Now, we will not go into the detail derivation of this approach, but we will present the salient features of this approach. The algorithm by itself is very simple. Now, in order to do that we will use simplified notations as follows.

(Refer Slide Time: 07:13)



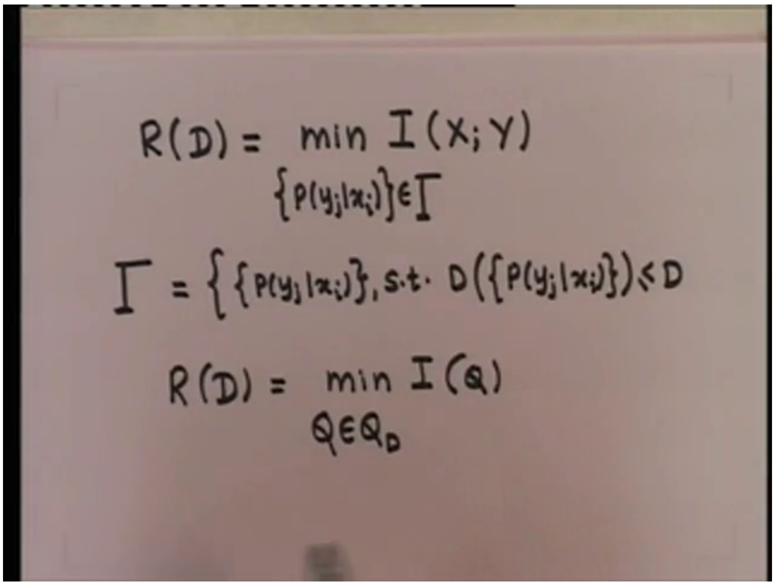
Conditional probabilities that is probability of y j given x i by a matrix Q, where we assume that the source alphabet is of size N, that is there are N elements and reconstruction alphabet is of size M. So, in this case Q matrix by definition is given by probability of y 1 given x 1, probability of y 2 given x 1 to probability of y M given x 1, probability of y 1 given x 2, probability of y 2 given x 2, probability of y M given x 2. And finally, the last row is probability of y 1 given x N, probability of y 2 given x N and the last term is probability of y M given x N. Now, using this notation, we can write the average distortion for a given source and specified distortion measure as the function of the matrix Q as follows.

(Refer Slide Time: 09:19)



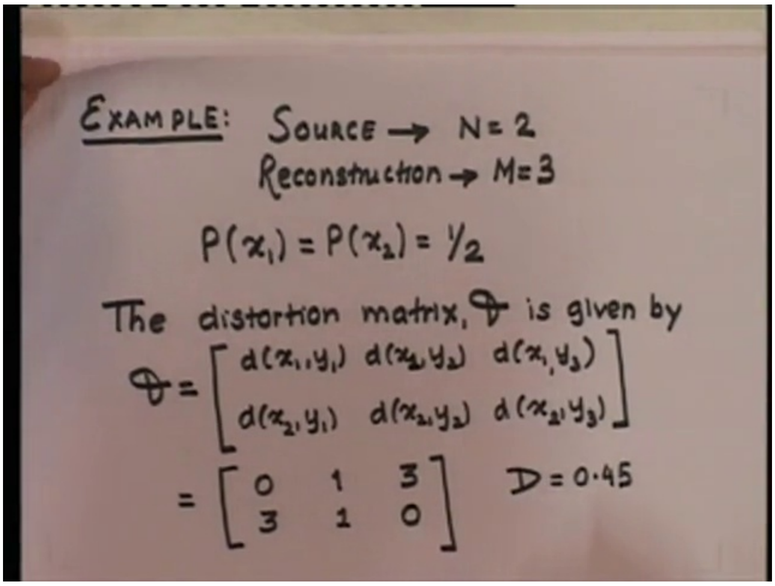
Distortion D, which we have seen is basically function of probabilities, can be written as D function of Q and similarly, the average mutual information which is again function of conditional probabilities can be written as the function of the matrix Q. Now, we introduce the set Q D which consists of the matrices Q of the conditional probabilities, for which distortion as the function of Q is less than or equal to some specified distortion level D, that is Q D is equal to set of matrices Q will satisfy the distortion criterion.

(Refer Slide Time: 11:46)



Now, to recollect the mathematical definition for the rate distortion function R D is equal to minimum of I X Y. This minimum is over conditional probabilities belonging to the set where the set is equal to set of conditional probabilities is such that distortion criterion is satisfied. So, based on our new notation, we can write the rate distortion function as R D is equal to minimum of I Q, where Q belongs to the set Q D. Now, based on this new notation, let us solve some few examples in order to understand these concepts in much better manner.

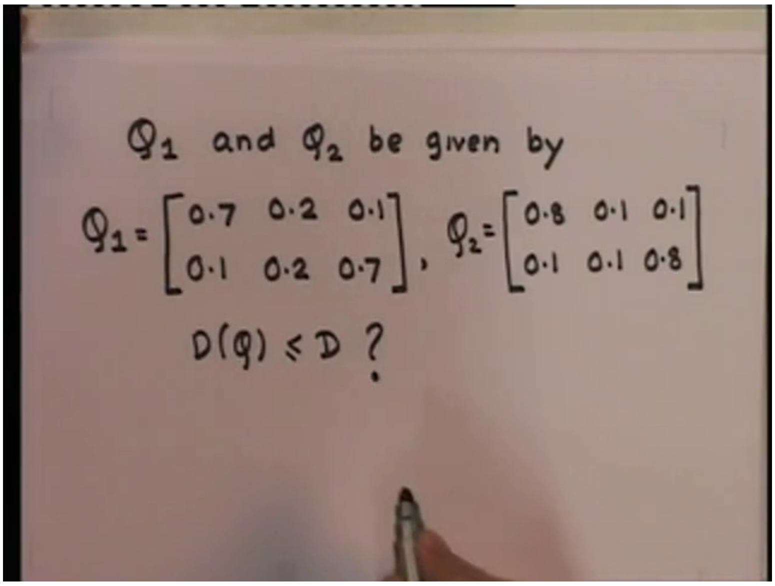
(Refer Slide Time: 13:28)



So, let us assume that we have a source alphabet of size N equal to 2 and reconstruction alphabet of size 3 that is M is equal to 3. Let us assume that the probabilities of the source symbols that is P of x 1, equal to P of x 2 is equal to half. Let us specify the distortions between a source symbol and a reconstruction symbol in the form of a distortion matrix given by. This is by definition, D x 1 y 1, D x 1 y 2, D x 1 y 3, D x 2 y 1, D x 2 y 2.

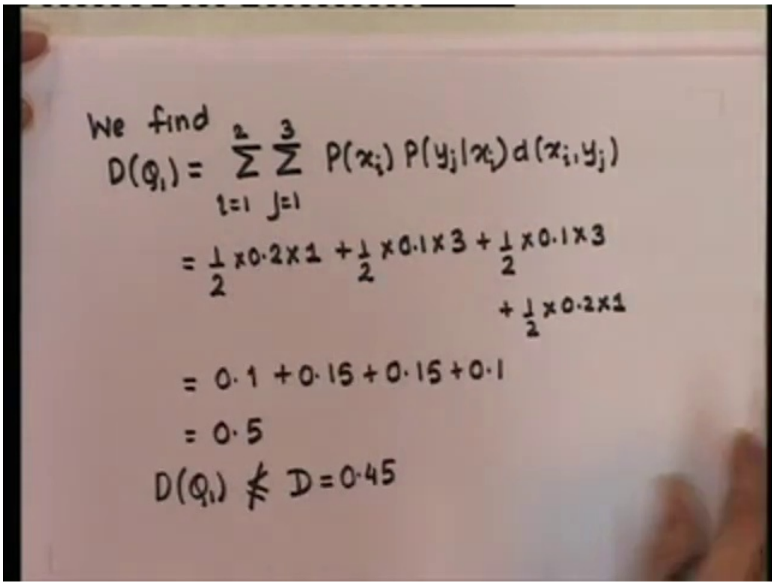
And finally, D x 2 y 3 and this let us assume to be given as 013310. What does distortion matrix means that there is no distortion if source symbol x 1 is encoded into symbol y 1 and x 2 is encoded into y 3. Otherwise, there is distortion and let us assume that the average permissible symbol distortion be less than equal to D equal to 0.5. Now, let the two conditional probability matrices Q 1.

(Refer Slide Time: 16:34)



And Q 2 be given as follows 0.7 0.2 0.1 0.1 0.2 0.7 and Q 2 is equal to 0.8 0.1 0.1 0.1 0.1 0.8. So, the sum of the row elements is always equal to 1. Now, for both these matrices, we may confirm whether or not they give an average symbol distortion satisfying the given constrain, that is Q of Q is less than or equal to D this is what we want to verify.

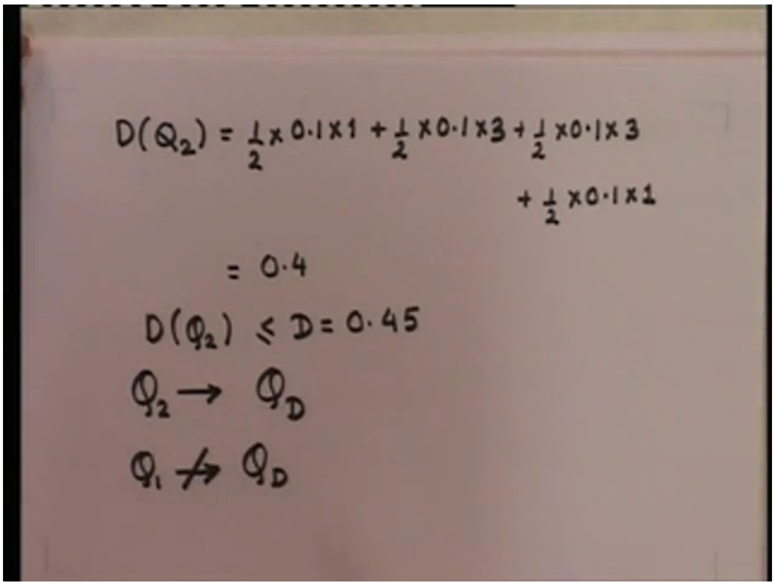
(Refer Slide Time: 17:54)



So, we find D Q 1 as follows. Double summation i is equal to 1 to 2, j is equal to 1 to 3, probability of x i, conditional probability of y j given x i and the distortion measure between x i and y j. Now, let us plug in the values which have been provided. This is equal to half multiplied by 0.2 multiplied by 1 plus half multiplied by 0.1 into 3 plus half multiplied by 0.1 into 3 plus half into 0.2 multiplied by 1.

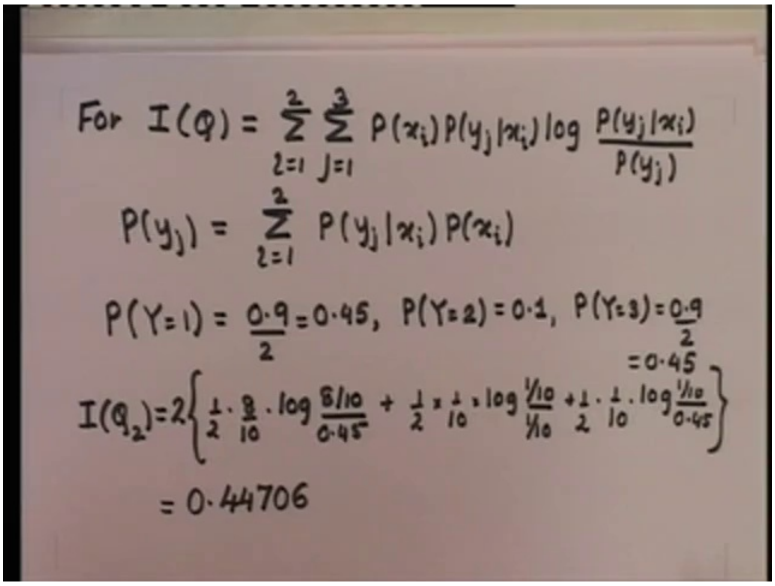
Other two terms are 0 because D x 1 y 1 is equal to 0 and D x 2 y (Refer Time 19:27) and this can be simplified as 0.1 plus 0.15, 0.15 plus 0.1 is equal to 0.5. This clearly shows that D of Q 1 is not less than the specified distortion level that is D equal to 0.5. Now, let us similarly, calculate the distortion for the given matrix Q 2.

(Refer Slide Time: 20:25)



If you plug in the values similarly to what we have done here then we can show that D of Q 2 is equal to half multiplied by 0.1 multiplied by 1 plus half multiplied by 0.1 multiplied by 3 plus half multiplied by 0.1 into 3 plus half into 0.1 multiplied by 1, and this is equal to 0.4. Clearly, D Q 2 satisfies the constrain. Therefore, the matrix Q 2 belongs to Q D as opposed to Q 1 which does not belong to Q d. Now, we can calculate the average mutual information for all those matrices which belong to Q D and select the lowest 1 of this. Now, for the given example, we have only 1 matrix that is Q 2 that belongs to the set Q d. So, let us calculate I of Q 2.

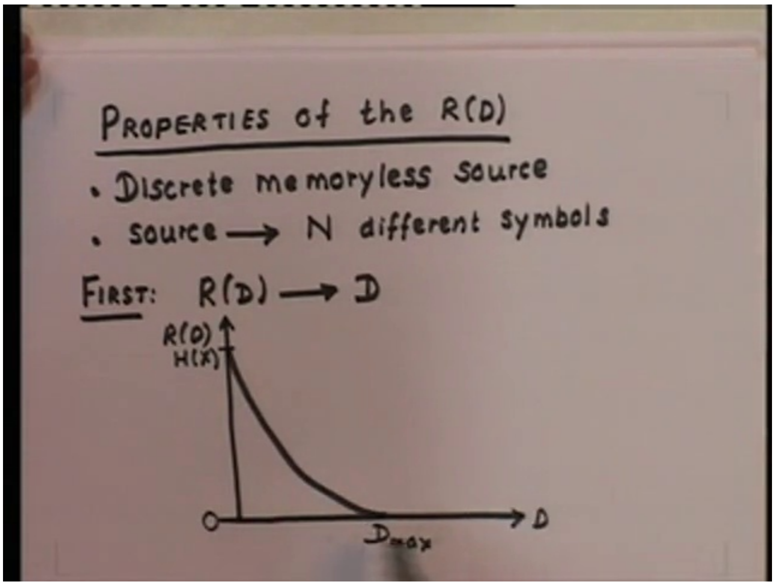
(Refer Slide Time: 22:02)



For I Q calculation, we use the following relationship. Probability of x i, probability of y j given x i, log of probability y j give x i divided by probability of y j and we also know that probability of y j is equal to summation of over i equal to 1 to 2 of probability y j given x i multiplied by probability of x i. So, we calculate I Q 2 for Q 2 since it belongs to Q D and probability of y equal to 1 from the given values can be evaluated as follows.

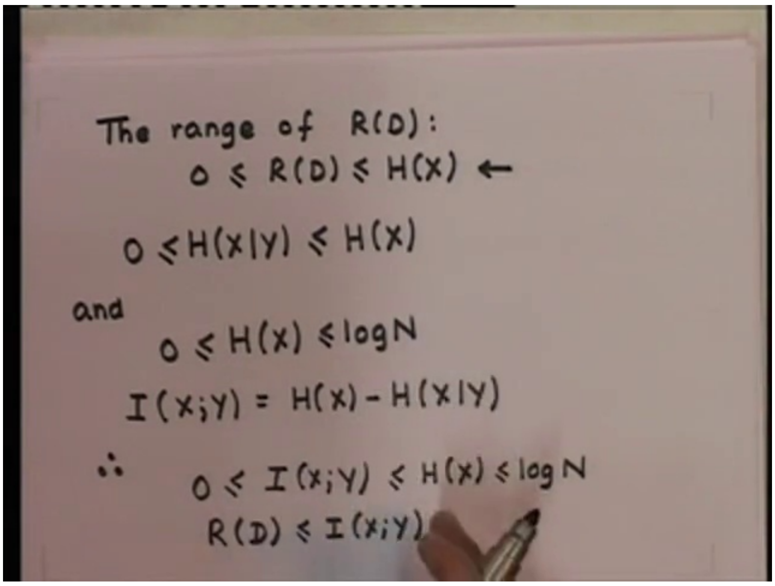
Probability of y equal to 2 is equal to 0.1 and probability of y equal to 3 is equal to 0.9 by 2 is equal to 0.45. And if we use this result and plug in for the relationship of I Q 2 we get the following result plus half into 110 th into log of 110 th by 110 th plus half to 110 th log of 110 th divided by 0.45 and this on simplification is equal to 0.44706. So, having solved this example, now let us proceed and try to look in general what are the properties of the rate distortion function.

(Refer Slide Time: 25:21)



So, we will investigate the properties of the rate distortion function. We will assume that we have discrete memory less information source and let us assume that the source has N different symbols. Now, first let us see how the rate distortion curve proceeds as a function of distortion level D. Now, we can show based on what we derive from the binary source and the Gaussian source that in general that the rate distortion function has the following shape. So, it is a monotonically decreasing function with increasing D.

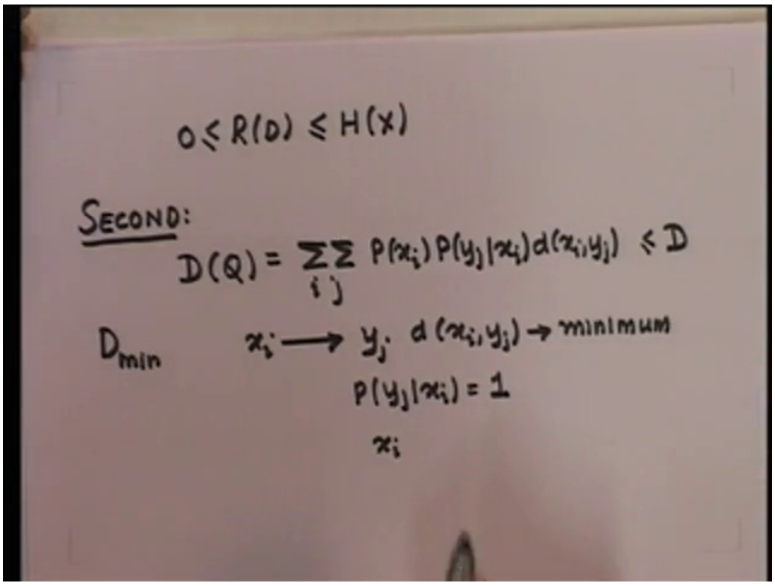
(Refer Slide Time: 26:45)



The range of rate distortion is given by R D is greater than equal to 0 and is less than equal to entropy of discrete memory less source. Now, this result can be shown as follows. We know that equivocation that is H of X given y is always greater than equal to 0 and is always less than equal to entropy of the variable X. And we also know based on our previous study that H of X is greater than equal to 0, and less than equal to log of N based on the assumptions that we have, a discrete memory less information source with the source alphabet of size N.

Now, average mutual information is equal to H X minus H of X given Y. Therefore, using the previous results it implies that I X Y is greater than equal to 0 is less than equal to H X is less than equal to log N. Now, since the rate distortion function, that is R D, is the minimum of the average mutual information, that is I X Y, we have rate distortion function less than equal to average mutual information, and based on the previous results we can say.

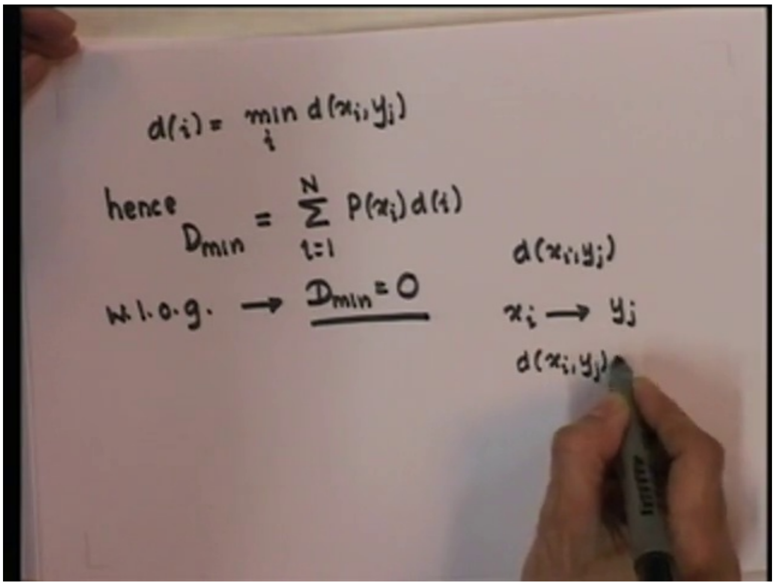
(Refer Slide Time: 29:20)



So, R D is greater than equal to 0 and less than equal to h x. Next, let us determine the permissible distortion D may take on. So, the next property of the rate distortion function is to find the permissible distortion D. So, we know that average distortion D Q is defined as double summation over i j of the quantity P x i, probability of y j given x i, distortion measure between x i y j, this quantity is less than equal to the specified distortion level D.

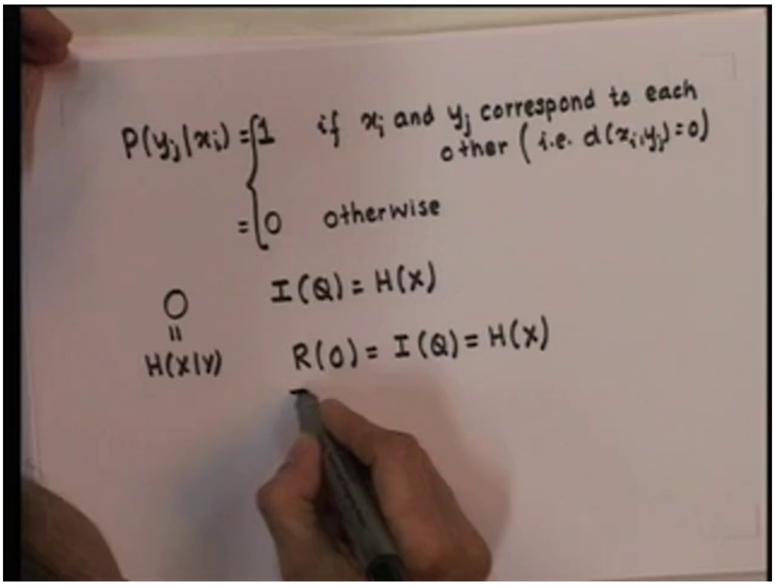
So, the smallest possible average distortion D minimum is obtained if for each source symbol, that is x i, we seek the reconstructed symbol y j where d x i y j is minimum and subsequently set this conditional probability that is y j given x i is equal to 1. Thus set the remaining conditional probabilities of the reconstructed symbol given that particular source symbol x i equal to 0.

(Refer Slide Time: 31:36)



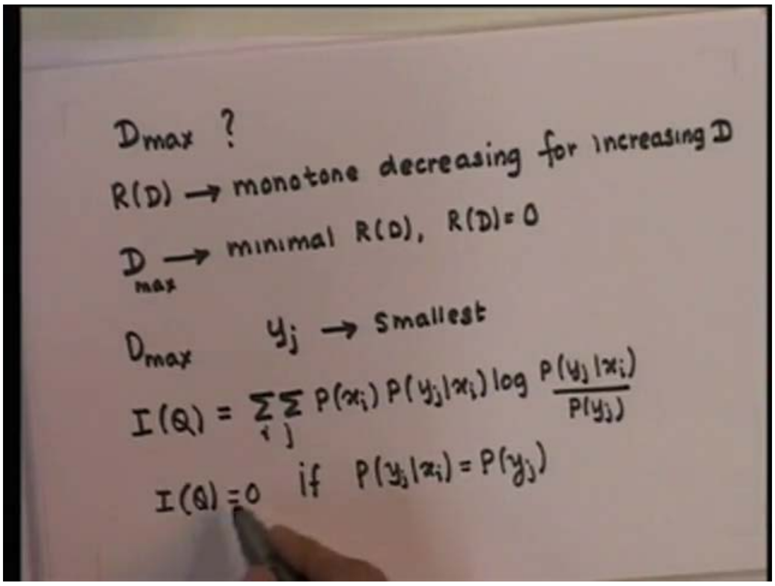
So, let us define di to be minimum of d x i y j over i. Hence, D minimum is equal to summation of p x i d i i is equal to 1 to N. Now, without loss of generality, we will assume that D minimum is equal to 0. If this is not the case, we modify the distortion d x i y j in such a manner that D minimum becomes equal to 0 again. This means that that for every source symbol x i there must be a symbol y j such that d x i y j is equal to 0.

(Refer Slide Time: 33:00)



Now, since for the conditional probabilities we have y j given x i equal to 1 if x i and y j correspond to each other that is distortion measure between x i and y j is equal to 0 is equal to 0 otherwise. The equivocation will be equal to 0. So, that I Q is equal to H X. Now, Q no longer appears here. So, that it directly follows that R 0 is equal to I Q is equal to H X. So, in order to achieve 0 distortion all of the information of the source must therefore, be constructed at the destination. Now, let us determine the maximum possible average distortion.

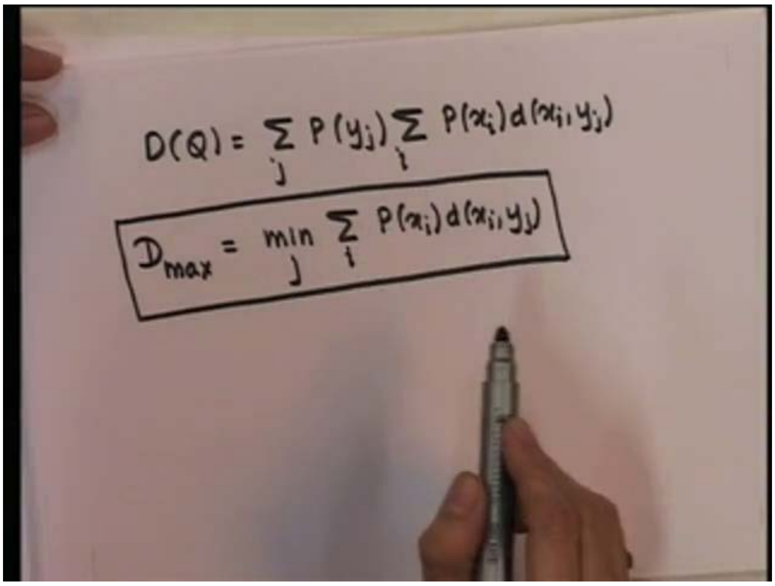
(Refer Slide Time: 34:44)



So, we are interested in finding out D max and this can be determined as follows. It can be shown that the rate distortion function is monotone decreasing for increasing distortion D. So, the maximum D then occurs for minimal R D that is R D equals to 0. This means that the destination receives no information from the source. Let us denote smallest possible distortion that is then still possible by D max. This value is achieved by constantly choosing that value y j for which, the average distortion between an arbitrary source symbol and the symbol y j in question is smallest and every other choice would lead to a larger average distortion.

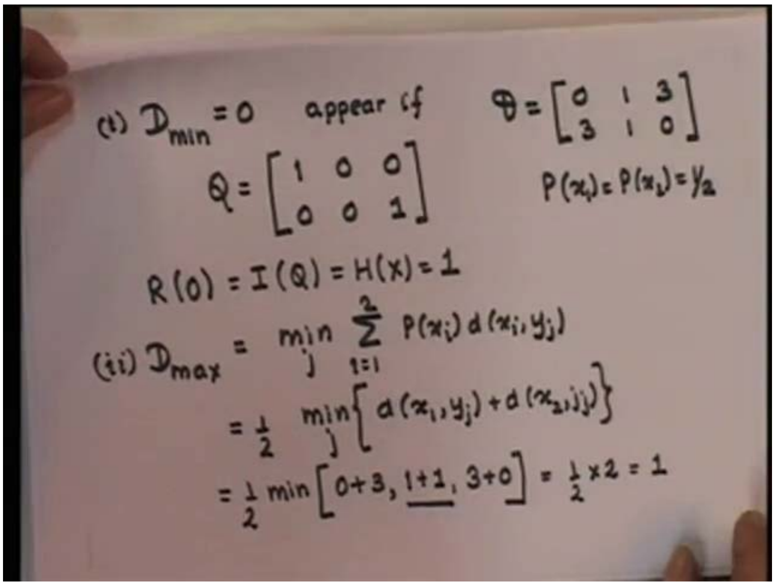
Now, the mutual information I Q is equal to P x i, probability of y j given x i, log of probability y j given x i over probability of y j. So, if I Q is equal to 0then probability of y j given x i is equal to probability of y j. So, if we choose conditional probability is equal to probability of y j which means, the occurrence of x i does not have any effect on the reconstruction of y j. So, that no information is reproduced in that case the average distortion D Q becomes.

(Refer Slide Time: 37:43)



So, the minimum value of D Q is obtained by setting probability y j equal to 1 for that value of j for which this quantity is smallest. So, this gives D max equal to minimum over j of the quantity P x i d x i y j summed over i. Let us try to appreciate this result with the help of an example. Let us consider again the earlier example, but now we want to find D minimum and D maximum, and the corresponding values of the rate distortion function.

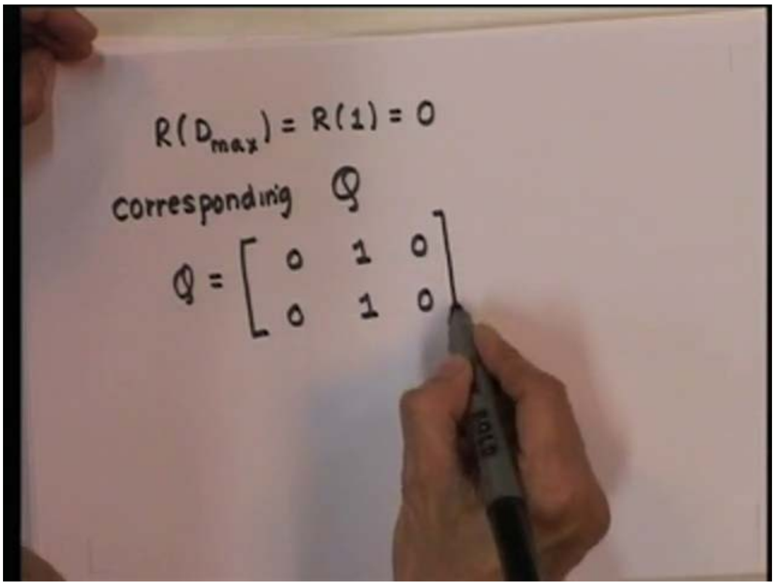
(Refer Slide Time: 39:25)



So, going back to the earlier example which we had discussed in the starting of the class, we can show that D minimum is equal to 0 and it appears, if the matrix Q is of the form 100001 where the distortion matrix was given as follows 013310 and probability of x 1 equal to probability of x 2 was specified to be half. Now, this corresponds to R 0 equal to I Q is equal to H X is equal to based on this one. So, the minimum value is equal to 0 and the rate for the minimum distortion is equal to the entropy of the source that is given by 1.

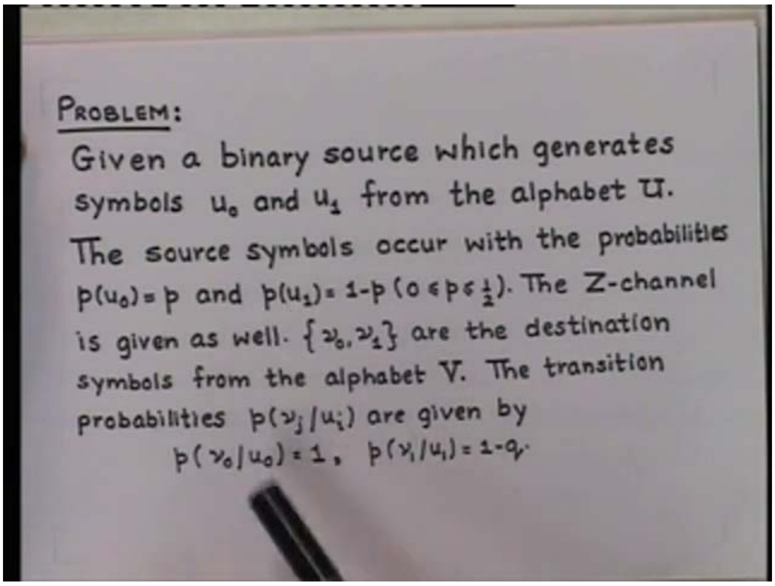
Now, let us evaluate the maximum value of the distortion. So, D max is equal to minimum over j from the quantity P x i d x i y j summed over 1 to 2 is equal to, p x i is equal to half therefore, is equal to half times minimum over j of d x 1 y j plus d x 2 y j and this is equal to half based on this matrix times minimum of 0 plus 3 1 plus 13 plus 0. So, the minimum is 2 is equal to half times 2 is equal to 1.

(Refer Slide Time: 42:32)



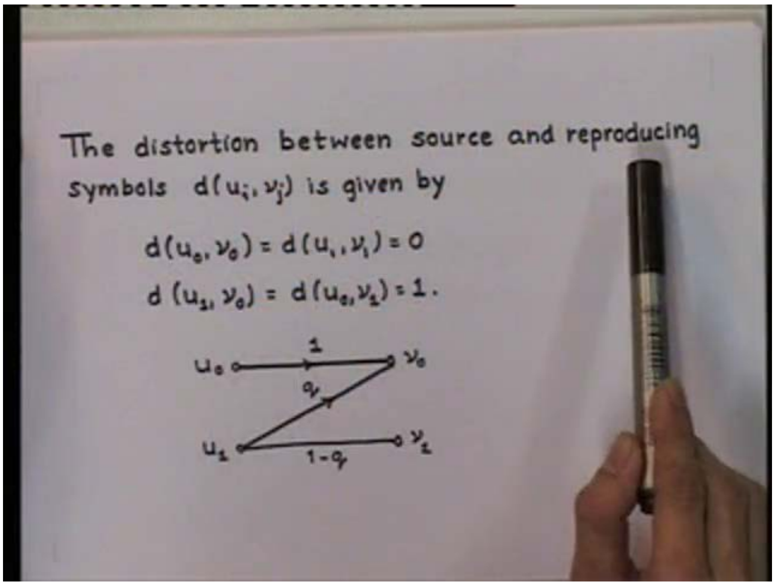
So, for this D max is equal to 1, the corresponding value of R D satisfies the relationship given here and the corresponding conditional probability matrix Q is given as shown here 010010. In order to get a better and a more clearer understanding of rate distortion function, let us solve another problem. The problem is given as follows.

(Refer Slide Time: 43:26)



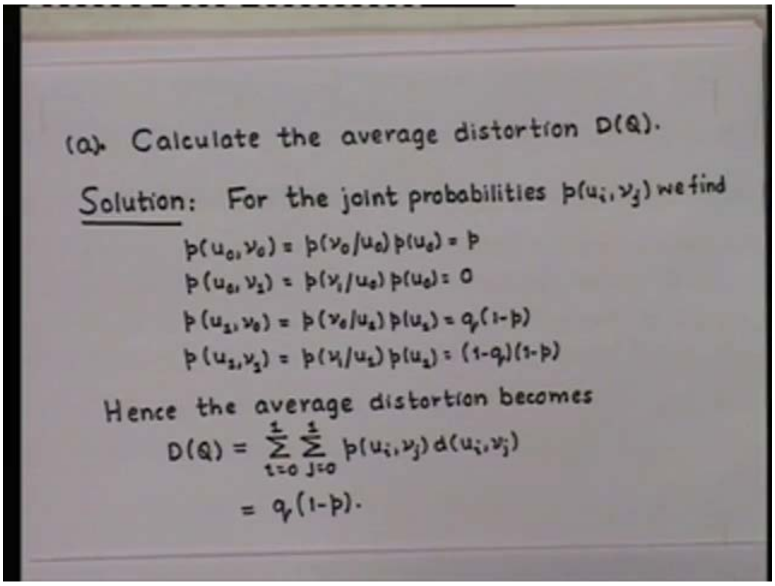
Given a binary source which generates symbols u 0, and u 1 from the alphabet u. The source symbols occurs with the probabilities, probability of u 0 equal to p and probability of u 1 equal to 1 minus p where p is greater than 0 and less than equal to half. The Z channel for this is given as well v 0 and v 1 are the destination symbols from the alphabet V. The transition probabilities, that is the probability of v j given u i are also given as follows. Probability of v 0 given u 0 is equal to 1 and probability of v 1 given u 1 is equal to 1 minus q.

(Refer Slide Time: 44:38)



We have also been provided the distortion between source and reproducing symbols that is distortion between u i and v j is given as follows. Distortion between u 0 v 0 is equal to distortion between u 1 and v 1 is equal to 0. And distortion between u 1 v 0 is equal to distortion between u 0 v 1 equal to 1. The Z channel is given as shown here u 0 and u 1 are the input symbols v 0 and v 1 are output symbols.

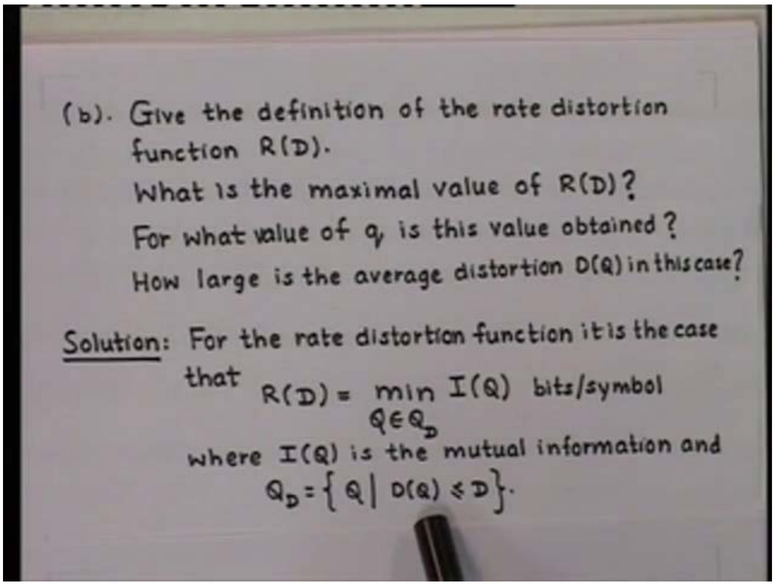
(Refer Slide Time: 45:39)



Given this, we are required to calculate the average distortion D Q, where Q is the conditional probability matrix and the solution is as follows. First, let us calculate the joint probabilities, that is the probability of u i and v j. We calculate this as follows. Probability of u 0 v 0 is equal to conditional probability, that is probability of v 0 given u 0 multiplied by probability of u 0 and based on the information provided, this is equal to p.

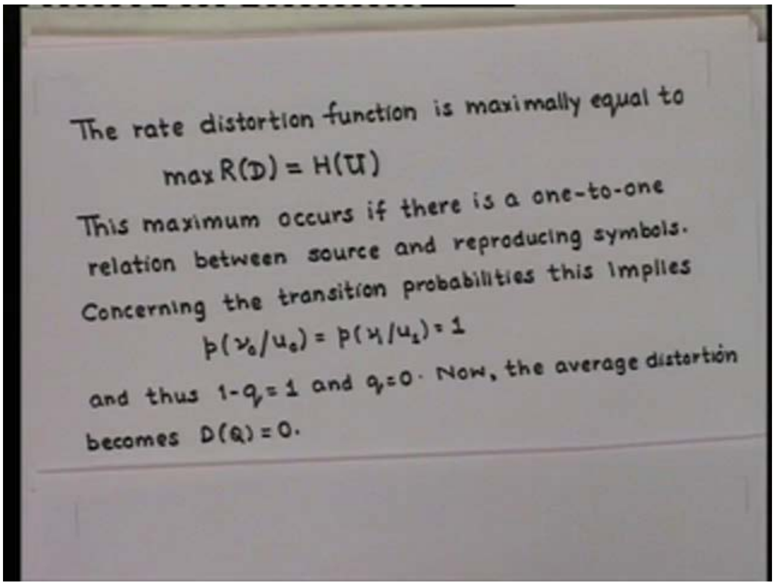
Similarly, we can calculate the joint probabilities p u 0 v 1 which comes out to be 0, probability u 1 v 0 is equal to Q into 1 minus p, probability of u 1 v 1 is equal to 1 minus Q multiplied by 1 minus p. Given these joint probabilities we can plug this along with the distortions given between the source and reproducing symbols and the average distortion becomes as follows. D Q given by this formula is equal to Q multiplied by 1 minus p.

(Refer Slide Time: 47:14)



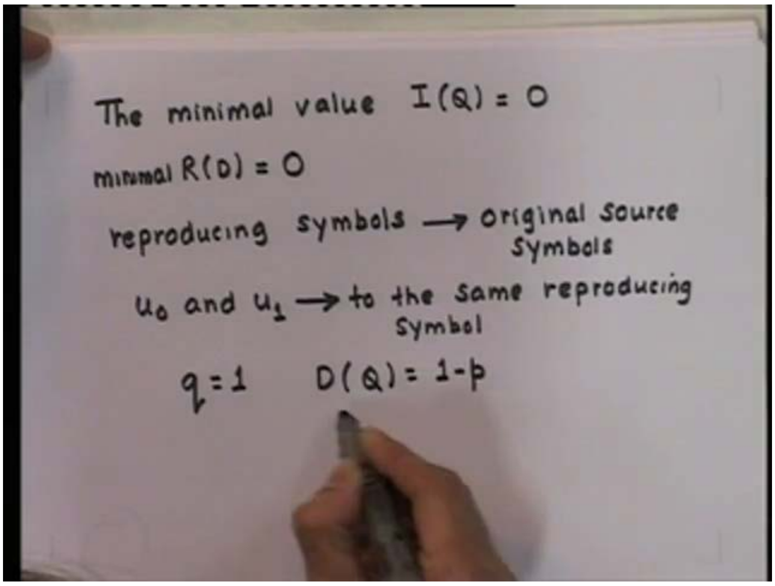
Next, we are required to provide the definition of the rate distortion function. Then for this problem what is the maximal value of R D for what value of Q is this value obtained? And, in this case, how large is the average distortion D Q? The solution to this task, the solution for the same is as follows. For the rate distortion function, it is the case that R D is defined as minimum of I Q. This minimum is over Q belonging to Q D where I Q is the mutual information and Q D is the set of all conditional probability matrix, such that distortion for that conditional probability matrix is less than, or equal to the specified distortion level D. So, this is the definition of the rate distortion function which we had seen earlier.

(Refer Slide Time: 49:38)



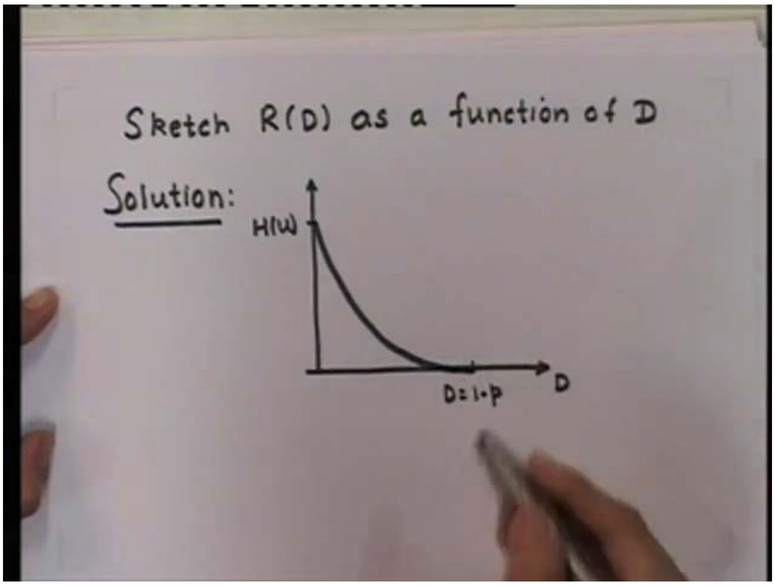
Now we have also seen that the rate distortion function is maximally equal to H U and this maximum occurs if there is one-one relation between source and reproducing symbols. So, concerning the transition probabilities, this implies that we should have probability of v 0 given u 0 is equal to probability of v 1 given u 1 equal to 1 which implies that 1 minus q equals to 1 which implies q is equal to 0 and. Now, in this case the average distortion becomes equal to 0. Next question is what is the minimal value of rate distortion function for what value of Q is this value obtained? And how large is the average distortion D Q in this case?

(Refer Slide Time: 49:58)



Now, the minimal value of the mutual information I Q is equal to 0. This we have proved earlier and thus the minimal value of the rate distortion function also equals 0. This minimum appears, if the reproducing symbols give no information about the original source symbols and this is the case if both symbols u 0 and u 1 give rise to the same reproducing symbol. So, in this case, it leads to Q equal to 1 and now distortion is equal to 1 minus p and the final question is…

(Refer Slide Time: 51:28)



Let us sketch rate distortion function as a function of D and the solution to this is as follows. We have seen the maximum value is H U, minimum value is 0 and that value is achieved for D equal to 1 minus p and in between this it will be monotonically decreasing function. So, for D equal to 1 minus p, R D is equal to 0. There are two ways to calculate the rate distortion function.

We have looked at one approach, which is to calculate the mutual information, and find the lower bound for this mutual information and show that this lower bound is achievable. The other approach is a computational approach. In the next class, we will discuss the computational approach to calculate the rate distortion function and also, explore the application of the same to a binary case.