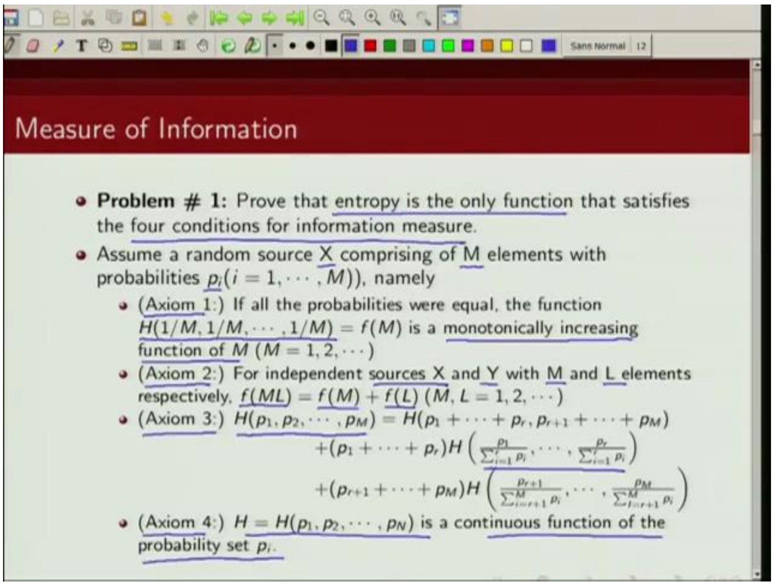
An Introduction to Information Theory
Prof. Adrish Banerjee
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kanpur

Lecture – 14B
Problem Solving Session-IV

Welcome to the course on Introduction to Information Theory. So, this is our last lecture. So, I thought why not solve some problems related to information theory. So, we are going to devote this session to some problem solving.

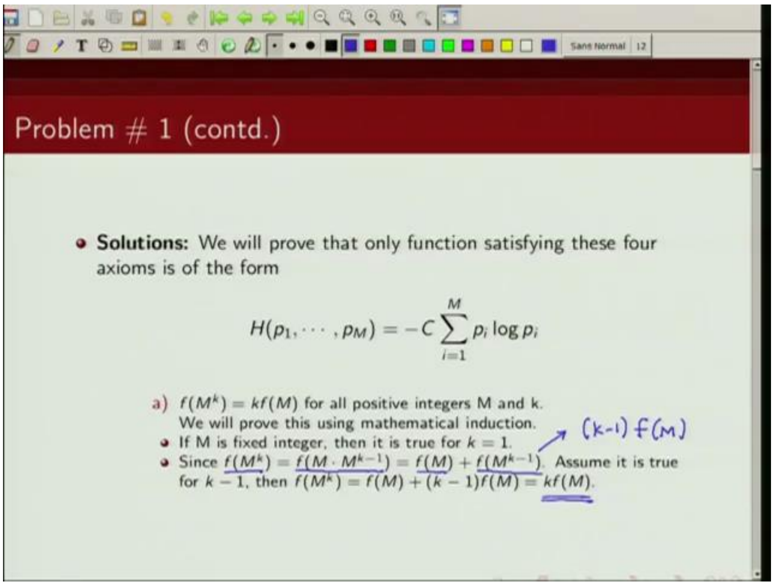
(Refer Slide Time: 00:36)



So, the first problem that we are going to solve is as follows. So, prove that entropy is the only function that satisfies the four conditions for information measure and what are those conditions. So, these are given in the form of axioms. So, assume we have a random source X that consists of M elements with probabilities given by p i. So, axiom 1 says if all the probabilities are equal, then the entropy function of our measure information should be monotonically increasing function of M. The second axiom says for two independent sources x and y, where x has M elements and y has L elements, then information measure f ML should be at the form f M plus f L.

Third axiom is what we call grouping axiom. So, uncertainty basically p 1 p 2 p 3 p m can be written as entropy of p 1 p 2 p r and p r plus 1 to p m plus p 1 p 2 p r. Entropy of p 1 divided by summation i 1 to r p i p 2 summation i going for 1 to r of p i and p r and similarly, plus p r plus 1 p r plus 2 up to p m and this is just entropy H of p r plus 1 divided by summation of p i, where i goes from r plus L to M. Similarly p f r plus 2 divide by summation of p i, where i goes from r plus 1 to M up to pfm divided by summation of p i where i goes from r plus L to M. This is a grouping axiom and the final axiom is basically is entropy function should be a continuous function of the probability set p i.

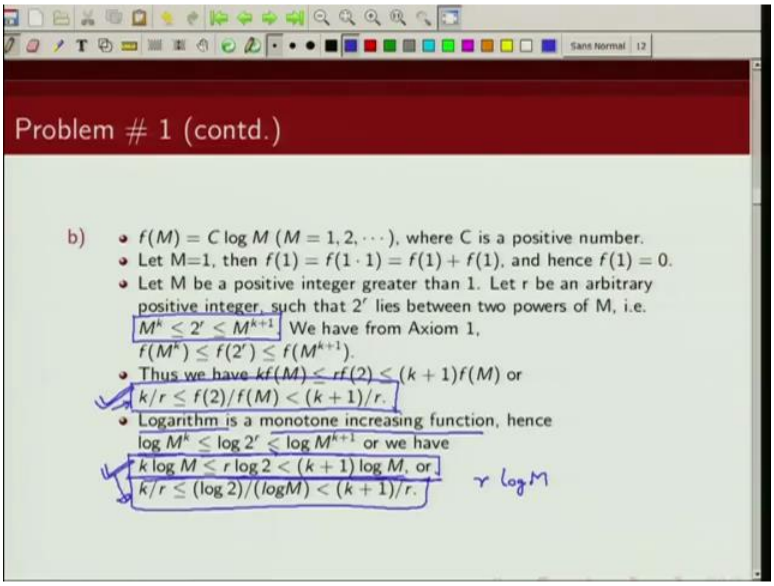
(Refer Slide Time: 03:34)



So, we are interested to prove that entropy is the only function that will satisfy these four conditions for information measure. So, we are going to show that only function that satisfies these axioms are of the form this. So, let us first prove this. So, function of M raise to power k should be k times f of M for all positive integers M and k. We can use mathematical induction to prove this. So, if M is a fixed integer, clearly for k equal to 1 that is true because f of M is equal to 1 times of fM, so that it holds true for n equal to 1. Let us say it holds for k equal to k minus 1. So, we want to prove that it holds for k for f of M k can be written as f of m into m k minus 1. Now, this from axiom 2 can be written as f of m plus f of k minus 1. Now, since this holds true for k minus 1, this would be given by k minus 1 f of m.

So, this when you add with f of M becomes k times f of M. So, this also holds for k. So, mathematical induction we have shown that this holds for positive number integers, M and k.

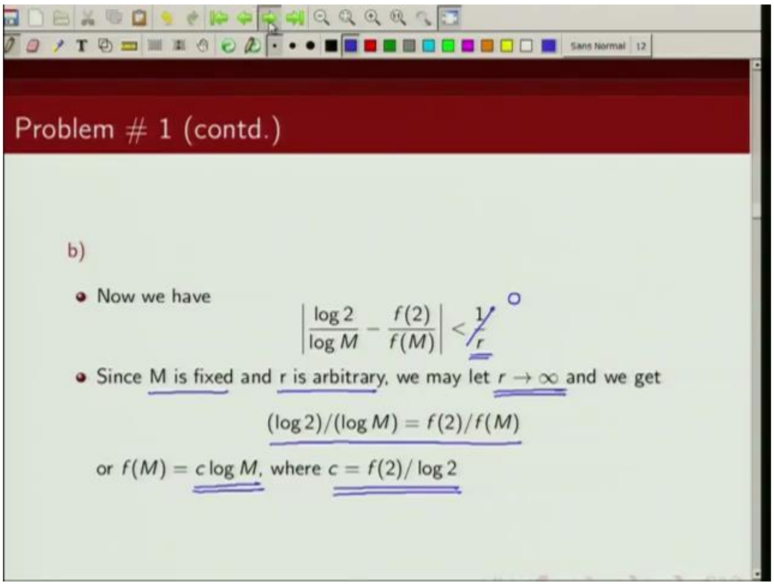
(Refer Slide Time: 05:10)



Now, f of M is the form C log of M. So, how do we show this? So, let us take M equal to 1. So, f of 1 is f of 1, but 1 which from axiom 2 can be written as f of 1 plus f of 1. So, we have f of 1 equal to f of 1 plus f of 1. Now, this holds true for only for case when f of 1 is 0. Next let M be a positive integer greater than 1 and let r be an arbitrary positive integer which is r in such a way, so that 2 raise power r lies between 2 powers of m. Now, what do you mean by that? So, there is some k such that M raise to power k is less that equal to 2 raise power r is less than equal to M raise to the power k plus 1. Now, we know from axiom 1 that information measures should be an increasing if all probabilities are same. It should be an increasing function of m. So, then from that we get f of M k should be less than equal to f of 2 r and that should be less than f of m raise to power k plus 1. So, this follows from axiom 1.

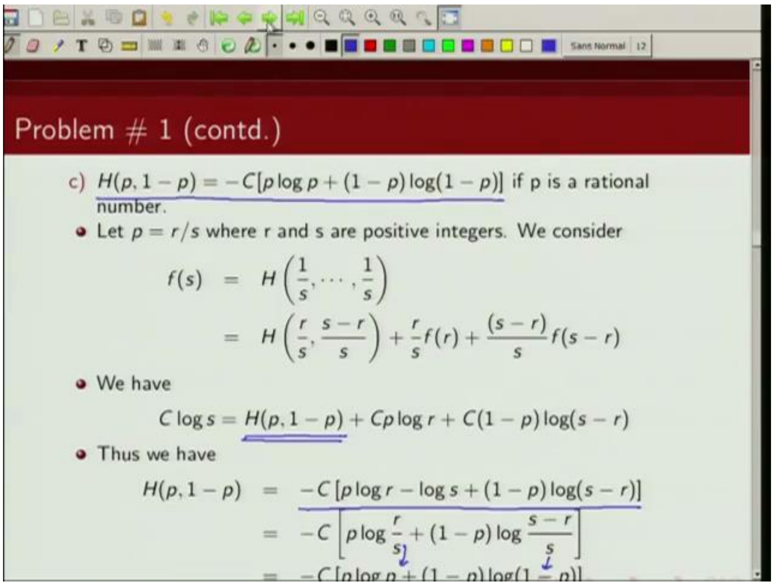
Now, what is f M k? We just showed if you recall f of M k is given by this, right. So, we can then write f of M k as k times f of M. We can write f of 2 raise to power r f r times f of 2 and f of m raise to power k plus 1 can be written as k plus 1 times f of M. Now, if you take this and divide it by all the clause by f of M times r, what we will get here is k by r is less than equal to f of 2 divide by f of M is less than k plus 1 divided by r. Now, we know that logarithm is also a monotonously increasing function. So, if we take log of this, just take log of this, what do we get? We get log of M k that is less than equal to log of 2 raise to power r is less than log of m raise to power k plus 1. Now, M raise to power k is k log m log of 2 raise to r is r log 2 and this is k plus 1 times log of m. Now, again we divide here by r times log of M. So, if you divide this by r times log of M, what I get is k by r is less than equal to log 2 by log M is less than k plus 1 divide by r.

(Refer Slide Time: 09:53)



So, from here I got k by r is less than equal to f of 2 by f of m is less than k plus 1 divide by r and here, I got k by r is less than equal to log 2 by log M is less than k plus 1 divided by r. So, from these two, I can write that absolute difference between log of 2 by log of m minus f of 2 by f of M must be less than 1 by r. So, this follows from this result and this result this one, this result and this result. Now M is fixed, but we can make r very large. If we make r very large, this 1 by r will become very small. So, then this will tend to 0 as r goes to infinity. This will tend towards 0. Then, basically what we will get is log of 2 by log of M is equal to f of 2 divide by f of M or in other words, f of M will be at the form c of log M, where c is f of 2 divide by log of 2.

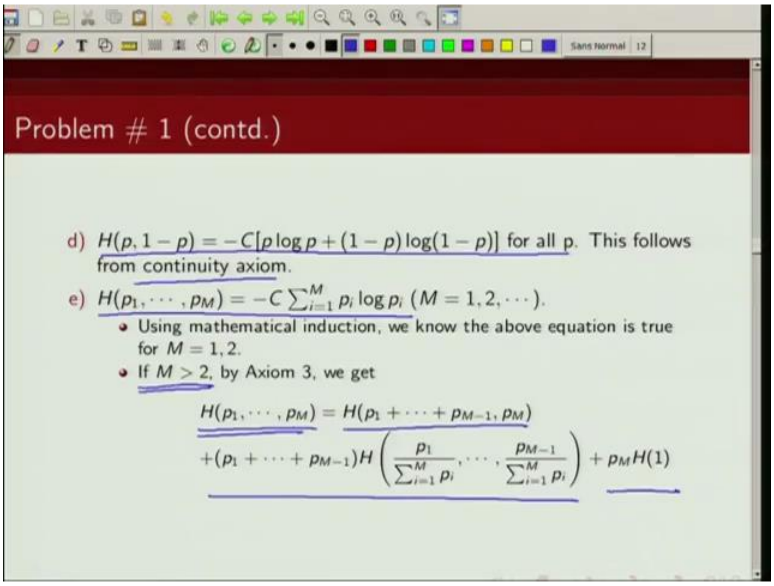
(Refer Slide Time: 11:20)



Now, ensure the entropy of p 1 minus p is the form this, where p is a rational number. So, let us take p to be equal to r by s, where r and s are positive integers. Now, from axiom 1 we can write f of s like this and from z1, we can write this entropy in this particular fashion now r by s is p. So, this p, this is 1 minus p, this is p. I mean this is 1 minus p. So, this we can write basically f of s of the form C log of this will become of the form like this. So, f of s is a form of c log s f of r form c log of r. So, what we get is this.

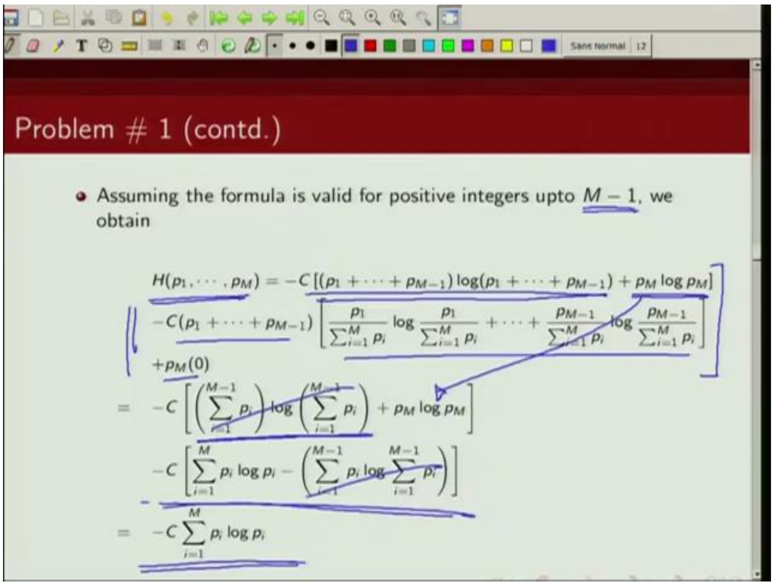
Next, from here we can write entropy of p and 1 minus p s. So, if we bring all of them here, what we get is something of this form and we get terms containing 1 minus p. If you do that, we get this of the form minus some constant times p log r by s, where r by s is this p 1 minus p log of 1 minus r by s this is 1 minus p. So, what we get is this is given by s.

(Refer Slide Time: 13:29)



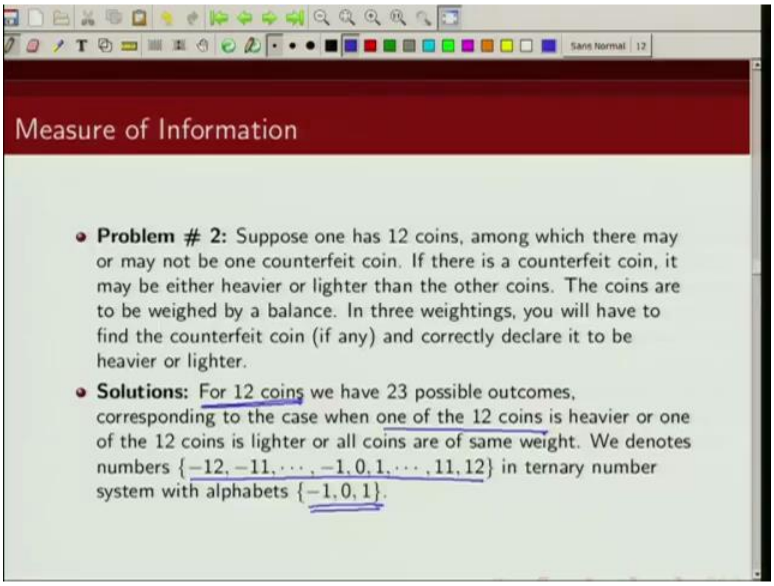
This holds true for all p. This follows from the continuity axiom. Now, finally we are going to show that joint entropy of p 1 p 2 p m can be written of the form minus c times summation p i log p i, where i goes for 1 2 M. Now, we are going to use mathematical induction to prove this. So, we can see very easily this holds for M equal to 1. If m equal to 1 p i basically just 1 that it hold for M equal to 1, it also holds for M equal to 2. So, let see what happens when M is greater than 2. Now, this joint entropy using grouping axiom can be written as this, plus this and plus this.

(Refer Slide Time: 14:27)



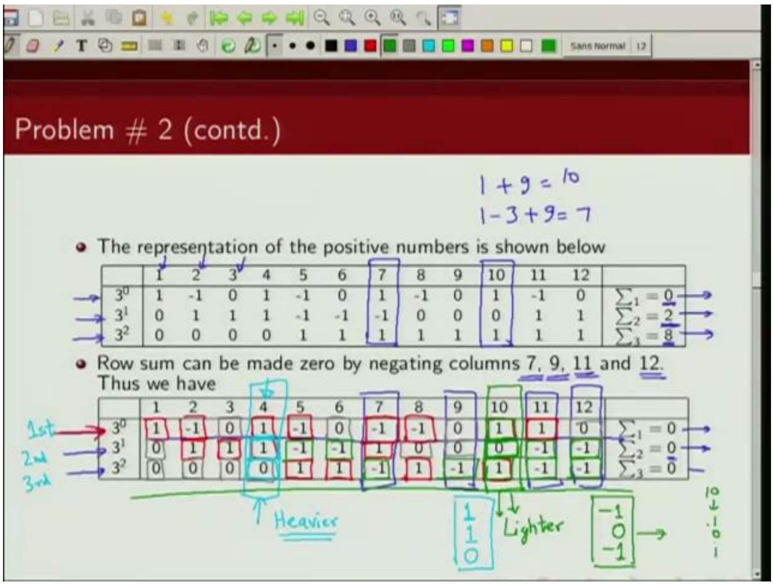
Now, assuming this holds true for some k up to M minus 1, we will try to see whether this holds for k equal to n. So, this can be written as now we are making use of if grouping axiom and we can write this as minus c times this term plus this minus c times, this term plus this plus p m and this h of 1 was 0. So, now we know that this term is given by this because this holds for M minus 1. So, this is the expression. Similarly, this expression comes here and this will be summation p i to the power minus M and the summation from p and this will be p of this will be 0. So, if you look at this term essentially what we are going to get is, of the form this will be some c times p i log of p i and there will be some term minus summation p i p i log p i. So, this term basically comes out to be this. You can verify this. Now, if you look at this, this term you can take this common out this, this cancels. So, what we have is minus c summation p i log p i and this is p M log p M. So, this becomes this. So, what we have shown is and entropy function basically satisfies the axioms of information measure.

(Refer Slide Time: 17:05)



The next question that we are going to solve is, so we are given 12 coins. Now, we do not know whether they are counterfeit coins or not. So, there might be some counterfeit coins. If there are counterfeit coins, the coin can be heavy or it can be lighter also that also you do not know now. So, you are supposed to weigh these coins using a weighed balance. You are going to weigh these coins to give a weighed balance and in 3 weighing, you will have to find out the counterfeit coin number 1. Second you have to correctly declare whether the counterfeit coin is heavy or lighter. So, how do we solve this? So, you have 12 coins and there are 23 possibilities. So, we are denoting these possibilities. So, these possibilities corresponds to the coin to the fact that one of these coins could be heavier or one of these coins could be lighter or all coins are of equal weight and we are denoting by these numbers, these 23 possibilities and we are going to write these numbers using ternary alphabet minus 1 0 and 1.

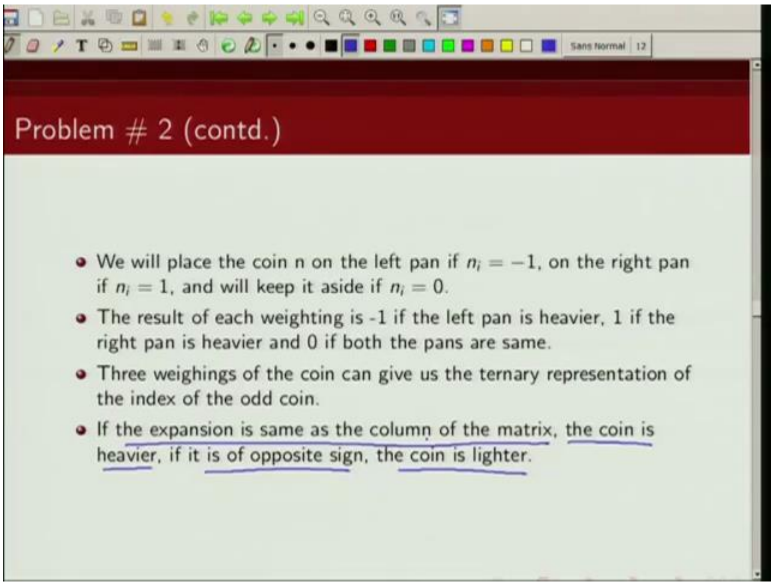
(Refer Slide Time: 18:50)



So, these 12 coins I am just writing this 12 in equivalent ternary notation. So, 1 would be 1 times 3 is to power 0 plus 0 times 3 is to power 1 plus 0 times 3 is to power 2. Similarly, you can see this 7 is nothing, but 1 minus 3 plus 9 plus 7. So, you can write any number. Let us take 10 10. I can write as 1 plus 9 that is 10. So, that is 1 times 3 is to power 0 plus 1 times 3 is to power 2 now. So, I have written all of these 12 numbers in the ternary notation minus 1 0 and 1. Now, if I sum up in each unit, it corresponds to 3 0e 3 is to power 1 3 is to 2. If I sum them up, I get here 0 2 and 8.

Now, what I do is, I negate columns 9 7 7 9 11 and 12. So, what I did was. So, 7 was 1 minus 1 1. I made it minus 1 1 and minus 1. Similarly, 9 was 0 0 1. I made it 0 0 minus 1 11 was minus 1 1 1. I made it 1 minus 1 minus 1 and 12 was 0 1 1. I made it 0 minus 1 minus 1. What I am doing is, you can see here when I sum up these numbers, I get 0. So, this row I sum them up, I am getting 0. Here I was getting earlier also 0, but if you look at these two rows, earlier I was getting 2 and 8, but now as a result of negating of these few columns, I am getting 0.

(Refer Slide Time: 21:08)



Next how I am going to weigh them? So, I am going to place a coin n on the left pan if n i is minus 1 and I am going to put it, on the right pan if n i is 1 and if n i is 0. I am not going to put it in the any pan. So, for example we are going to weigh it 3 times plus corresponding to this row, second time corresponding to this row, third time corresponding to this row. So, when I am doing this weighing corresponding to this row, I am going to put this in the right pan and I am going to put this in the left pan. I am going to put this in the right pan, I am going to put this in the left pan, I am going to put coin number 7 in the left pan, coin number 8 in the left pan, coin number 10 in the right pan, coin number 11 in the right pan and one's that you see which are 0, that coin number 3, coin number 6, coin number 9, coin number 10, 12, I am not going to put in any of the pan. This is for the first weighing.

Similarly for the second weighing, I am going to put coin number 2 3 4 7 in right pan. I am going to put 5 6 11 and 12 in the left pan and I am not going to put coin number 1 8 9 and 10. Likewise for a third weighing I am going to put coin number 5 6 8 10 in the right pan and coin number 7 9 11 and 12 in the left pan. Why coin number 1 2 3 4 are not put in any pan? Now, the claim I am making is the result of each weighing is if the result of each weighing is minus 1, it means the left pan is heavier and if the result of weighing is 1, it means the right pan is heavier and if it is 0, then both the pans are of same weight. So, what I am going to do is, if I weigh them and I find the left pan is heavier, I am going to give number minus 1. If right pan is heavier, I am going to give number 1 and if both the pans are of same weight, I am going to give number 0. This is a result of my weighing.

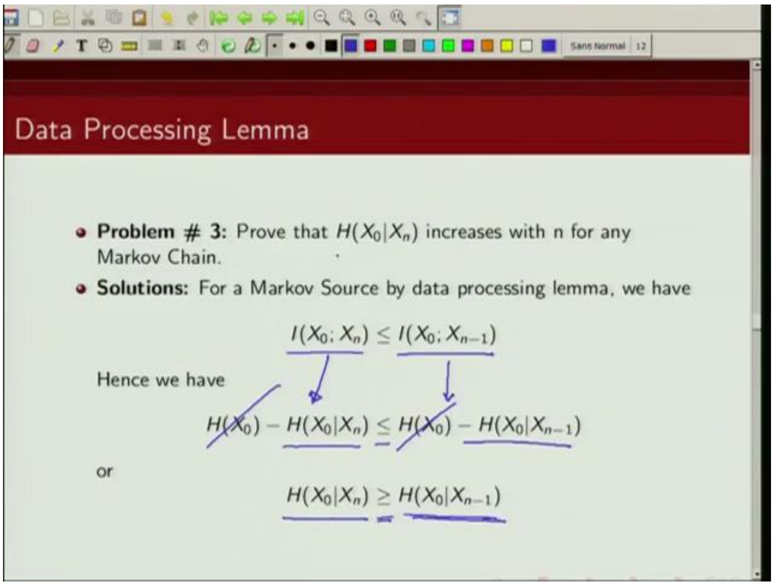
Now, three weighing of the coin can give me the ternary representation of the index of the odd coin. How if the expansion is same as the column of the matrix? Then the coin is heavier and if it is of opposite sign, then the coin is lighter. So, what I am saying is as a result of my weighing when I give this numbering minus 1 1 and 0 depending upon which pan is heavier if that numbering comes out to be same as is that expansion comes out to be same as a column of this modified matrix that I had, then the coin is heavier and if it is of the opposite sign, then the coin is lighter. So, let us stick to this result and pick up one of the coin. Let us see we pick coin number 4 and let us see this is heavier. Let us say coin number 4 is heavier. Now, if coin number 4 is heavier when we do this first weighing, what would you notice because this expansion is 1? So, we would have put it in the right pan and since coin n umber 4 is heavier, the right pan would have come out to be heavier.

So, the result of outcome would have been 1. Similarly for the second weighing corresponding to this row, you should look at coins, 4th coin this is 1. So, we would have put this coin in the right pan and then, again outcome of our weighing would have been 1 in the final weighing which is corresponding to the third weighing. Since this is 0, we would not have put this coin in our weigh. So, we would have observed that both left and right pan are of same weight. So, the result of our outcome of weighing would have been 1 1 0. Now, if you compare 1 1 0 with the columns of this matrix, we see that this column corresponding to coin 4 has the same expansion and as I said. So, we are able to identify the odd coin is coin number 4 and since the expansion of this is same as expansion of column 4 in this matrix, the column 4 is heavier which it indeed was.

Now, let us take another example. Let us take coin number 10 and let us say this coin is lighter like this coin is lighter. So, in the first weighing this is 1. So, we would have put it on the right pan since this is the lighter coin and left would have come out as heavier. So, result of first weighing would have been minus 1. Now, the second weighing and since the expansion corresponding the term here is 0, we would not have put this coin anywhere. So, result of weighing would have been that, both left and right pan are of equal weight.

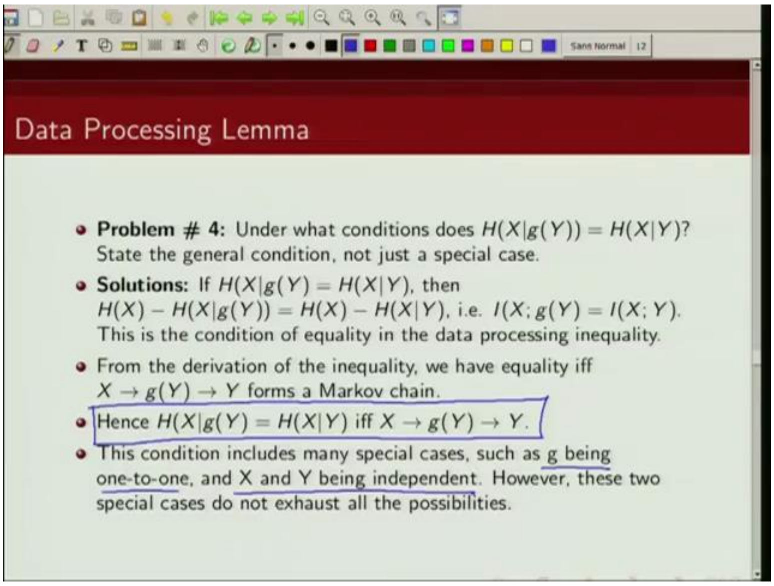
So, then this result would have been 0 and finally, for the third weighing since this is 1 here, we would have put this coin in the right pan. So, left pan would have come out as heavier. So, this result of our outcome of weighing would have been minus 1. For the third weighing now look at do we have any column of this matrix which of this expansion minus 1 0 minus 1. We do not, but if you look at expansion for 10 if you multiply it by minus 1, you get this. So, then we are able to identify that the coin number 10 is counterfeit and we are also able to identify because the expansion of 10 is 1 0 1. So, and this is minus 1 0 minus 1. So, we are also able to identify that not only coin 10 is counterfeit, but it is of lighter weight.

(Refer Slide Time: 29:35)



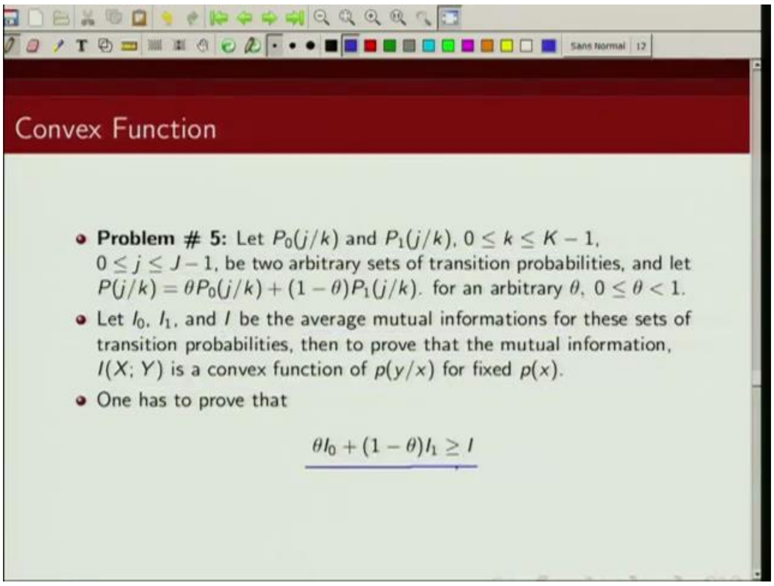
The next problem that we are going to solve is as follows. So, we want to show that entropy of x 0 given x n, it increases with n for a Markov chain. So, if x 0 x 1 x 2 x n forms a Markov chain, then we know x 0 x 1 x 2, they form a Markov chain and then from the data processing lemma, we know that mutual information between x 0 and x n is less than mutual information between x 0 and x n minus 1 because we know that further processing of data does not increase the mutual information. So, then from the definition information, we can write this as uncertainty in x 0 minus uncertainty in x 0 given x n. Similarly, this we can write as uncertainty in x 0 minus uncertainty in x 0 given x n minus 1. So, this is common. So, what we get is minus of this is less than equal to minus of this. So, then multiply by minus both side, we get entropy of x 0 given x n is greater than equal to uncertainty in x 0 given x n minus 1. So, as n increases, you can see this entropy increases.

(Refer Slide Time: 31:17)



The next question is under what condition does uncertainty in x given g of y is same as uncertainty in x given y. So, if uncertainty in x given g of y is same as uncertainty in x given to y, then h of x minus h of x given g y can be written as h of x minus h of x given y or in other words, the mutual information between x and g y is same as mutual information between x and y. Now, we know x y and g of y forms a Markov chain, then mutual information between from data processing lemma we know mutual information between x and g y should be less than equal to mutual information between x and y and here, it is given it is equal and when does equality happen, the equality happens when x g y and y also forms a Markov chain. So, this result comes from the property of data processing lemma. So, we know that this equality will happen if g x g y and g also forms a Markov chain and that is a general condition. I am talking about under which uncertainty in x given g y is equal to uncertainty in x given y and there are many special cases. For example, g being 1 to 1 x and y independent, but the general condition is this, ok.

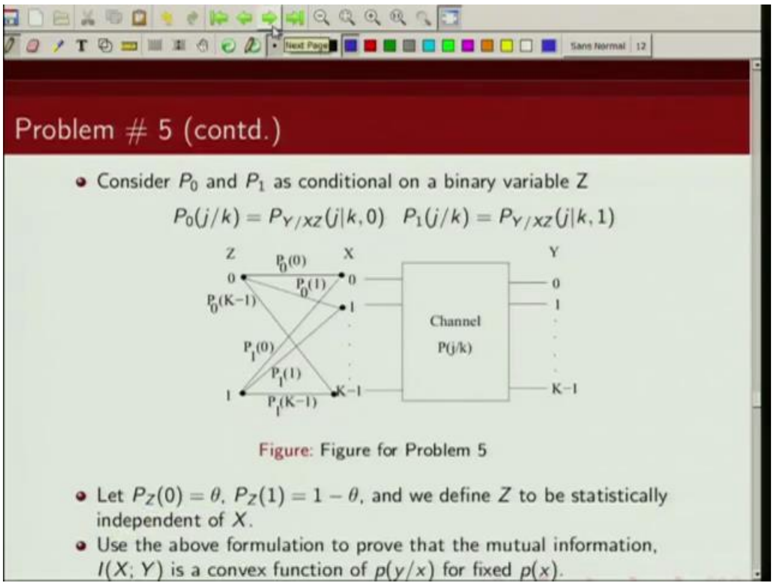
(Refer Slide Time: 33:43)



The next question is on proving whether a function is convex or concave. We want to show that mutual information is a convex function of p of y given x for a fixed p of x and how are we going to prove this. So, let's p 0 j by k and p 1 j by k, where k varies from 0 to k minus 1 and j varies from 0 to j minus 1 by 2 arbitrary set of transition probabilities. It defined p of j k as theta times p 0 j given k minus 1 minus theta times p 1 j given k for some theta which lies between 0 and 1. Now, let i 0 i 1 and i be the average mutual information for these three sets of transition probabilities and then, p 0 p 1 and this p j.

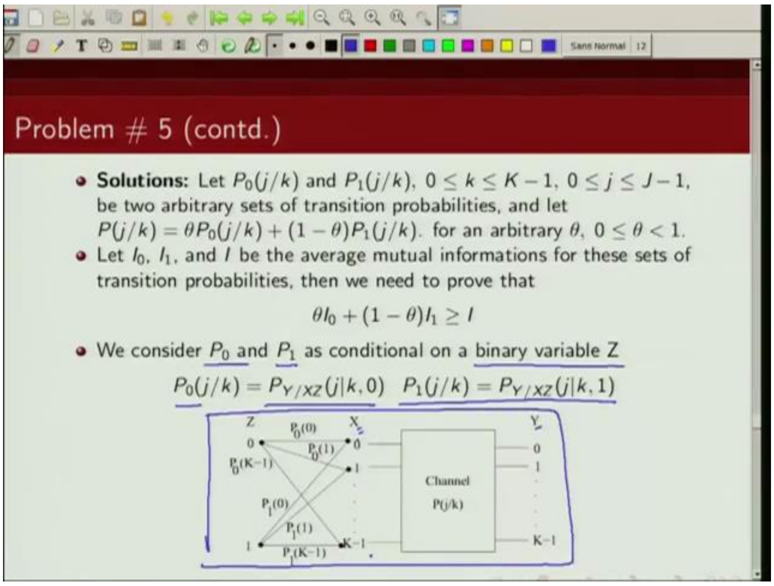
Now, we want to prove that mutual information is a convex function of p of y given x for a fixed p of x. So, to prove it, we have to show that theta times i 0 plus 1 minus theta times i 1 that is greater than equal to this average mutual information corresponding to this transition probability.

(Refer Slide Time: 35:26)



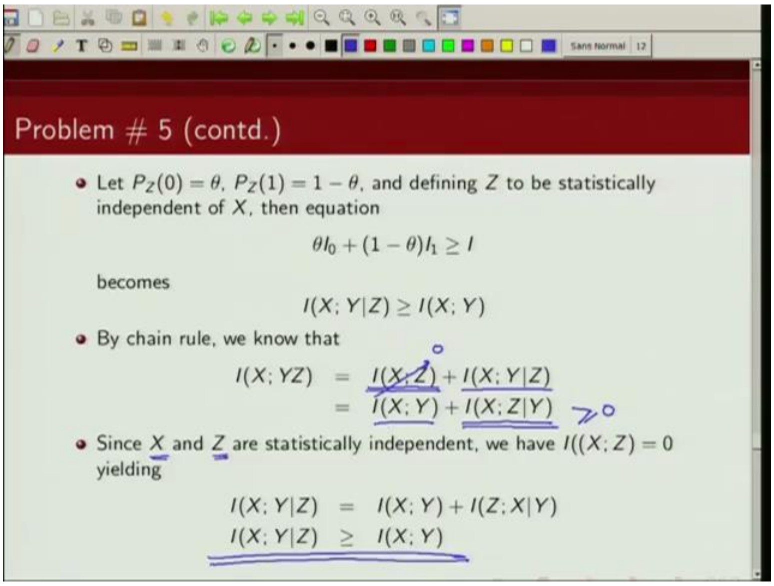
So, let us consider p 0 and p 1 to be conditioned on some binary random variable z. So, we have a binary random variable z. It takes value 0 and 1 and p of 0 can be written as probability of y given x and z, where z is 0 and probability p of 1 can be written as probability p of y given x z, where z is plus 1. Let p of z being 0 is given by theta. So, p of z being 1 will be 1 minus theta and z is considered to be statistically independent of x. Now, we want to show that mutual information is a convex function of p of y for a fixed p of x.

(Refer Slide Time: 36:28)



So, let's see as I said we have to prove this. We already showed that we are considering p 0 and p 1 as conditioned on this binary random variable z. So, p 0 can be written like this and p 1 can be written like this. This is a diagram corresponding to that to think of it. This x and y that condition probability of y given x is condition on probability y given x and z whereas, z is 0 z is 1 and these are the channel transition probabilities.

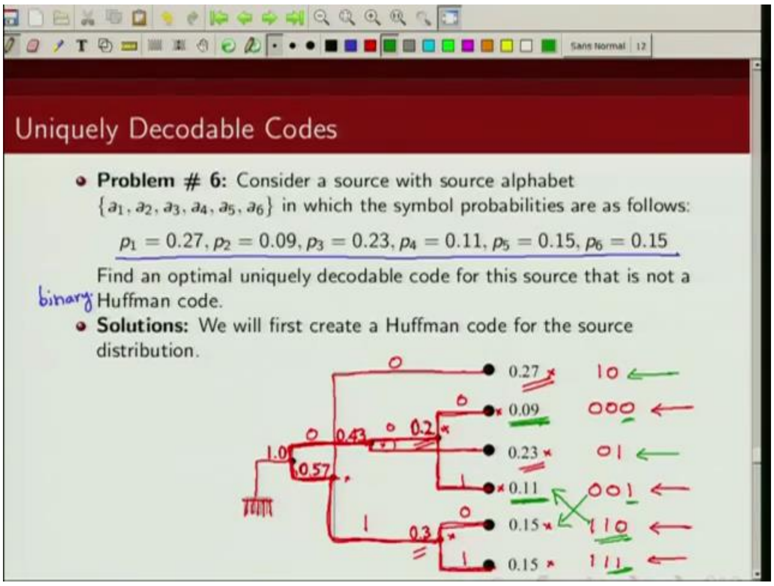
(Refer Slide Time: 37:07)



Now, what is this term? Theta is nothing, but p of z being 0 and i of 0 is corresponding to this transition probability p of 0 and 1 minus theta as a probability of p of z being 1 and this is the mutual information corresponding to transition probability p of 1. So, this can be written as mutual information being x and y given z. What about this? This is nothing, but mutual information between x and y. So, to prove that mutual information is a convex function of p of y given z for a fixed p of x, we will have to show that this relation holds.

Now, using chain rule we can write mutual information being x and y z as mutual information being x and z and mutual information plus mutual information between x and y, given z. Now, if you apply chain rule in another way, we get this as mutual information being x and y plus mutual information between x and z, given y. Now, since x and z are statistically independent mutual information between x and z will be 0. So, this term will be 0. So, then what we get is mutual information be x and y, given z is equal to mutual information between x and y plus mutual information between x and z, given y and we know that mutual information is greater than equal to 0. So, this shows that mutual information between x and y, given z is greater than equal to mutual information between x and y. So, we have proved that this relation holds and hence, we have proved that mutual information is a convex function of p of y given x for a given p of x.

(Refer Slide Time: 39:37)



Now, in this question you have been asked to design an optimal uniquely decodable code which is not a Huffman code. Now what are the properties? So, here you have been given a source, the source alphabet is given by this and the symbol probabilities are given by this. I forgot to mention basically I mean which is not a binary Huffman code. Now, what are the properties of a binary Huffman code? We know that there are two light symbols and they differ in only one bit location and when I see find an optimal uniquely decodable code, I want the expected code word length of this uniquely decodable code to be same as that of Huffman code. However, it is not Huffman code. So, how do we solve this?

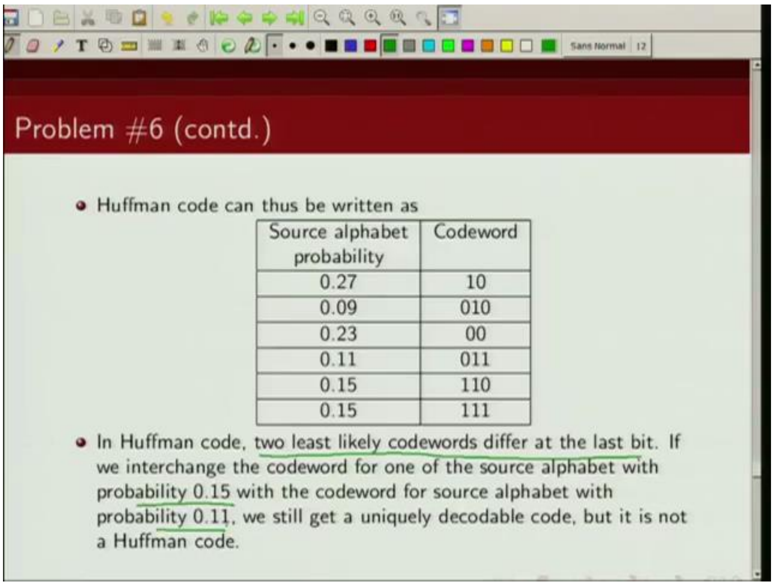
So, first we are going to do is, we are going to create a Huffman code using these symbol probabilities. So, you can see here two smallest symbols. If probability of this is 0.01 and o.11, so join them get probability 0.2. So, I deactivate these nodes and activate this node. Now, the two nodes with least probability are this time is point this one which has probability 0.15 and this one which is probability 0.15. So, I deactivate these nodes and this is probability 0.3.

Next I have this is 0.3, this has 0.23, this one has 0.2 and this one has 0.27. So, the two least lightly active nodes are this one ith probability 0.2 and this one with probability 0.23. So, I deactivate these nodes. Now, I have this node with probability 0.43. Now, among the three nodes, remaining this one has probability 0.27 and this one has probability 0.3. I join them, I get this as node as probability 0.57. I deactivate these nodes and finally, I join this node and this node and I get here, this is probability root is probability 1 and deactivate this.

So, if you now assign code words, you can assign 0 1 0 0 1 0 1 0 1 0 1 0 1. So, what you will notice is, this has code word 1 0, this is 1 0, this is 1 and this is 0, this one has code word. Let's see 0, this one 0 and then 0 and 0, this one has 0 and then 1 0 and 1, this has 0 and then 0 and then 1. So, 0 0 1, this has 1 1 0 1 1 0 and this one has 1 1 and 1. So, if you notice in the optimal Huffman coding for this source with these source probabilities, I have four code words of length 4 and I have two code words of length 2 and remember the two least lightly symbols which are probability 0.09 and 0.11, they are differing only in one bit location, correct.

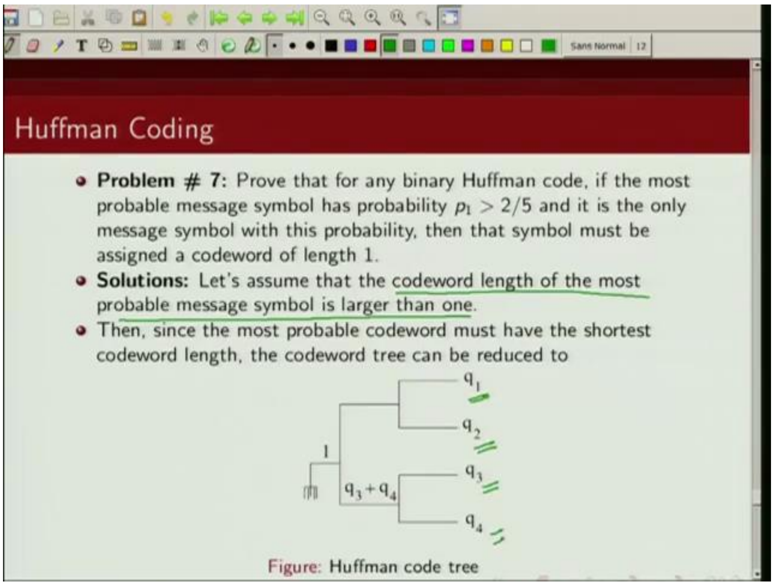
Now, if I swap these two code words are also having 3 bits, if I swap any one of them with, so that will if I make this code word as 0.001 and I make this one as 1 1 0, then it does not change the expected code word length. However, it is now no longer Huffman code. Why? It is because now the least lightly decode word this one has 0 0 0 and this one has 1 1 0. So, two least lightly code words are differing in two positions. So, it is no longer a Huffman code. So, this way you can solve this problem.

(Refer Slide Time: 45:29)



This is a Huffman code as I said and in case of Huffman code, two least lightly code words differs only in the last bit. So, if we interchange the code word for one of the source alphabet with probability 0.15 to the code word for source alphabet with 0.1, we still get a uniquely decodable code and it is not Huffman code. In fact, we get a prefix free code, but it is not a Huffman code. One such example is this.

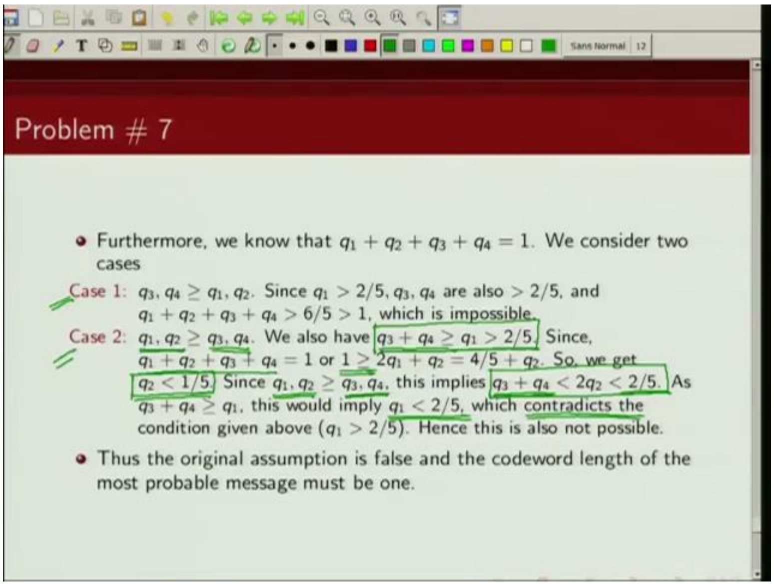
(Refer Slide Time: 46:07)



The next problem is prove that for a binary Huffman code, if the most lightly message symbol has probability greater than 2 by 5, then it is the only message symbol and it is the only message symbol with this probability, then that symbol must be assigned a code word of length 1. So, what I am saying is the most probable message symbol has probability greater than 2 by 5 and it is the only message symbol which has this probability, then show that this symbol must be assigned a code word of length 1.

Now, we are going to use method of contradiction to prove this result now does method of contradiction work. So, we are going to assume that lets say this particular message symbol requires more than 1 length code word and then, later on we will show that this is not possible. So, our initial assumption that this message symbol has code word of length greater than 1 is incorrect and that is how we will prove it. So, let us assume that code word length of the most probable message symbol is larger than 1. If that is a case, this is a kind of Huffman tree, binary Huffman tree you will get. So, I am just referring to this message symbol by has q 1 here of q 1 q 2 q 3 q 4.

(Refer Slide Time: 48:02)

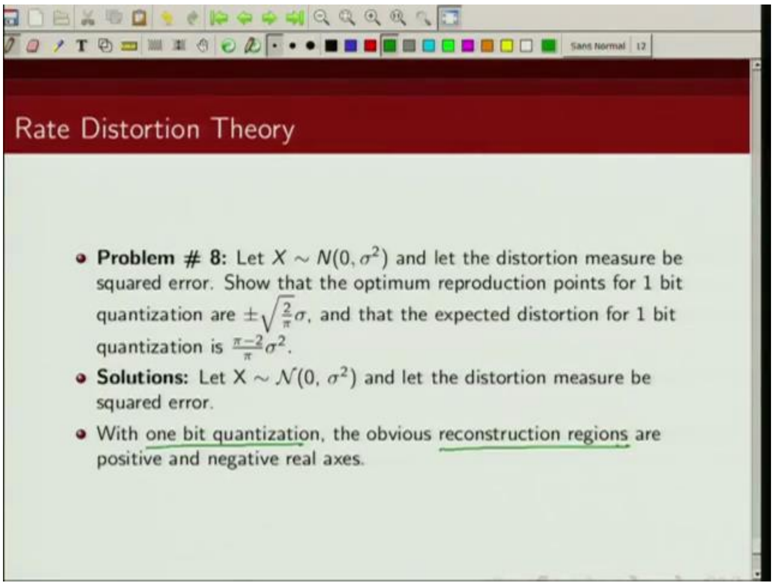


Now, these probabilities q 1 q 2 q 3 q 4, they should add up to 1, right because some of probability of the root should be 1. So, let us say I have at depth 2. This is how I have these probabilities q 1 q 2 q 3 q 4. Now, consider two cases. In case 1, we consider that q 3 and q 4 their probability is more than q 1 and q 2. So, what we are considering is that these probabilities are more than these probabilities that the first case and the second case we will assume these two probabilities are more than these probabilities. So, if q 3 and q 4 are more than q 1 and q 2 and since q 1 that message symbol with probability 2 by 5 is there q 1, so q 3 and q 4 must also be greater than 2 by now. If we add up q 1 plus q 2 plus q 3 plus q 4, we get because q 3 and q 4 are 2 by 2 by 5 and q 1 is also 2 by 5. So, q 1 plus q 3 plus q 4 is greater than 6 by 5 which is not possible. So, we cannot have this case 1.

Now, let us look at case 2. So, here we assume q 1 and q 2 is greater than q 3 and q 4. Now, we also have q 3 plus q 4 to be greater than q 1, otherwise we would have joined them. So, q 1 plus q 4 is greater than q 1 which is greater than 2 by 5. Now, since q 1 plus q 2 plus q 3 plus q 4 should add up to 1, what we get from here is two times q 1 plus q 2 is less than equal to 1. So, 4 by 5 plus q 2 is less than equal to 1 or we get q 2 is less than 1 by 5. Now, since q 1 and q 3 are greater than q 3 and q 4, q 3 plus q 4 if we add them up, they must be less than two times q 2. What is q 2? In this case, q 2 is less than 1 by 5. So, then if this condition works, what we have shown is q 3 plus q 4 is less than 2 by 5, however you also said q 3 plus q 4 is greater than q 1 which is greater than 2 by 5, but here we are getting q 3 plus q 4 is less than 2 by 5.

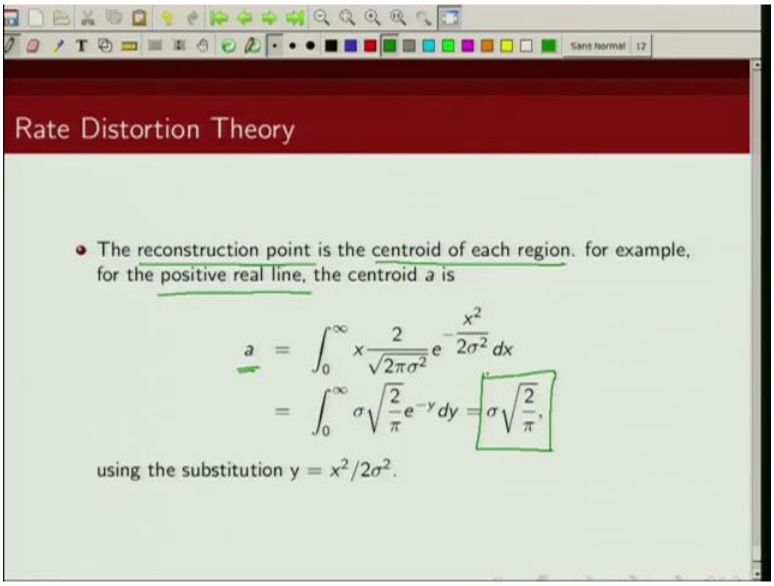
So, this contradicts because from here we are getting condition that q 1 is less than 2 by 5, but we are given that q 1 has probability greater than 2 by 5. So, hence this contradicts our assumption that the message symbol of the probability 2 by 5 is a sign code word of length greater than 1 since we have considered all possible cases. So, we know that it is not possible for a message symbol with probability greater than to 2 by 5 and if that is the only message symbol with that probability, it is not possible for this message symbol to have a code word greater than length 1. So, since the original assumption is false, by method of contradiction we know that the code word length of this message symbol has to be 1.

(Refer Slide Time: 52:31)



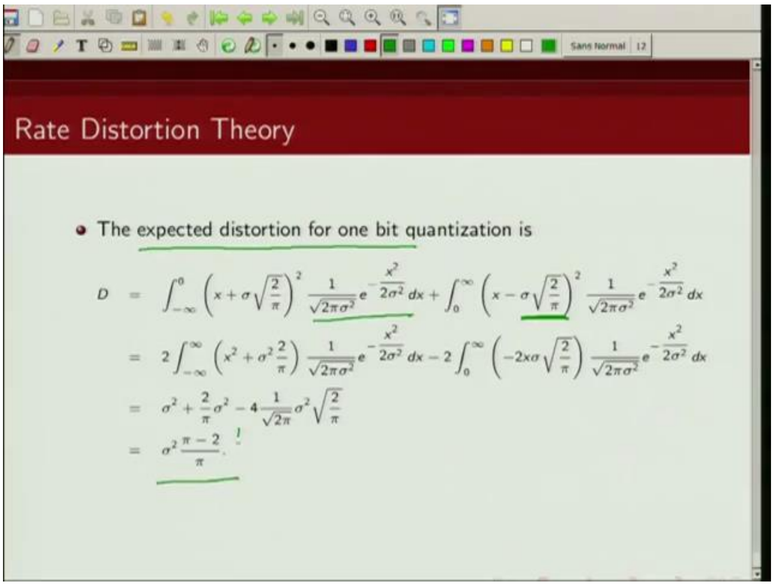
Finally we conclude with one example. So, we have a Gaussian source with 0 mean variance sigma square and let the distortion measure be squared distance. So, we have been asked to show that the optimum reproduction point for 1 bit quantizer is given by this and the expected distortion for 1 bit quantizer is given by this expression. So, this is a Gaussian distribution random variable, right. It is zero mean Gaussian distribution random variable symmetric across around 0. Now, with 1 bit quantizer, it is obvious that the reconstruction regions are one is negative real axis and other one is positive real axis.

(Refer Slide Time: 53:53)



So, to find out the optimal reproduction point, we need to find the centroid of these regions. So, the reconstruction point is nothing, but centroid of this negative of region or the positive of region and they are symmetric, right. So, if we consider example, the real axis, then the centroid point comes out to be, this is integration from 0 to infinity two times exp1ntial of minus x square by 2 sigma square divided by and the root 2 pi sigma square and this comes out to be sigma times under root 2 by pi. Similarly for the negative real axis, the centroid point will come out to be minus of sigma under root 2 by pi.

(Refer Slide Time: 51:43)



Now, to find out the expected distortion, for the real negative axis the reproduction point is minus sigma under root 2 2 by pi for whereas, for the positive, this is the reproduction point. So, distortion is given by x minus, minus of sigma under root 2 by pi and since, the distortion is squared error, we take square of this. There is PDF of the source integrating for minus equal to 0. Similarly, this is the centroid point for the positive real axis. So, we integrate from 0 to infinity and there is PDF of the source. So, by these algebraic manipulations, what we get finally is this that the expected distortion is given by sigma square into pi minus 2 divided by pi. So, with this we will conclude this lecture.

Thank you.