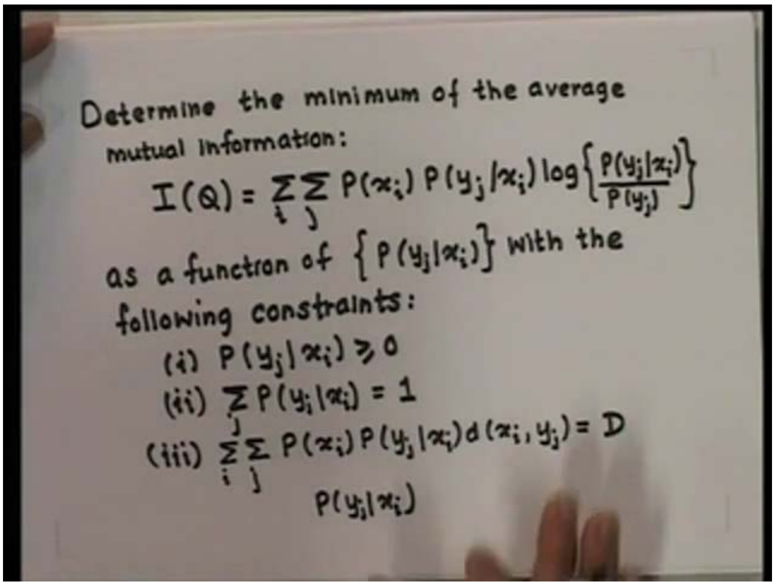
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 34
Computational Approach for Calculation of Rate-Distortion Functions

In the earlier classes, we have studied an approach to evaluate the rate distortion function for both discrete and continuous sources. In this approach, essentially we calculated the lower bound for the average mutual information, and then showed that this bound is achievable. In today's class, we will look at the computational approach to evaluate the rate distortion function. We will restrict our discussion to discrete sources, without going into the details of the derivation, we will examine the salient features of this derivation. The determination of the rate distortion function can be regarded as a problem from the calculus of variations. We can formulate the problem as follows.

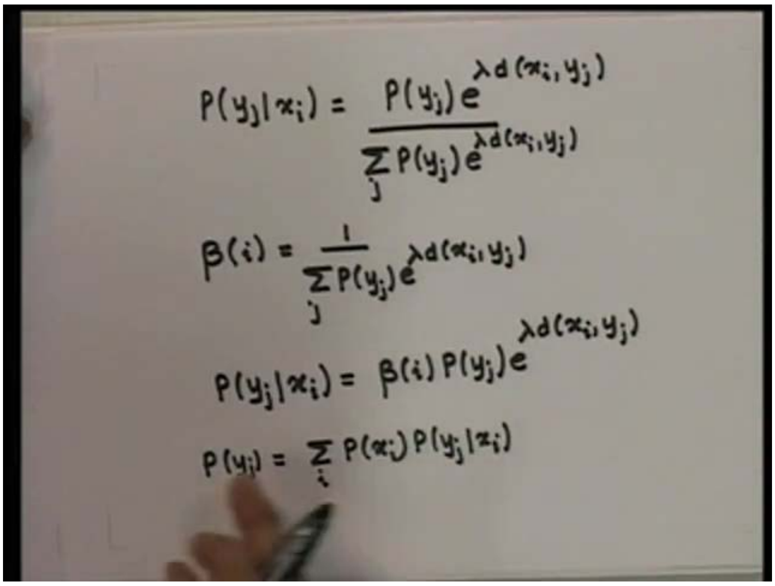
(Refer Slide Time: 02:20)



Determine the minimum of the average mutual information that is I of Q, where based on the earlier notation, Q is the conditional or transitional matrix is equal to double summation over i j of probability of the source multiplied by the conditional probabilities log of. So we have to minimize I Q as a function of conditional probabilities with the following constraints.

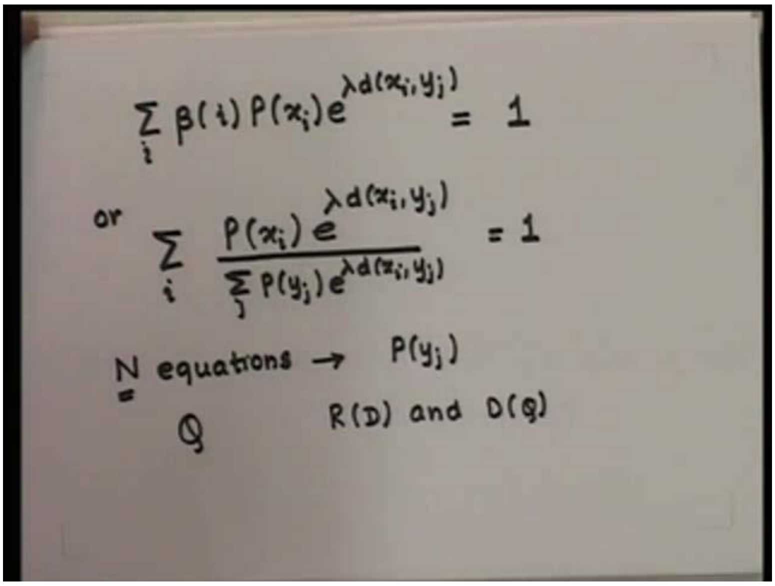
First conditional probability should be all greater than equal to 0, summation of the conditional probabilities, that is probability of y j given x i is equal to 1 and the average distortion, which is double summation over i j of probability of the source multiplied by conditional probabilities distortion between the source and reconstructing symbols, this is equal to specified distortion level D. So, if we ignore the constraint one, a solution is possible by differentiating the average mutual information with respect to conditional probabilities that is, probability of y j given x i and then equating this result to 0. Now, we will not go into the derivation of this, but using Lagrangian's method, we will obtain the following expression for the conditional probabilities.

(Refer Slide Time: 06:02)



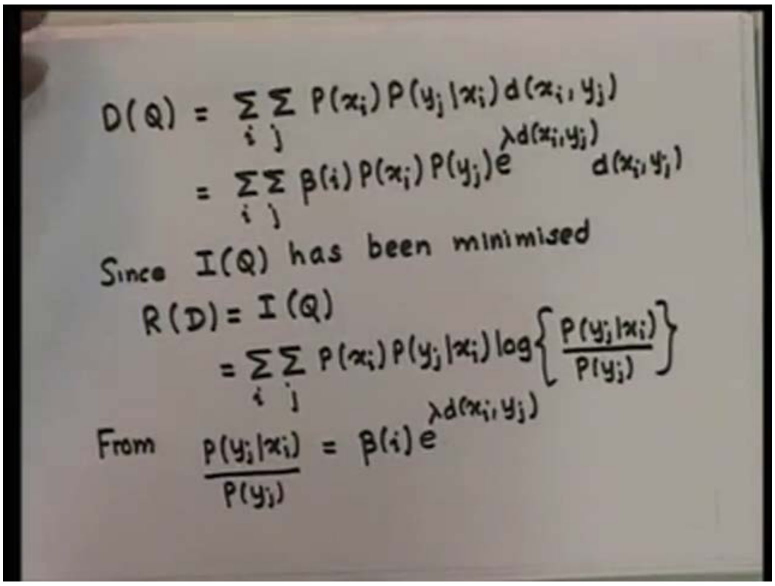
Probability of y j given x i is equal to probability y j multiply e raise to exponential. Now, if we introduce another parameter called beta i, which is equal to 1 by summation of P y j, e raise to lambda d x i, y j, then this expression that is conditional probability can be written as probability of y j given x i is equal to beta i times probability y j, e raise to lambda distortion between x i, y j. So, this procedure yields a set of equations for each i and j where probability of y j given x i is expressed in terms of probabilities of y j it therefore, now remains to determine the probability p y j and in general probability p y j is equal to summation over i of P x i multiplied by conditional probabilities P y j given x i.

(Refer Slide Time: 08:30)



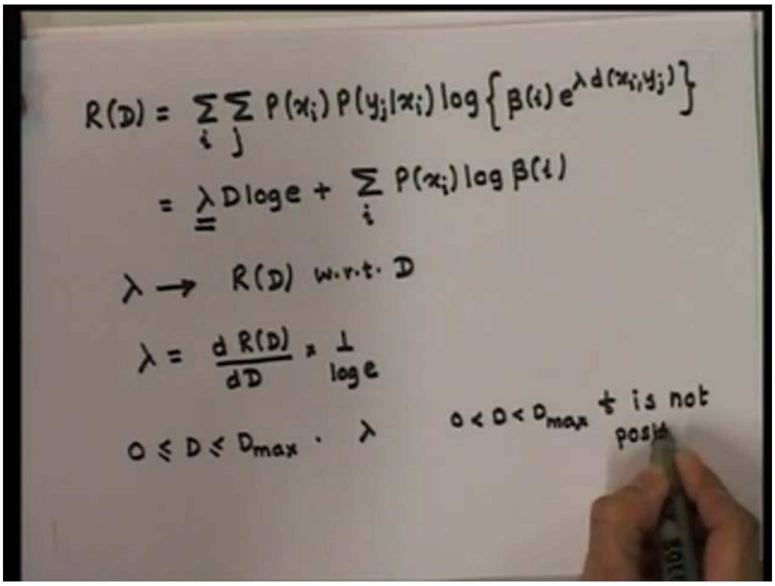
So, if we divide this equation by probability of y j and use this relationship, we get the following relationship is equal to 1 or same thing can be rewritten as summation over i of P x i, e raise to lambda d x i, y j. Now, it must still be verified, if constraint one is satisfied. So, in this manner we get n equations assuming that the size of the reconstruction alphabet is N for the N probabilities P y j, which now may be solved and after this the conditional or transitional probability matrix q can be determined. So, with the equation so obtained it is now possible to obtain an expression for R D and D Q as follows.

(Refer Slide Time: 10:23)



We know that the distortion as a function of conditional probability matrix Q is equal to double summation i j probability x i multiplied by conditional probability, multiplied by distortion between x i, y j. This can be rewritten based on the previous results as summation over i j beta i, P x i, P y j, e raise to lambda d x i y j, d x i y j. Now, since I Q has been minimized, we can write the rate distortion function R D equal to I Q, which is equal to double summation over i j of the quantity P x i, P y j given x i log of P y j given x i divided by probability y j. From the relationship, which we have derived from P y j given x i, over P y j equal to beta i, e raise to lambda d x i, y j. We can write the rate distortion function R D as follows.

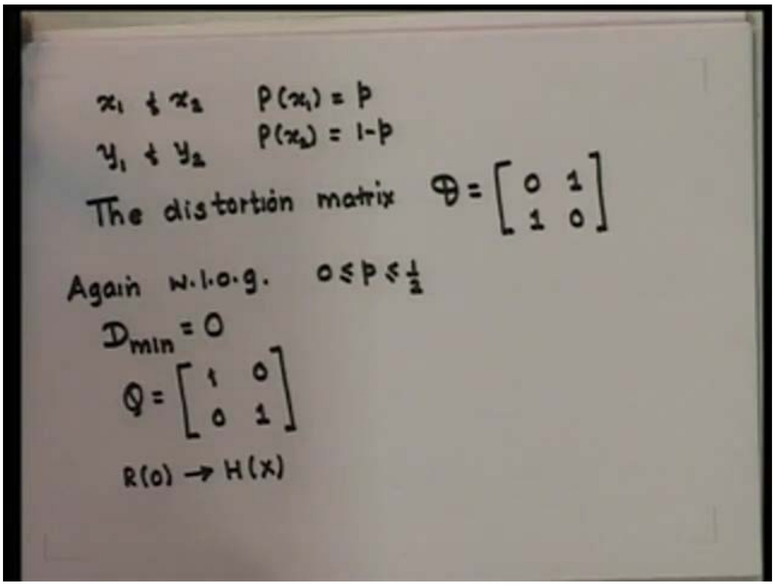
(Refer Slide Time: 13:05)



R D is equal to double summation of i j, P x i, p of y j given x i log of beta i, e raise to lambda d x i y j and this can be further simplified to get the following result. So, we have now found an expression for the rate of the source as a function of the permissible distortion D. Now, the solution is found in implicit form, why the parameter lambda? And explicit relation is not possible excepting in a few simple cases. Now, a value of lambda yields a value for D and for R D and thus gives a point on the R D curve.

Now, it can be shown that the parameter lambda is proportional to the first derivative of R D with respect to D and is therefore, connected to the slope of the R d curve at a certain point. So, we have the relationship as lambda is equal to derivative of r D with respect to D with a constant multiplying factor 1 by log e. Therefore, the R D is a continuous monotone decreasing function for D greater than equal to 0 and less than equal to D max. The parameter lambda is continuous for D greater than 0, less than D max and is not positive. In order to get a better understanding of the concepts of the computational approach, let us apply this approach to evaluate the rate distortion function for a binary source. Earlier we have already derived the rate distortion function for a binary case; we will redo the problem using the computational approach.

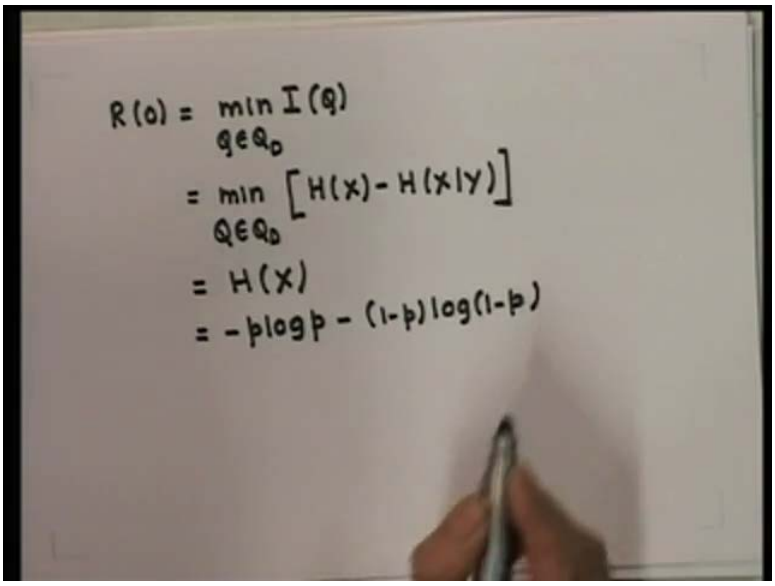
(Refer Slide Time: 16:46)



So, let us assume that the source generates symbols x 1 and x 2 with the probabilities P x 1 equal to small p and P of x 2 equal to 1 minus p. Let us assume that the reproducing symbols are y 1 and y 2, again using the earlier notation. Let us assume that the distortion matrix is given by 0 1, 1 0, which means that if x 1 is reproduced as y 1, there is no distortion, but if x 1 is reproduced as y 2 there is a distortion 1 and similarly, for x 2.

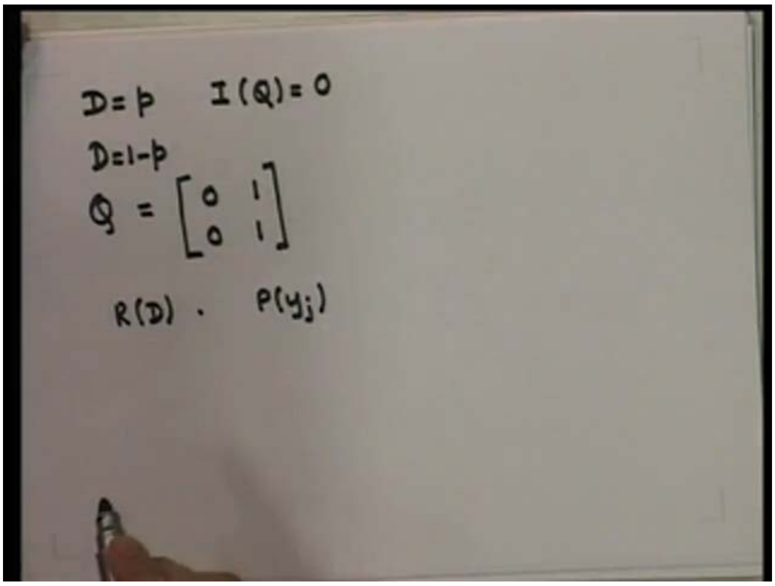
Now, again without the loss of generality, we assume that p is greater than 0 and less than equal to half. So, in this case it is easy to see that, D minimum is equal to 0, that is the minimum distortion is equal to 0 and this occurs, if we choose for the conditional probability matrix Q equal to 1 0 0 1. The R D function R 0 is equal to H x in this case.

(Refer Slide Time: 19:20)



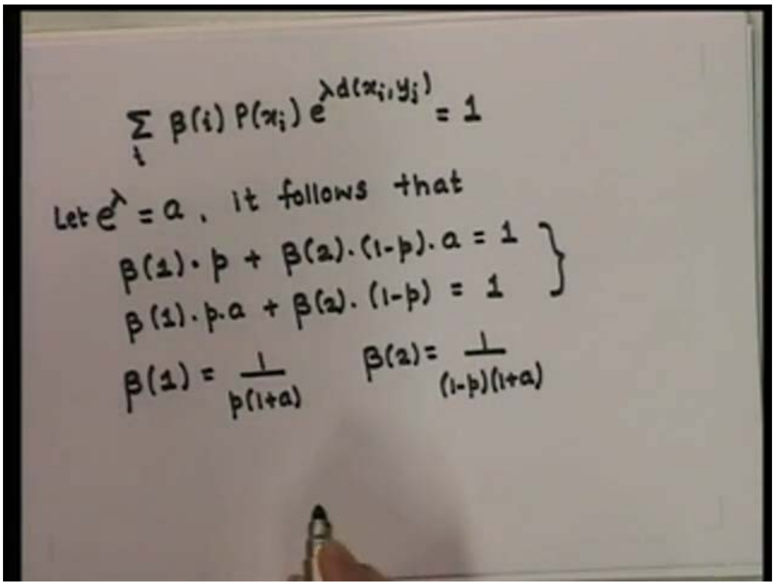
Now, this is not difficult to show R 0 is equal to minimum I Q, Q belonging to Q D and if D is equal to 0 this is equal to minimum Q belonging to Q D of H x minus H x given Y. Since, the equivocation is 0 this is equal to H x, which is equal to minus p log, p minus 1 minus p log of 1 minus p.

(Refer Slide Time: 20:12)



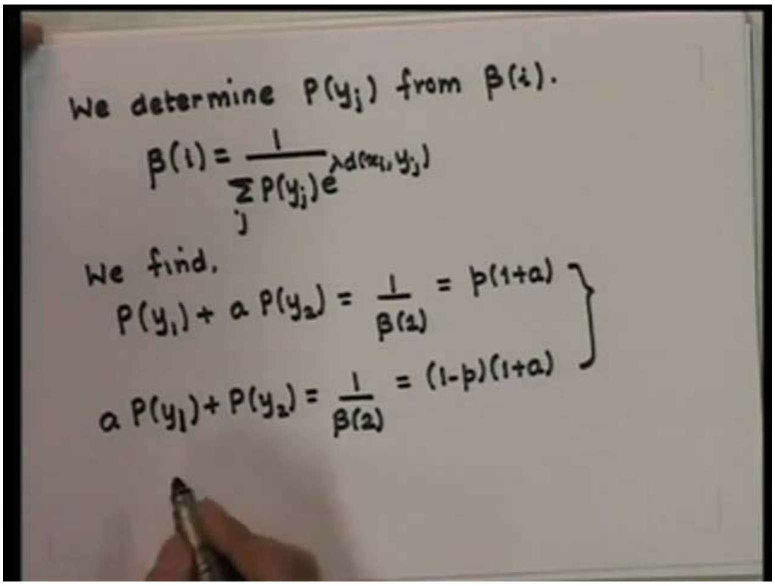
Now, the maximum distortion is D equal to p now, this is the best choice that can be made if I Q is equal to 0, because the alternative D equal to 1 minus p, always yields a larger distortions, since p is less than equal to half. So, the matrix Q of the conditional probabilities, in this case is equal to 0 1 0 1. Now, in order to determine the other points of the R D curve, the probabilities P y j must first be found.

(Refer Slide Time: 21:12)



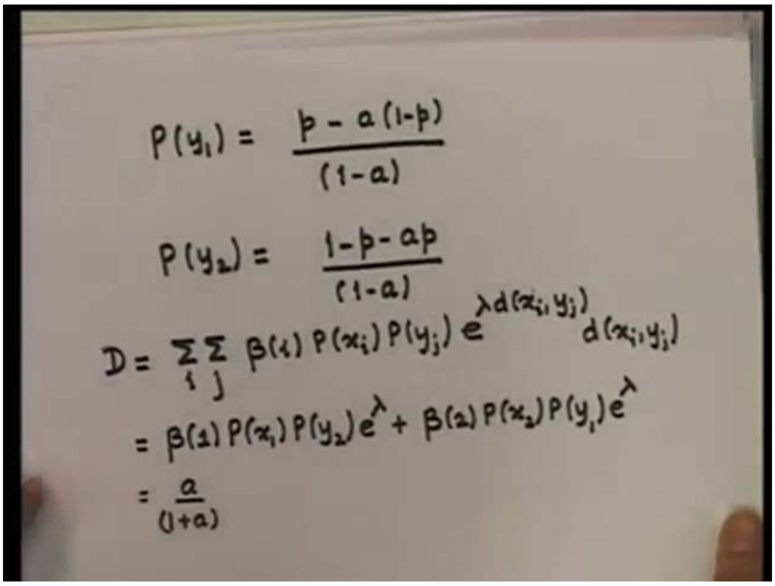
So, we have seen that summation over i of b i, P x i, e raise to lambda d x i y j is equal to 1. Now, let us assume e raise to lambda equal to a, then it follows that beta 1 into small p plus beta 2 into 1 minus p multiplied by a is equal to 1 and beta 1 multiplied by p multiplied by a plus beta 2 times 1 minus p y 2 is equal to 1 by beta 1. If this two equations are solved simultaneously, we get beta 1 is equal to 1 by p into 1 plus a, and beta 2 is equal to 1 by 1 minus p times 1 plus a. Next, we determine probability y j from beta i now, this can be done on the bases of the relation beta i is equal to 1 by summation over j of P y j, e raise to lambda d xi y j. So, using this relationship we find, P y 1 plus a times P y 2 is equal to 1 by beta 1, which is equal to p times 1 plus a.

(Refer Slide Time: 23:02)



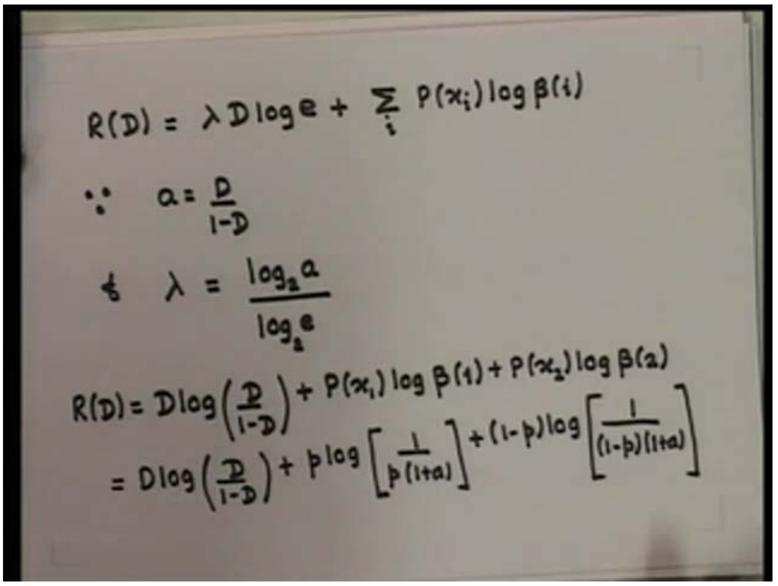
Another relationship which we get as follows, a times probability y 1 plus probability y 2 is equal to 1 by beta 2 is equal to 1 minus p times 1 plus a. If this two equations are solved we get probability of y 1 is equal to p minus a times 1 minus p divided by 1 minus a, and probability of y 2 is equal to 1 minus p minus a times p divided by 1 minus a.

(Refer Slide Time: 25:02)



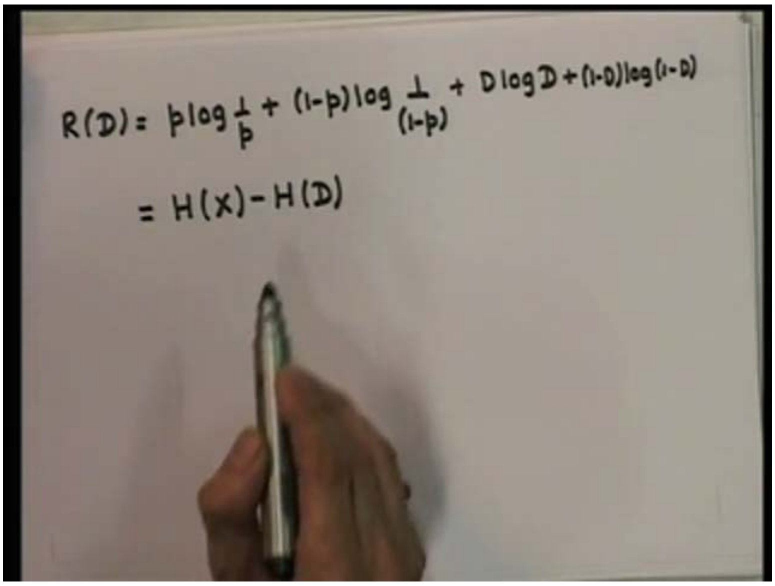
Now, if we use this result and substitute in the expression for the permissible distortion, we get the result as follow D is equal to double summation over i j of beta i, P x i, P y j, e raise to lambda d x i y j times d x i y j and this can be simplified as beta 1 time P x 1 times P y 2, e raise to lambda plus beta 2, p x 2, p y 1 e raise to lambda. If we substitute the values for p y 1, p y 2 beta 1 beta 2, p x 1, p x 2, we can show that this simplifies to a by 1 plus a.

(Refer Slide Time: 27:31)



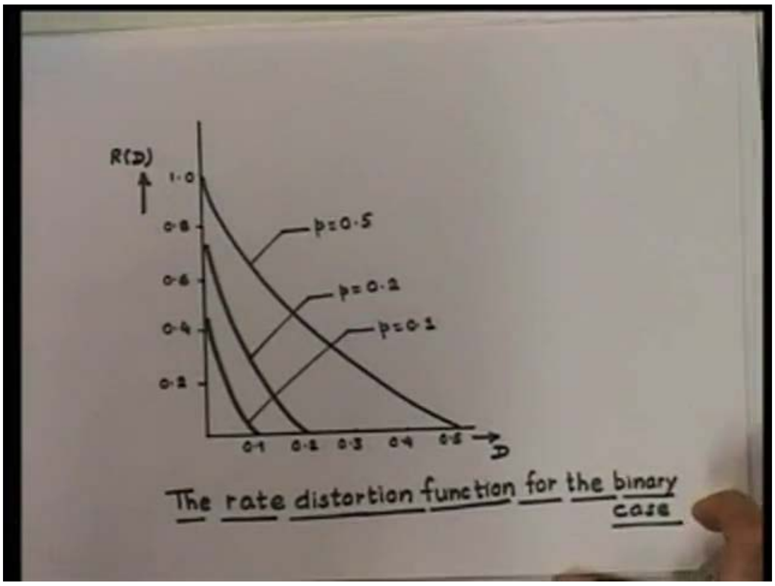
Now, we know that R D is equal to lambda times D multiplied by log to the base 2 of e plus summation of P x i times log beta i. Now, because a is equal to D by 1 minus D, from this relationship and lambda is equal to log 2 a divided by log of e to the base 2 R D can be rewritten as follows, is equal to D times log of D 1 minus D, plus P x 1, log beta 1 plus, p x 2 log beta 2. This can be further simplified as D log of D 1 minus D plus p times log of 1 by p times 1 plus a plus, 1 minus p times log of 1 by 1 minus p times 1 plus a.

(Refer Slide Time: 30:10)



Now, it is not very difficult to show that this can be reduced to the following expression, R D is equal to p times log 1 by p, plus 1 minus p times log of 1 minus p, plus D log D, plus 1 minus D times log 1 minus D and this is equal to H X minus H D. So, we have thus found a relation from which we can directly determine the mutual information, for a given value of D which must be conveyed in order to achieve an average distortion D.

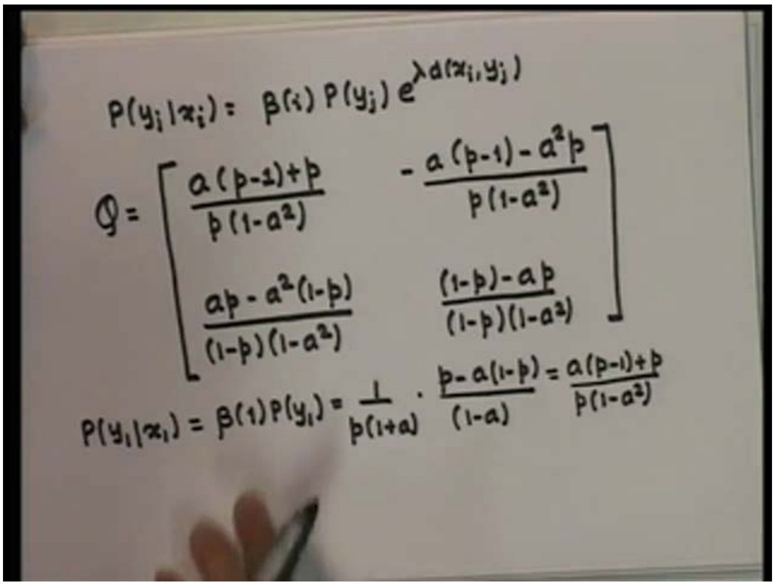
(Refer Slide Time: 31:26)



So, this figure provides the rate distortion curve for a few values of p, from the figure it can be seen that a smaller average distortion can only be achieved by increasing the rate.

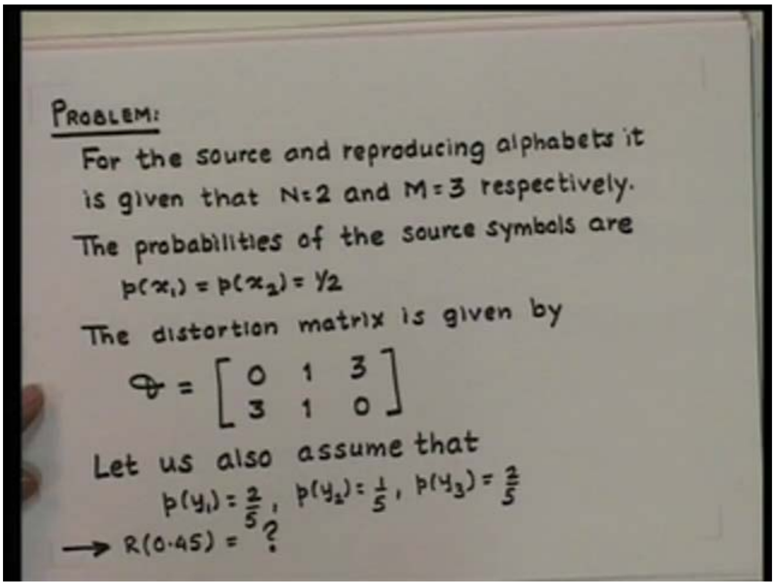
The rate distortion function R D for p equal to 0 point 5 is larger than for p less than 0 point 5 for every value of D, which is again intuitively true. So, each point on the curve is reached by a matrix of conditional or transitional probabilities, which give rise to both an average distortion D Q equal to D and average mutual information R D.

(Refer Slide Time: 32:51)



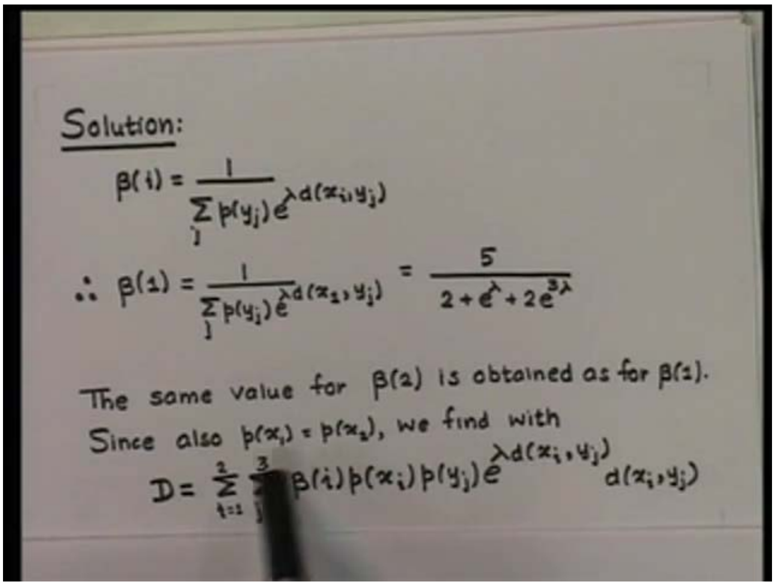
Now, for the conditional or transition probabilities, we have probability of y j given x i is equal to beta i times probability y j, e raise to lambda d x i, y j and this yields the following matrix Q and finally, 1 minus p minus a p whole over 1 minus p 1 minus a squared. Now, the procedure to evaluate any of this term is very simple for example, probability of y 1 given x 1, that is this quantity is equal to beta 1 times p y 1. This is equal to 1 time 1 by p, 1 plus a multiplied by p minus a, 1 minus p whole over 1 minus a, and this simplifies to a p minus 1 plus p divided by p times 1 minus a squared. So similarly, other terms can be evaluated. Now, let us extend the application of the computational approach to a discrete source which is not a binary.

(Refer Slide Time: 35:26)



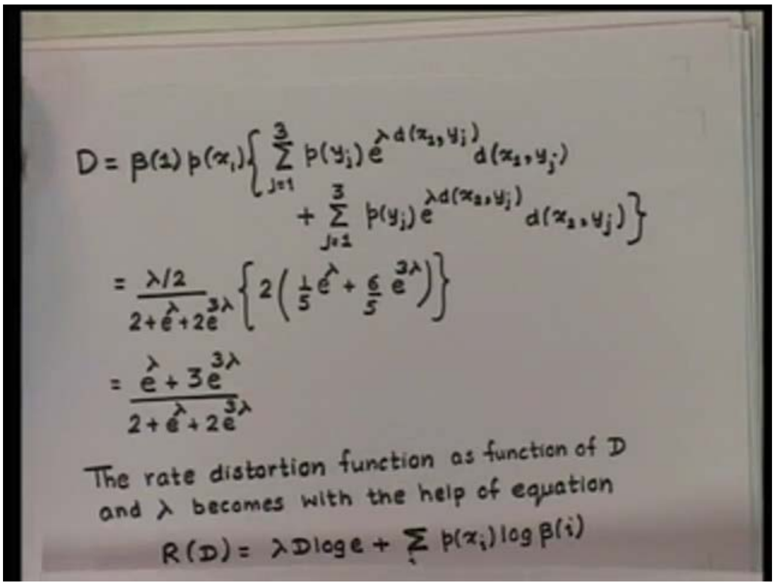
So, let us take another problem, where we have a source and reproducing alphabet. The source alphabet size is 2 and reproducing alphabet size is equal to 3. The probabilities of the source symbols are given as follows, probability of x 1 is equal to probability of x 2 is equal to half and the distortion matrix is given as shown here. Let us also assume that, the output or the reproducing symbols probabilities are specified, this is to simplify the problem. So, p of y 1 is equal to 2 by 5 p of y 2 is equal to 1 by 5 and p of y 3 is equal to 2 by 5 and the problem is to calculate the rate distortion function for D equal to 0.45.

(Refer Slide Time: 36:12)



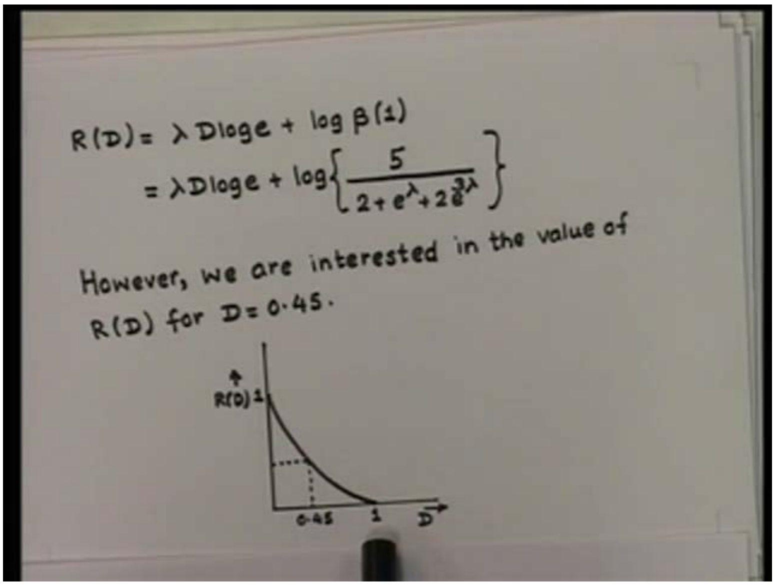
So, the solution is as follows, we know that beta i is given by this expression therefore, beta 1 can be rewritten as based on the fact that probability of P y j has been specified and distortion between x 1 and y j has been also specified. Now, the same value for beta 2 is obtained as for beta 1 since also p x 1 is equal to p x 2, we find that the distortion D which is given by this expression can be rewritten as follows.

(Refer Slide Time: 37:34)



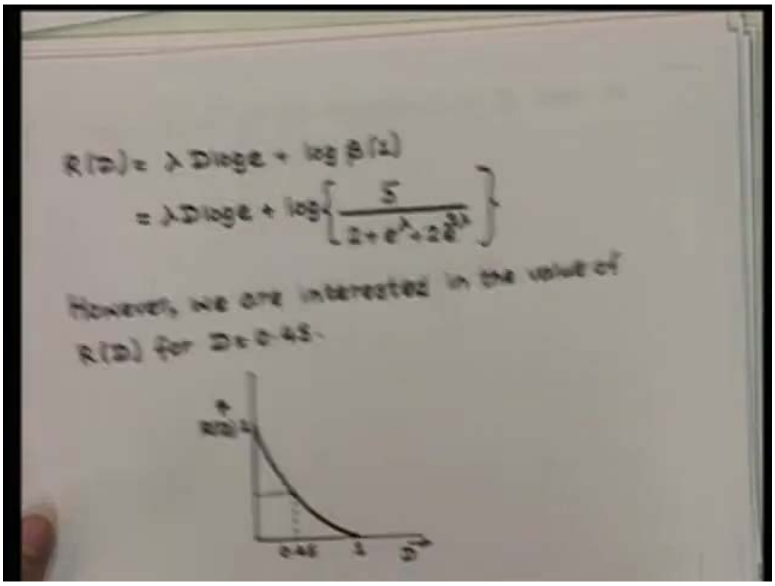
D is equal to beta 1 times probability of x 1 multiplied by this two terms and if we substitute the values for this terms, we get the following expression. So, the rate distortion function, as function of D and lambda becomes with the help of the equation R D equal to lambda D log e plus summation of p x i log b i as follows.

(Refer Slide Time: 38:19)



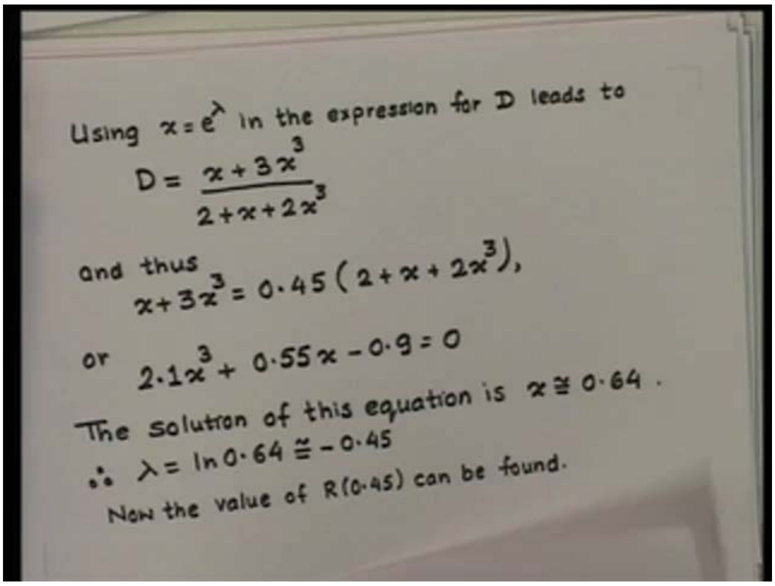
R D is equal to lambda D log e plus log beta 1; because beta 1 and beta 2 are same and this can be simplified to this expression. Now, however we are interested in the value of R D for D equal to 0.45 the rate distortion curve would be as follows, the for D equal to 0, R D is equal to 1 and for R D equal to 0 maximum distortion is equal to 1.

(Refer Slide Time: 39:15)



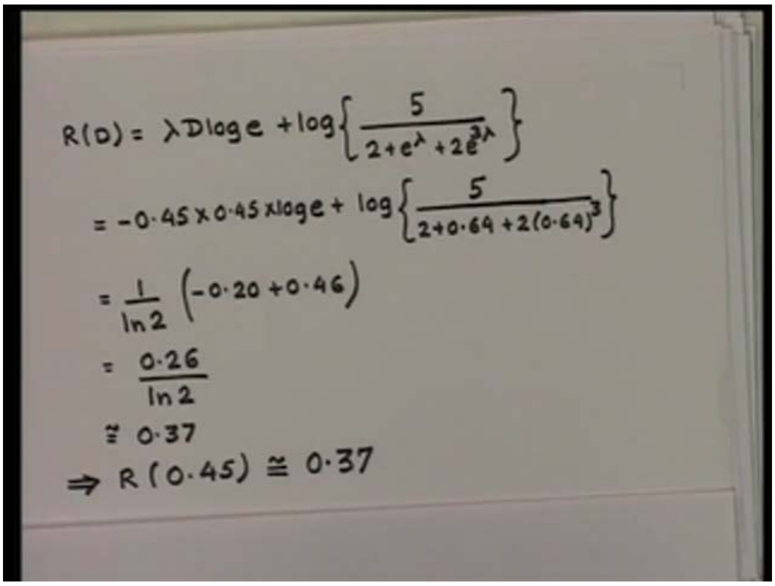
So, if we use x equal to e lambda, in the expression for D, this leads to D is equal to this quantity and we can solve this quantity as follows.

(Refer Slide Time: 39:18)



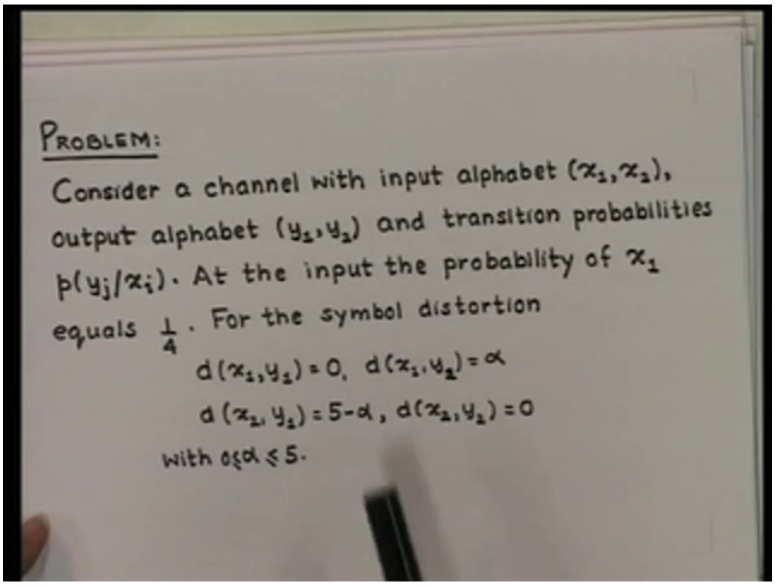
The solution of this equation is, x is approximately equal to 0.64, therefore log is logarithmic of this term, which is equal to minus 0.45. Now, the value of rate distortion function for D equal to 0.45 can be found as follows.

(Refer Slide Time: 40:02)



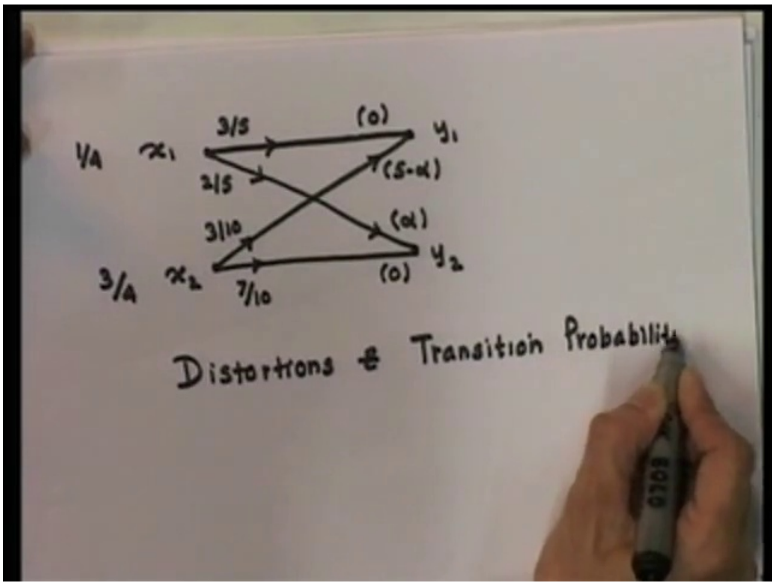
RD is equal to this expression, substitute the values for lambdas and D, we can show that this reduces to 0.37. So, this shows how to apply the computational approach to evaluate the rate distortion function in the example discussed. Let us take one more example to understand the concepts of evaluation of the rate distortion function.

(Refer Slide Time: 40:48)



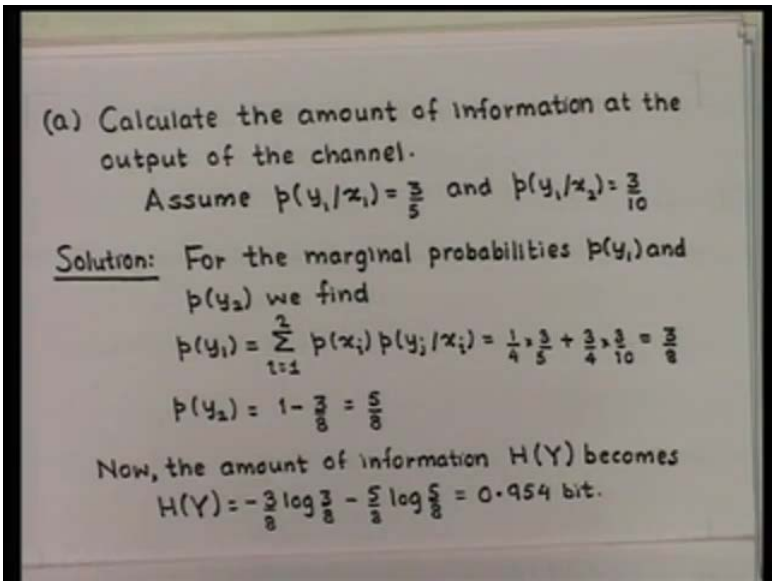
So, let us consider a channel with input alphabet given by x 1 x 2, output alphabet y 1 y 2 and transition probabilities. Probability y j given x i at the input the probability of x 1 equals one-fourth. For the symbol distortion we assume that distortion between x 1 y 1 is equal to 0 that between x 1 and y 2 is given by a variable factor alpha and between x 2 y 1 is equal to 5 minus alpha and distortion between x 2 y 1 is equal to 0, where we assume that alpha is greater than equal to 0 and less than equal to 5.

(Refer Slide Time: 41:38)



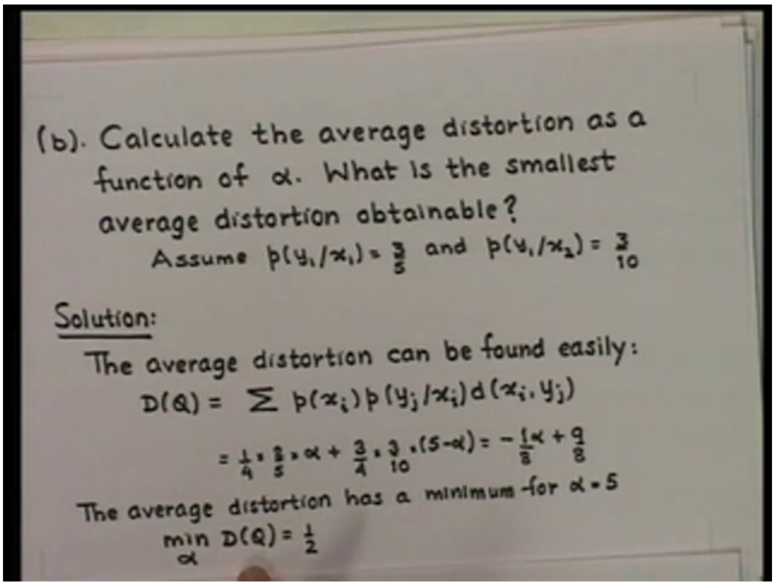
So, distortions and transition probabilities between the inputs and the reproducing symbols can be denoted by this diagram, x 1 x 2 y 1 y 2. Probability of this is one-fourth probability of x 2 is three-fourth probability of y 1 given x 1 is 3 by 5. This would be 2 by 5 this is 3 by 10 and this is 7 by 10 and distortion between x 1 and y 1 is denoted in brackets this 0. This is 5 minus alpha between x 1 and y 2 is alpha and that between x 2 and y 2 is 0.

(Refer Slide Time: 43:15)



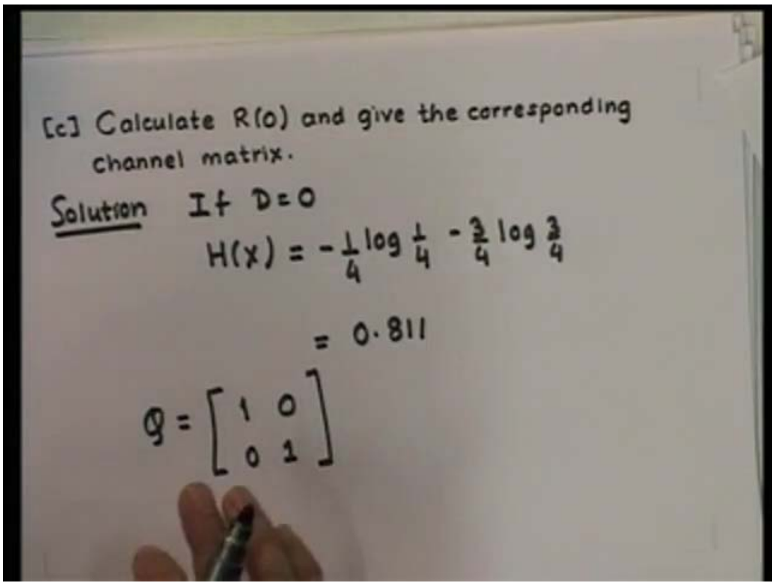
So, this is distortions and transition probabilities, problem is calculate the amount of information at the output of the channel assuming, that probability of y 1 given x 1 is 3 by 5 and probability of y 1 given x 2 is equal to 3 by 10 and this is very simple. So, for the marginal probabilities p y 1, p y 2, we find using this expression is equal to 3 by 8, 5 by 8. So, the amount of information H Y becomes 0. 954 bit.

(Refer Slide Time: 44:05)



The next problem is, calculate the average distortion as a function of alpha and what is the smallest average distortion obtainable? Again assume conditional probabilities as specified here. The average distortion can be found easily D Q is equal to this expression and when we plug in the values, we get D is equal to function of alpha and the average distortion has a minimum value for alpha equal to 5, so minimum D Q is equal to half.

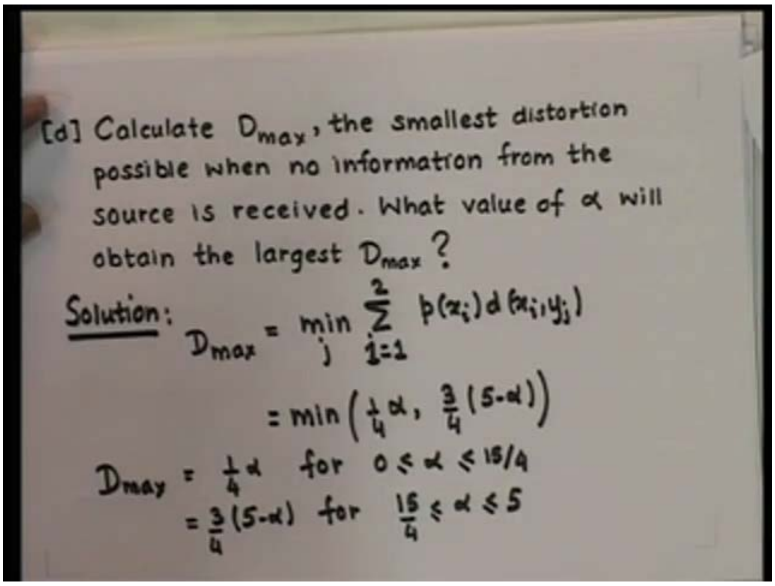
(Refer Slide Time: 44:34)



Next is calculate R 0 and give the corresponding channel matrix. Now, the solution to this is, if D is equal to 0, the rate distortion function achieves a maximum value equal to

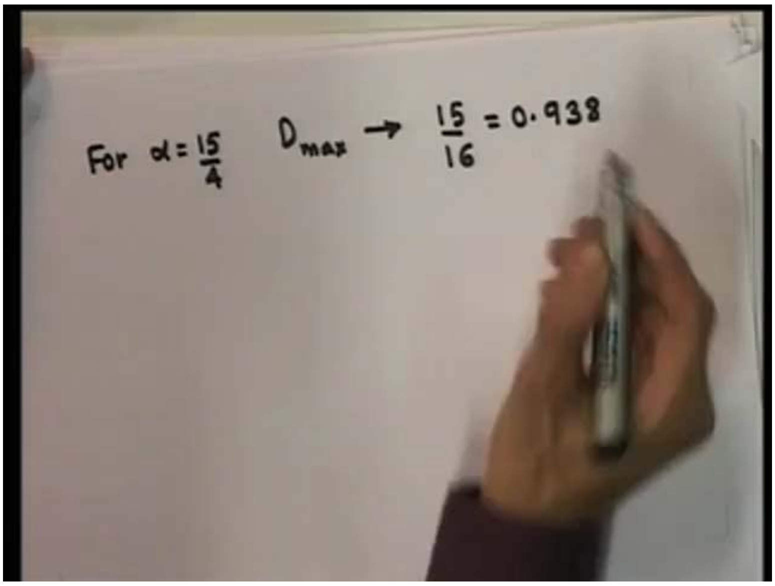
H X, which is the information of the source. In this case, this is equal to minus one-fourth, log one-fourth minus three-fourth, log three-fourth is equal to 0.811 and the corresponding channel matrix Q is equal to 1 0 0 1. That means there is a 1 to 1 correspondence between input and output symbol.

(Refer Slide Time: 45:36)



Finally, let us calculate D max, which is the smallest distortion possible, when no information from the source is received and what value of alpha will obtain the largest D max. Now, the solution again is as follows, on the basis of the theory we have discussed D max is equal to minimum over j of summation i equal to 1 to two, P x i, d x i y j and in the present case we get this, minimum of 1 by 4 alpha 3 by 4 times 5 minus alpha. Therefore, D max is equal to 1 by 4 alpha, for alpha between 0 and 15 by 4 and is equal to 3 by 4 times 5 minus alpha for alpha between 15 by 4 and 5.

(Refer Slide Time: 47:19)



For alpha equal to 15 by 4 D max will achieve its absolute maximum, which is 15 by 16, is equal to 0.938. So, after having examined some of these problems, we can come to the following conclusion, that determining the rate distortion function is generally not easy. And because of this a lower limit is often used for calculating the rate distortion function, numerical techniques also do exist for calculation of rate distortion function. Over next few classes, we will examine how the concepts of rate distortion function can be utilised to design efficient lossy compression schemes. In the next class, we will begin this study with contestation of a continuous source in finite discrete levels.