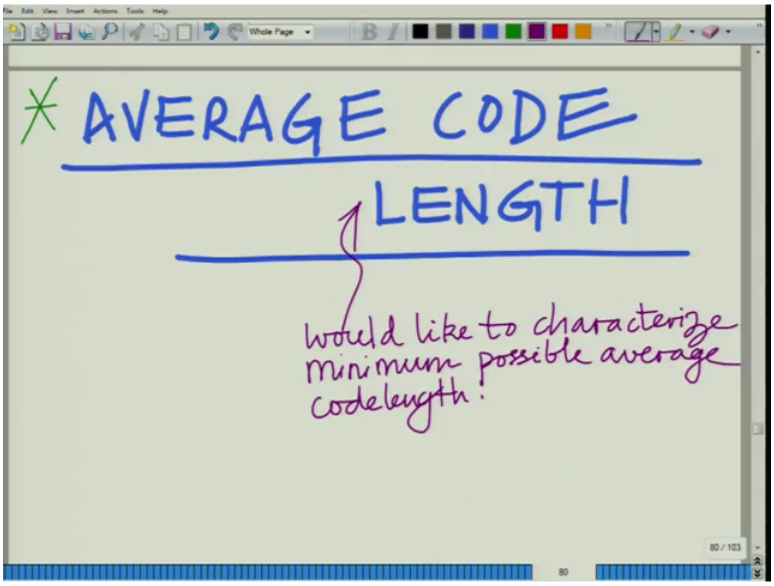
Principles of Communication Systems - Part II
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 44
Lower Bound on Average Code Length, Kullback-Leibler Divergence

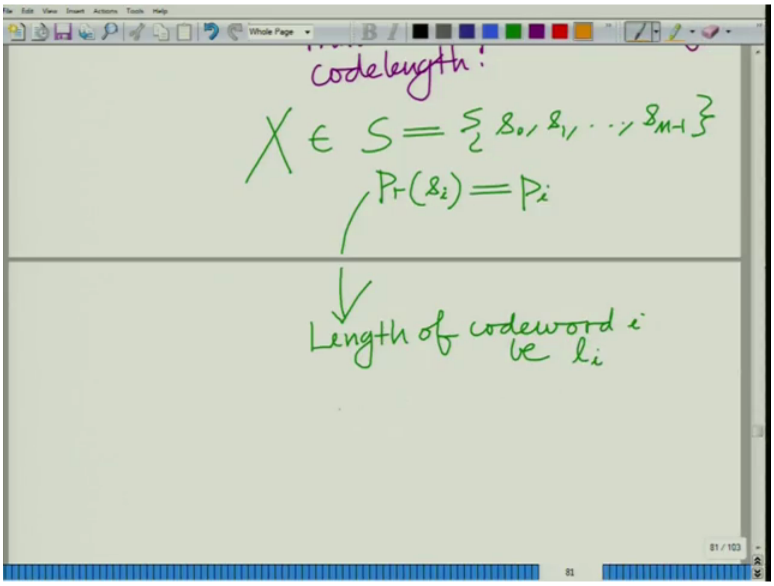
Hello. Welcome to another module in this massive open online course. So, we are looking at source coding, various aspects of source coding, different kinds of source codes in particular. We are interested in prefix free or instantaneous codes and in the previous module, we have looked at an important inequality that has to be satisfied by the lengths of the code words of a prefix free code that is given by the kraft inequality. We will use this kraft inequality to derive now a fundamental bound on the minimum possible average code length for a prefix free code, ok.

(Refer Slide Time: 00:50)



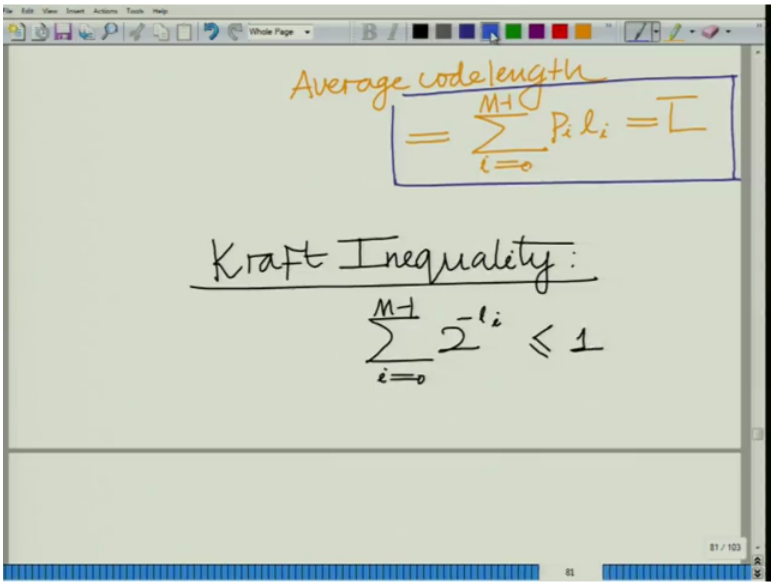
So, in this module, we would like to focus again on the average code length. As I have specified before, this is one of the fundamental aspects of a code that is we would like to characterize correct, we would like to characterize the minimum possible average code length. What is the minimum possible average code length? We would like to ask this question again towards this.

(Refer Slide Time: 01:54)



We would like to consider X drawn from a source with the alphabet s naught s1 up to sm minus 1 and we have probability of each symbol s i is equal to probability of each alphabet s i is purely symbol s i is p i. Now, the length we have seen, we have defined average length of the codeword.

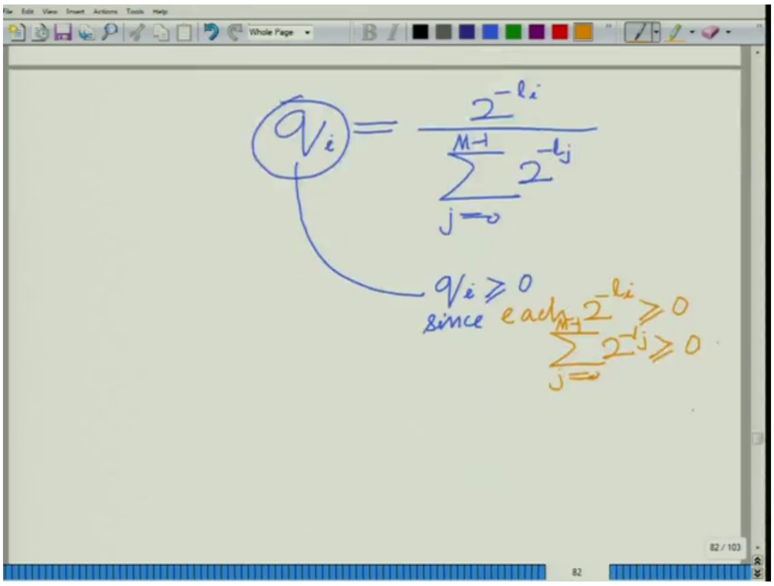
(Refer Slide Time: 02:53)



Let the length of codeword be l i, then we know that for the average code length or average codeword length is nothing, but expected value of l which is i equal to 0 to m minus 1 p i, that is the summation of the lengths l i weighted by the probabilities p i, correct. This is expected value of, this is the average codeword length which we are denoting by l bar. We have seen this before that the average codeword length is summation p i l i equal to 0 to M minus 1.

Yesterday we have also defined, we have also seen the kraft inequality that the codeword lengths of any prefix free code have to satisfy and the kraft inequality, this is a fundamental inequality which we had derived from a binary representation. Let me just refresh your memory. So, I have i equal to 0 to M minus 1 to the power of minus l i less than or equal to 1.

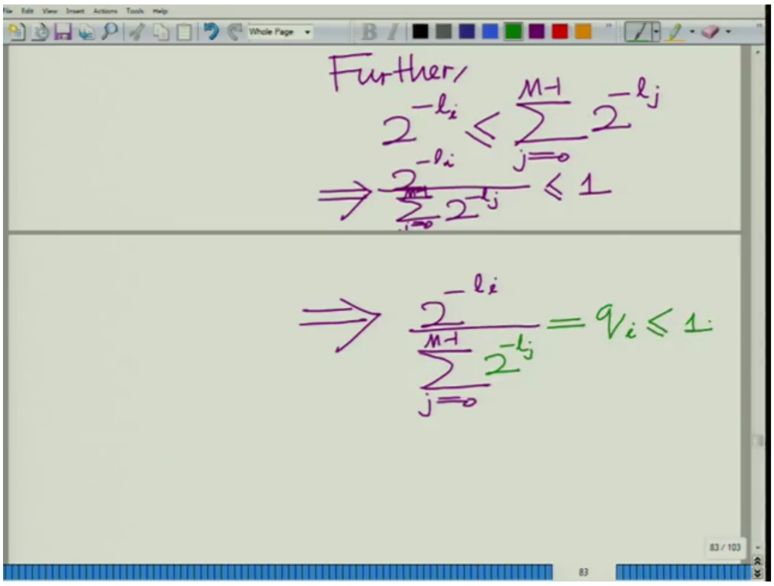
(Refer Slide Time: 4:20)



Now, what we would like to define is, we would like to define this quantity q i which is equal to 2 to the power of minus l i by summation j equal to 0 to M minus 1 2 to the power of minus l j. Just changing the index using a different index, instead of I am using j because I am using i equal to j equal to 0, j equal to 0 to M minus 1 2 to the power of minus l j.

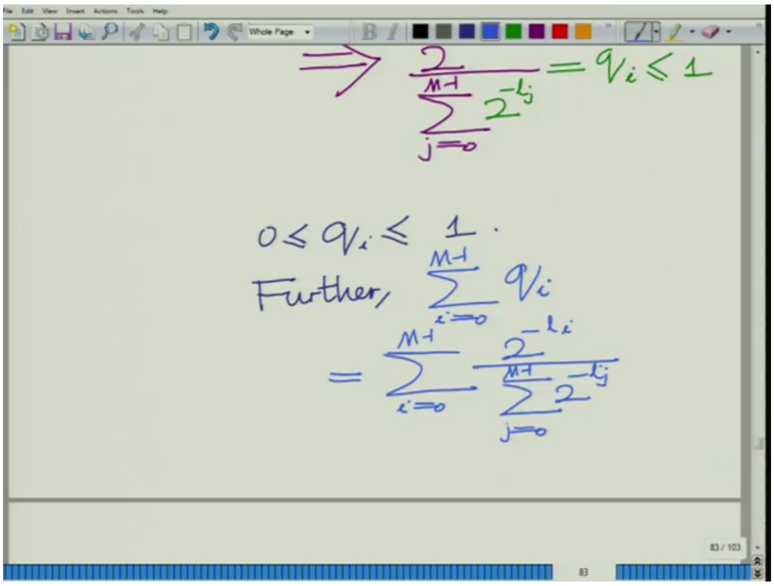
Now, what can we say about this q i? Now, we have defined this q i. Now, if you can look at this q i, the first thing you will observe is that q i is greater than or equal to 0. All the quantities involved are positive. Q i is greater than 0. Since each 2 to the power of minus l i greater than equal to 0 summation j equal to 0 to M minus 1 2 to the power of minus l j is greater than equal to 0. So, q i is less than equal to 0.

(Refer Slide Time: 05:43)



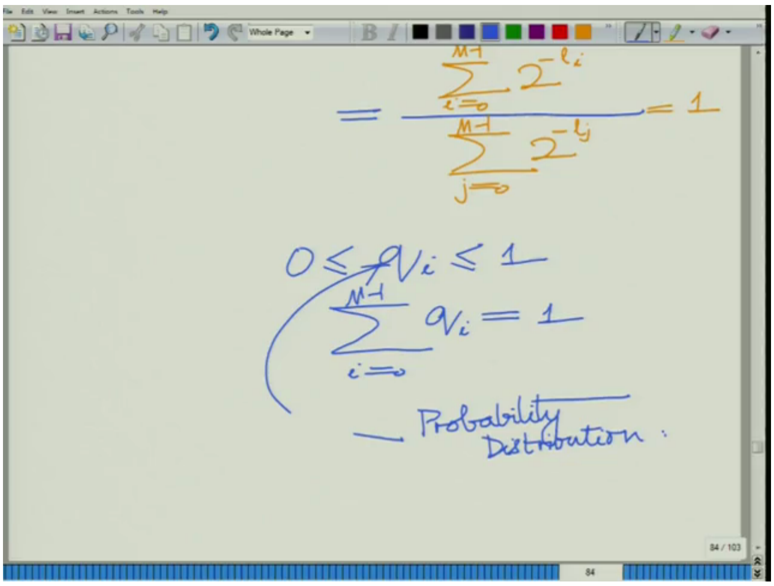
Further, we have 2 to the power of minus l i, each 2 to the power of minus l i. Remember all the quantities are positive. This is less than equal to summation j equal to 0 M minus 1 2 to the power of minus l j because 2 to the power of minus l i is one of the components in this summation, all right. So, all the quantities are positive and 2 to the power of minus l i is in fact one of the quantities in the components in the summation on the right hand side. Therefore, 2 to the power of minus l i is less than and equal to summation j equal to 0 to M minus 1 2 to the power of minus l j which implies 2 to the power of minus l i divided by summation j equal to 0 to M minus 1 2 to the power of minus l j. This is less than or equal to let me just write this a little bit more clearly. This implies 2 to the power of minus l i divided by summation j equal to 0 to M minus 1 2 to the power of minus l j less than or equal to 1 or rather this is our, in fact nothing, but our q j or q i. This is q i which is less than equal to 1.

(Refer Slide Time: 07:22)



So, we have q i 0 less than or equal to q i less than or equal to 1. Further summation i equal to 0 to M minus 1 q i equals summation i equal to 0 to M minus 1 2 to the power of minus l i divided by summation j equal to 0 to M minus 1 2 to the power of minus l j, the denominator is a constant.

(Refer Slide Time: 07:59)

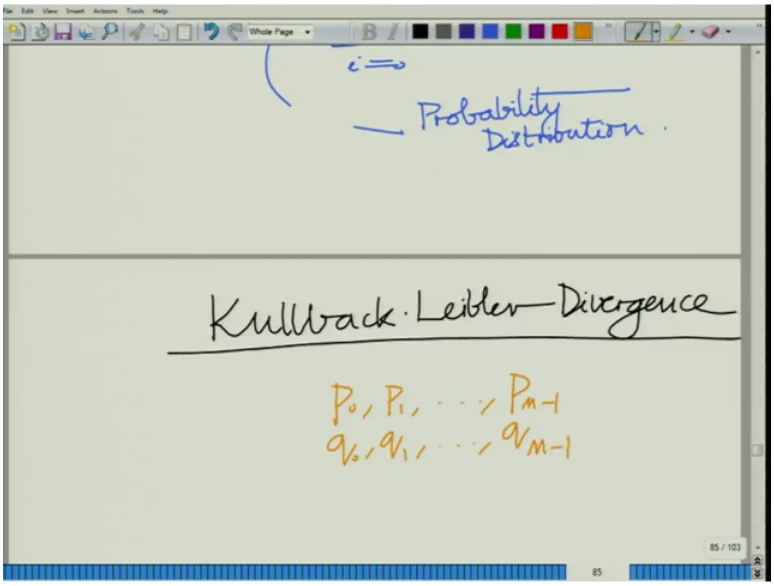


So, taking summation over the numerator, what we have is summation i equal to 0 to M minus 1 2 to the power of minus l i divided by summation j equal to 0 to M minus 1 2 to the power of minus l j. This is equal to 1.

So, what we have is each q i is positive, that is it is non-negative. 0 is less than equal to lies between 0 and 1 and summation of all q i is equal to 1. So, naturally the q i form a probability mass function on the probability distribution correct. So, q is correct. So, we have 0 less than or equal to each q i less than or equal to 1 and we also have summation i equal to 0 to M minus 1 q i equal to 1. So, q i is from a probability distribution.

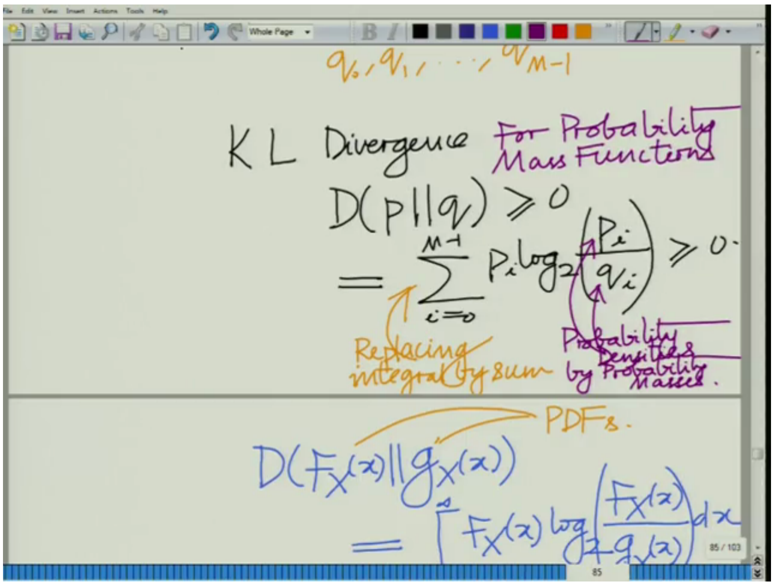
Now, let us use the concept of the Kullback Leibler Divergence which we have seen before.

(Refer Slide Time: 09:17)



So, now we have a probability distribution Pi P0, we have a probability distribution P0 P1 Pm minus 1 q0 q1 qm minus 1 and therefore, the KL divergence between these Kullback Leibler Divergence between these two probability.

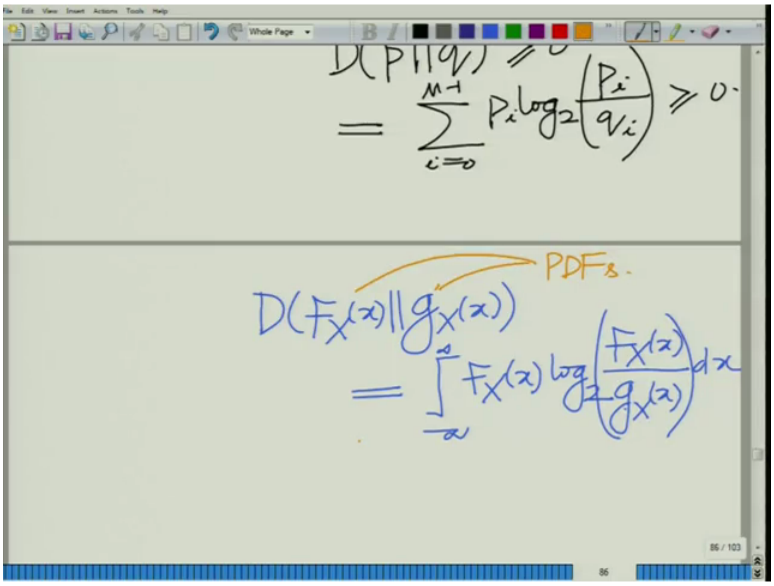
(Refer Slide Time: 10:07)



So, we have two probability distributions, the KL divergence which is defined as D which is equal to summation i equal to 0 to M minus 1 p i log to the base 2 p i divided by q i. This must be greater than or equal to 0. This is the KL divergence.

Remember we had looked at KL divergence between two probability density functions.

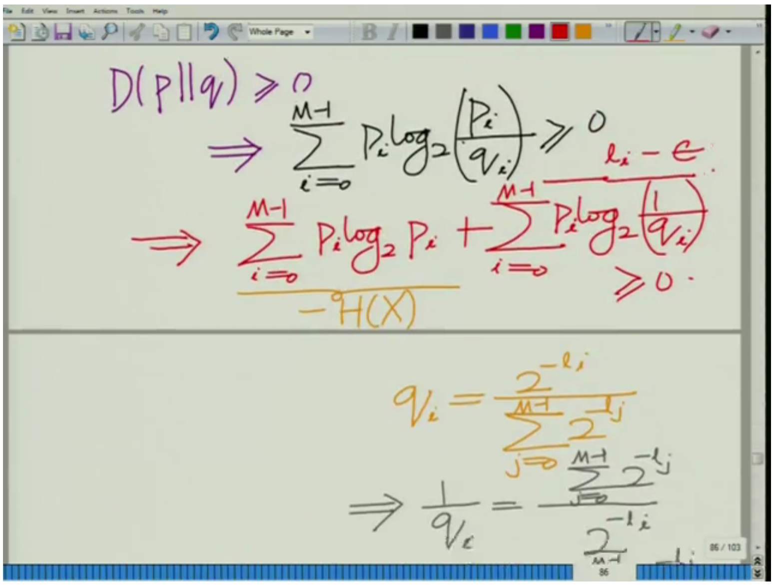
(Refer Slide Time: 10:47)



We had defined it something like this, the KL divergence between, for instance if F and g which are two probability density functions, F of x and g of x, we had defined it for continuous probability density functions as two probability density functions correspond into the random variable x as f of x g of x log 2 to the base 2 f of x. This was a definition, where f of x and g of f are probability density functions.

Now, what we are doing is, we are again doing the same thing for discrete probability mass functions. P and q are probability mass functions correct, probability distributions on discrete symbols s0 s1 s minus 1. So, what we have done is, we have taken the definition of KL divergence which we have defined for probability density functions and we have in fact now given the equivalent definition for probability mass functions which is obtained by of course replacing this integral by summation, that is replace continue integral which is the continuous sum replacing my integral by sum and of course, probability density functions by probability mass functions probability densities by probability masses. So, we have this is basically KL divergence for probability and you can say probability mass functions. So, that is what we have over here.

(Refer Slide Time: 13:12)

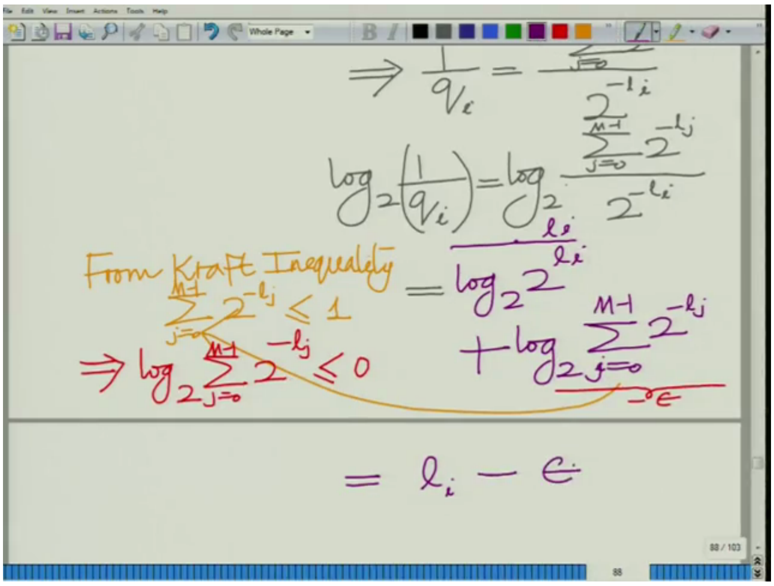


Now, therefore, we have KL divergence greater than or equal to 0. Remember we said KL divergence and in fact, we had established proved using log concavity, right using concavity of the log function that KL divergence is always greater than equal to 0 and the same property of course we are done it in the case context of probability density functions, but the same is also valid for probability mass functions, ok.

This implies that summation i equal to 0 to M minus 1 Pi log to the base 2 Pi log to the base 2 Pi greater than equal to 0 which implies summation i equal to 0 to M minus 1 Pi log to the base Pi minus plus summation i equal to 0 to M minus 1 Pi log to the base 2 1 over q i greater than equal to 0. I can equivalently write in this fashion.

Now, if you look at this quantity Pi log to the base 2 Pi, you will realize that this is nothing, but minus H of x and now q i remember equals to the power of minus li divided by summation j equal to 0 to M minus 1 2 to the power of minus lj which implies 1 over qi is simply summation j equal to 0 to M minus 1 2 to the power of minus l j divided by 2 to the power of minus l i.

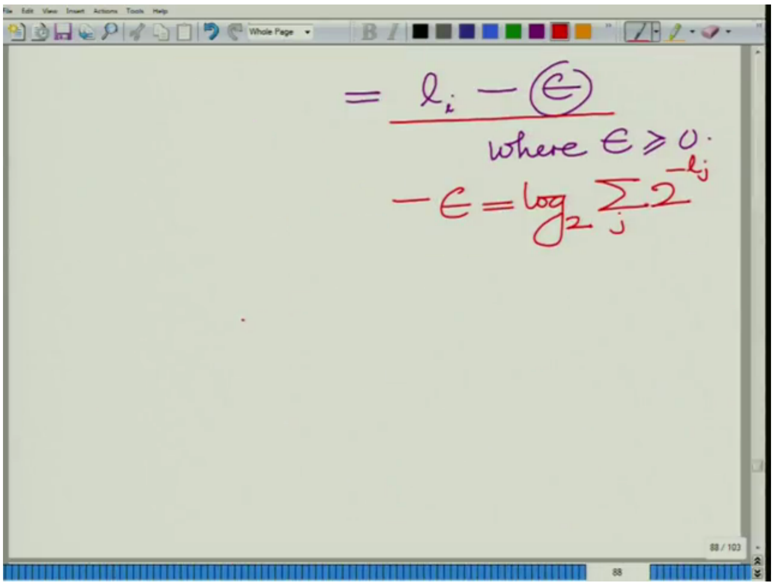
(Refer Slide Time: 15:06)



So, log 2 to the base log to the base 2 1 over q i, this is nothing, but log to the base 2. Well, summation j equal to 0 to M minus 1 j equal to 0 to M minus 1 2 to the power of minus l j divided by 2 to the power of minus l i 2 to the power of minus l i can come to the numerators or that becomes 2 to the power of l i. So, this is simply log to the base 2 2 to the power of l i. Let me just write that also log to the base 2 2 to the power of l i plus log to the base 2 summation j equal to 0 to M minus 1 2 to the power of minus l j.

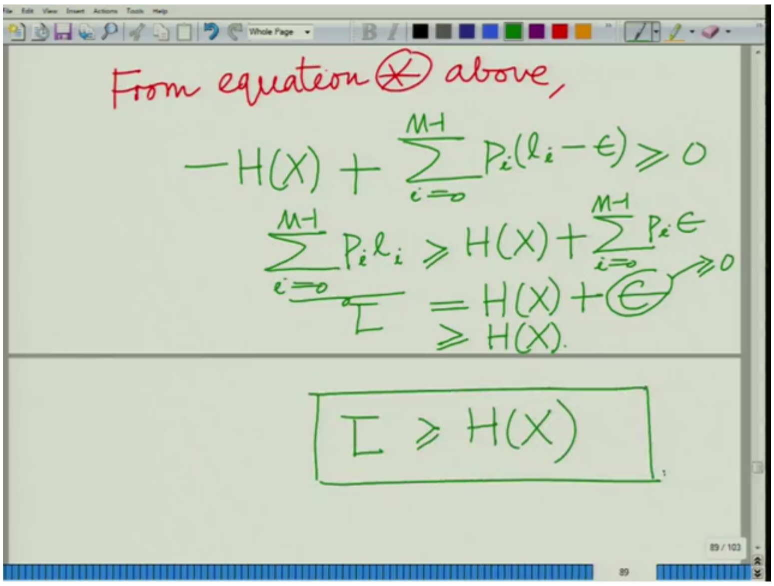
Now, if you look at this, this quantity is nothing, but l i. So, this will be l i plus now if you look at this quantity here, we know from Krafts inequality or from kraft inequality, we know that summation j equal to 0 for any prefix code 2 to the power of minus l j is less than and equal to 1 which means implies natural log to the base 2 summation j equal to 0 to M minus 1 2 to the power of minus l j. This has to be less than or equal to 0 correct from Kraft inequality. We know that summation j equal to 0 to M minus 1 2 to the power of minus l j is less than or equal to 1. Therefore, if we take the logarithm of that quantity, the log has to be negative.

(Refer Slide Time: 17:46)



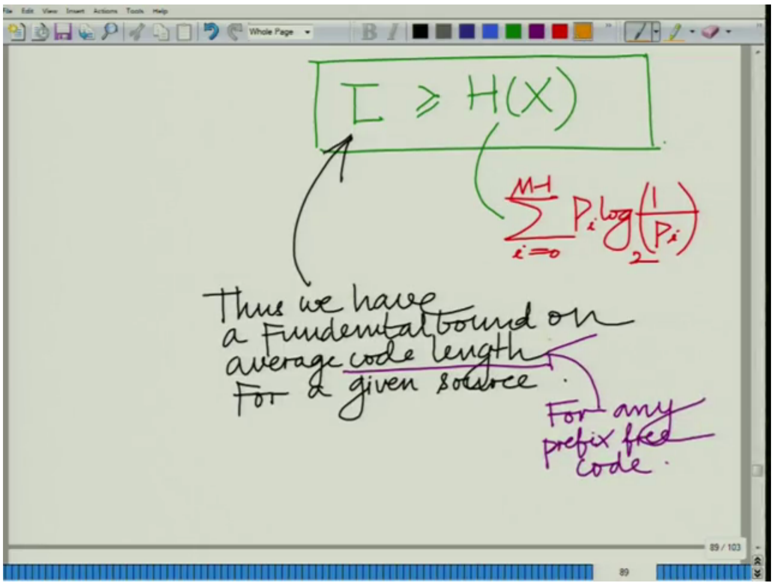
So, I can write this as minus epsilon, where epsilon is some positive quantity. So, that is the whole idea. So, I can write this as minus epsilon, where epsilon is some positive quantity and that is the reason for that is because, what is minus epsilon. Minus epsilon is log to the base 2 summation over j 2 to the power of minus l j and therefore, this is equal to l i minus epsilon. So, now, if you substitute these quantities here, this is equal to l i minus epsilon and now, what we have interestingly is if we call this equation as star.

(Refer Slide Time: 18:35)



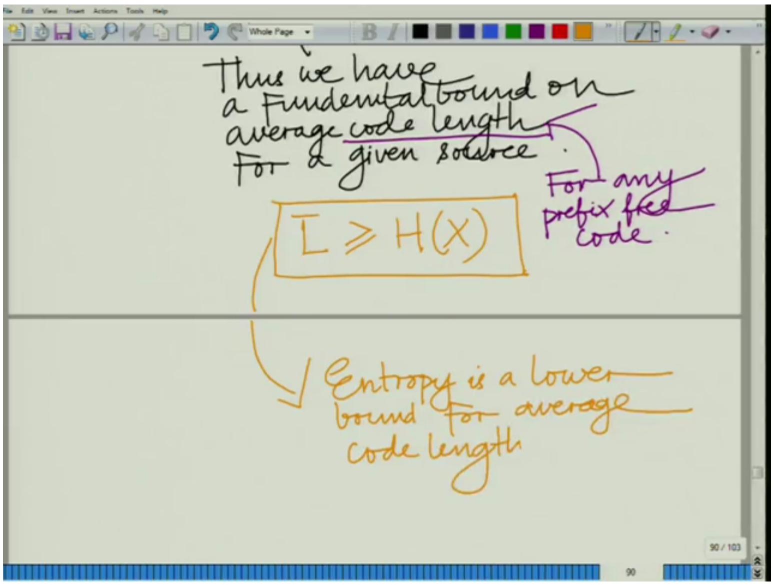
If we call this equation as star, from star what we have from equation star above, what we have is minus H X plus minus H X plus minus H X plus, well P i into l i minus epsilon minus H X plus summation i equal to 0 to M minus 1 P i into l i minus epsilon is greater than or equal to 0 from KL divergence implies summation i equal to 0 to M minus 1 P i l i greater than or equal to H X plus summation i equal to 0 to M minus 1 P i into epsilon. Epsilon is constant that comes out summation P i which is equal to 1. So, this is simply H X plus epsilon implies this is greater than or equal to and epsilon is positive. Recall epsilon is greater than equal to 0 which implies this is greater than equal to H X, ok

(Refer Slide Time: 20:29)



So, this is nothing, but l bar. So, what we have is l bar greater than or equal to H X which is the entropy of the source, that is summation P i log q over P i and there we have a very fundamental result, a fundamental bound on the average length of any prefix free code using the Kraft inequality. We have shown that the average length of any prefix code has to be lower bounded by the entropy of the source and this is a very fundamental elegant and interesting result. So, what this shows us is that, what this is telling us is that the average length of any prefix code, thus we have a fundamental bound on average code length for a given source and remember not any code, average code length, we have to qualify this. This is for any prefix free code, average code length for any prefix free code and we have shown that this is bounded by the entropy.

(Refer Slide Time: 22:17)



So, the average code length l bar greater than equal to H X, that is average code length that is entropy is a lower bound. What this tells us that entropy is a lower bound for the average code length. So, this is very fundamental. What this says is no matter what prefix free code you design, all right this has to satisfy the kraft inequality. What that tells us is that the average code length cannot be lower than the entropy. At most, it can be equal to entropy. It has to be greater than and equal to entropy.

So, the efficiency of a code can now be judged by how close is the average code length of the entropy. So, we have a convenient means to judge the efficiency of a code. Remember we said that lower the average length of the code, the more efficient it is and now, we have shown that you cannot arbitrarily reduce it to any non-zero quantity. This is lower bounded by the entropy. So, you can approach entropy. I mean one can desire or one can design a code to approach the entropy as closely as possible, but cannot of course make it lower than the entropy and therefore, the closeness of this average code length to the entropy you can characterize, can be used as a measure to characterize the efficiency of the designed prefix free code for a given source, all right.

So, now how closely can you approach this entropy and how to approach, it is something that we are going to see in the subsequent modules.

Thank you very much.