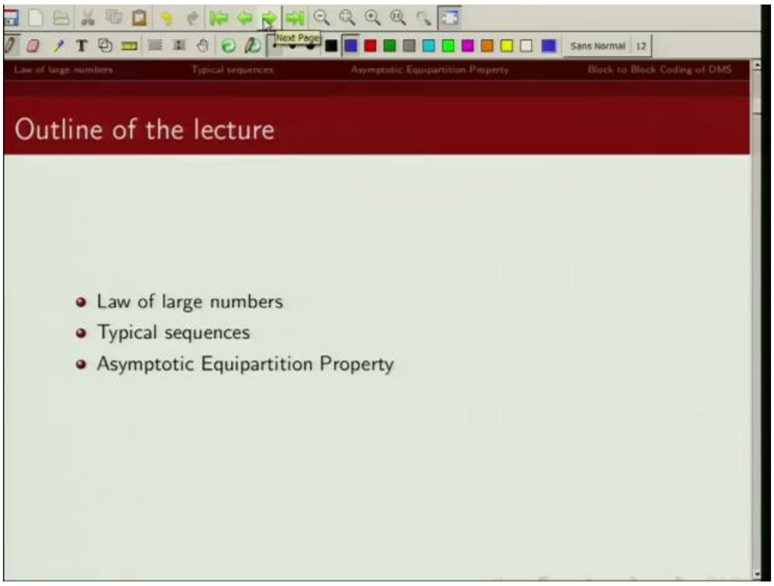
An Introduction to Information Theory
Prof. Adrish Banerjee
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kanpur

Lecture – 06A
The Asymptotic Equipartition Property

Welcome to the course on an introduction to information theory. I am Adrish Banerjee. So, in this lecture we are going to talk about block to block length coding. So, far we have talked about block to variable length coding, variable to block length coding and both of these were examples of loss less forced compression. Now to do block to block length coding, we can only get compression, because the input block size is also fixed, output block size is also fixed. So, we can only get compression if it is a basically lossy source compression, because our output block size has to be smaller than the input block size, which is meant both of them are fixed. Now in case of lossy source compression of course, we would like to minimize our probability of error. So, the way we will do this block to block length coding is, we are going to assign code words to those sequences that are more likely to happen.

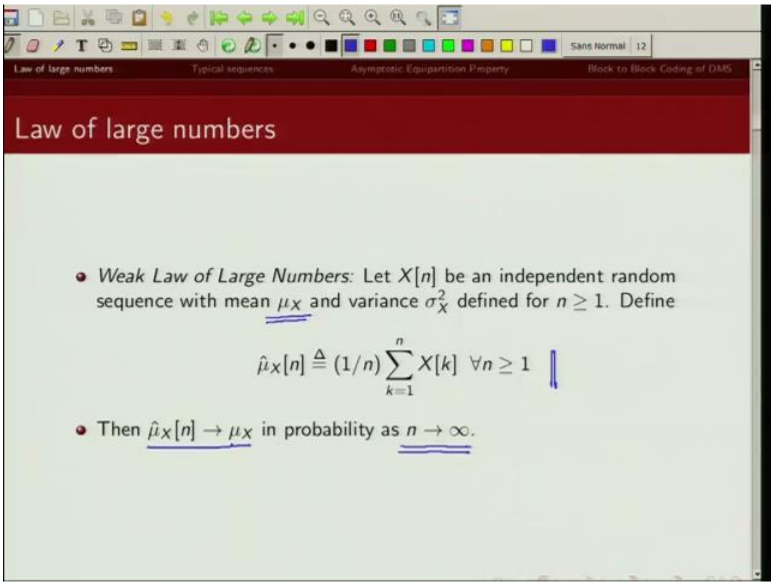
We will assign a unique code word to all those sequences, and the sequences which are less likely to happen, for all those sequences we are going to assign only one code word. So, if we get a typical sequence which is a more likely sequence coming out of that particular source. Since we are assigning a unique code word to those sequences, or the decoder we would be able to find out what our message bit's were, and we will encounter a lossiness or we will encounter distortion, when a non typical sequence, or sequence which is less likely to happen that is transmitted. So, to talk about block to block length coding, we will first talk about what is the typical sequence expected out of a source. If I sa let us say a binary source with probability of u a 0 being 0.5, and probability of 1 being 0.5, what are the sequences which are likely to happen from that particular source, if I take a very large block size. We are going to talk about that that is the concept of typical sequence, and then we will show some properties of these typical sequences, which are collectively known as asymptotic equipartition property.

(Refer Slide Time: 02:46)



We will prove these properties. So, this is the plan for today's lecture. Now before we go into what is a typical sequence, we need to build up some mathematical preliminaries to prove our results. So, we will talk about those mathematical preliminary, we will talk about re-class large number and ((refer:3:10)) inequality, and then we will use those results to prove some properties of typical sequence, and these are as I said collectively. Known these three properties of typical sequence are collectively known as asymptotic equipartition property, and then we will talk about what is the consequence of these properties for block to block length coding

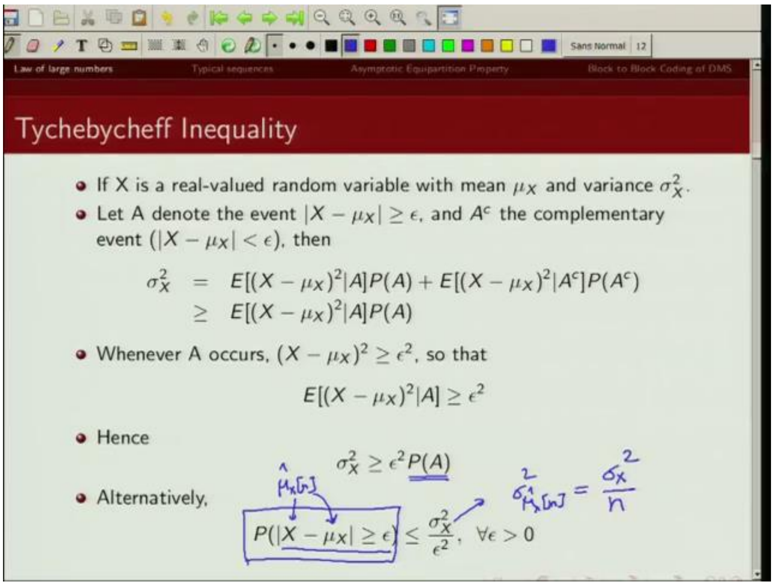
(Refer Slide Time: 03:35)



.

So, we start with weak law of large numbers. So, let x of n be independent random variables sequence with mean mu of x and variance given by sigma x square. So, let x of n be a independent sequence of random sequence with mean mu of x and variance sigma x square. Then the sample mean basically is given by this expression; summation of x k divided by 1 by n, this summation divided by 1 by n. Now, weak law of large number says, this converges to this mean mu of x in probability as n tends to infinity. So, let us first show what do you mean by mu of x converges to this true mean mu x, when n is very large. What do you mean by this converges to your mu of x in probability.

(Refer Slide Time: 04:49)



.

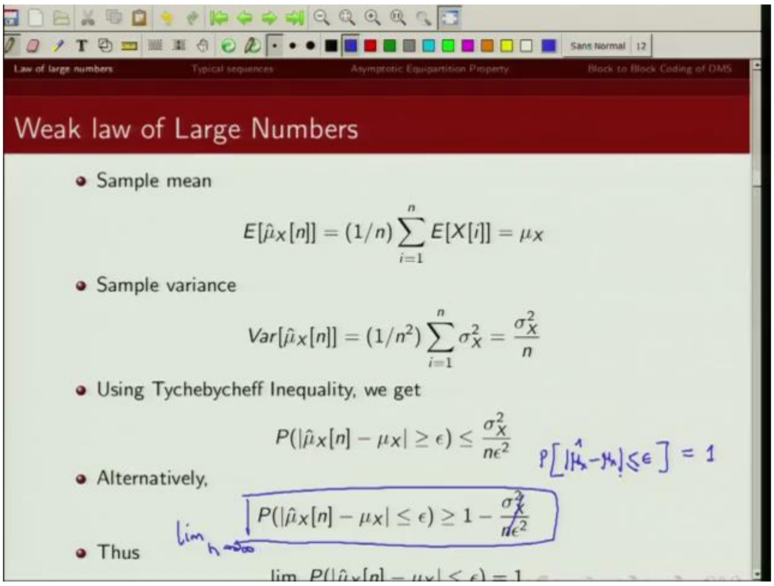
So, to prove this let us talk about, let say x is the real value valued random variable, with mean given by mu of x, and variance given by sigma x square. Now we denote an event a as follows. So, event a i's, this event that absolute value of x minus mu of this random variable x, is greater than equal to some small value epsilon. So, we can similarly define the complimentary event, which would be the event that absolute value of x minus mean of x is less than epsilon. So, we know the variance sigma square, is given by expected value of x minus mu x whole square. So, this we can write as expected value of x minus mu x whole square given event a as occurred multiplied by probability of event a plus expected value of x minus mu x square given the complimentary event of a has happened and multiplied by probability of the complimentary event.

Now, look at these quantities separately, what is probability of this complimentary event. There is a probability, so this will be a number between zero. And one and this is expected value of x minus mu x square given the complimentary event of a. So, this is the expected value of x minus mu x whole square. So, this expected value will also expected to be greater than equal to zero. So, then this whole term, this is greater than equal to zero, this is between zero and one. So, this whole term will be greater than equal to zero. So, then we can write that sigma square is greater than equal to this term, because the second term here is non negative. Now let us look at what happens when event a occurs. So, when event a occurs. What is event a? Event a is this condition the absolute value of x minus mu x is greater than epsilon.

So, when event a occurs x minus mu x whole square, should be greater than equal to epsilon square. So, we can write then expected value of x minus mu x square given event a has happened is greater than equal to epsilon square. Hence if I plug this value of, expected value of x minus mu x whole square given a. If I plug this value in here, what I get is this expression; mu x square variance of x is greater than equal to epsilon square into probability of occurrence of this event a, which is this event; that absolute the difference of x minus mu x the absolute value of that, is greater than equal to epsilon. So, this event a as I said is basically the event, that absolute value of x minus mu of x is greater than equal to epsilon. So, from here I can write probability of a, which is this, is less than equal to sigma x square by epsilon square, and this simulation holds for all positive epsilon.

(Refer Slide Time: 09:13)

.

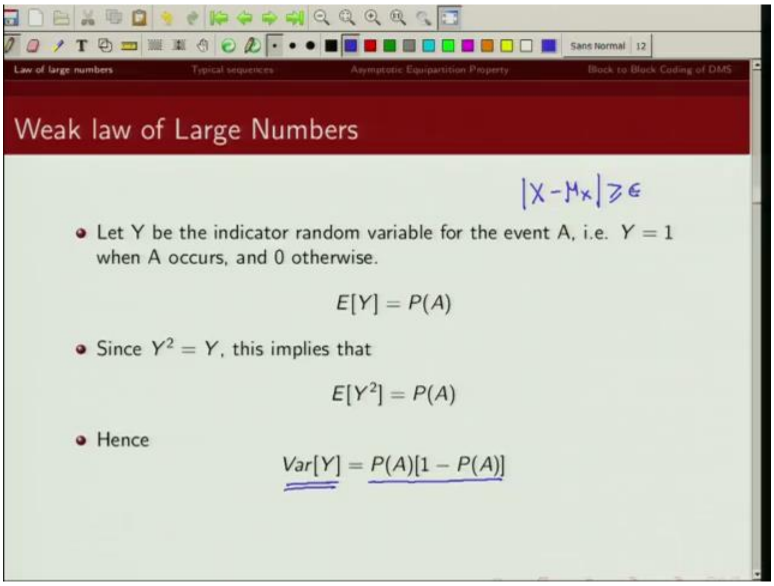


Now, let us define this expected value of mu x mu hat x this is nothing, but we can show that this is equal to mu of x which is the mean of x. similarly the variance of mu hat, this is equal to sigma x square by n. So, the mean of mu hat x is given by this, and the variance is given by this expression. now if we invoke this thing probability of x minus mu of x, the absolute difference, absolute value of that is greater than equal to epsilon, this probability is upper bounded by sigma x square by epsilon square. So, in place of x you put mu hat of x, and the mean of mu hat of x is from this expression is given by mu x. So, that is my mean of mu hat of x. And what is the variance of mu hat of x; that is this quantity. Remember I had this probability of x minus mu x. This is greater than equal to epsilon, I had this as upper bounded by variance by n. So, here this is my mu hat of x that is this, and it is mean is given by mu x that follows from here. And what is the variance of mu hat of x. Variance of mu hat of x is given by sigma square by sigma square x by n.

So, when I put this variance value here I get sigma square by n by, sorry this was epsilon square. Again just go back and look at this expression. So, I am replacing this by mu x of n, and replacing this by variance of mu x of n, and mean of this is given by this expression, and this variance of mu hat of x we just proved, it is sigma square by n. So, if you plug these values and this expression, what we get is, this. probability that mu hat of x minus mu x, the absolute difference of that being greater than equal to epsilon, is upper bounded by sigma x square by n epsilon square. Now the same thing I can write in this particular form. So, the probability that absolute difference between mu hat of x minus mu x is less than equal to epsilon is, greater than equal to 1 minus sigma x square by n times epsilon square.

And here if I let n tends to infinity, then what I would get is probability that mu, the absolute difference between mu hat of x minus mu x, is less than some small value epsilon, is basically greater, is almost equal to one. So, that essentially shows that, this quantity when we take limit n tends to infinity, probability that absolute difference between mu hat of x minus mu x, this being less than equal to epsilon, this will basically become one, because this term will go towards zero, this term will go towards zero. So, this probability will be almost equal to one. So, this mu. So, mu hat of x approaches mu x in probability, converges in probability; that is what we said earlier that, if you just go back, we said mu. If we define mu hat of x like this, this converges to mu x in probability as n tends to infinity.

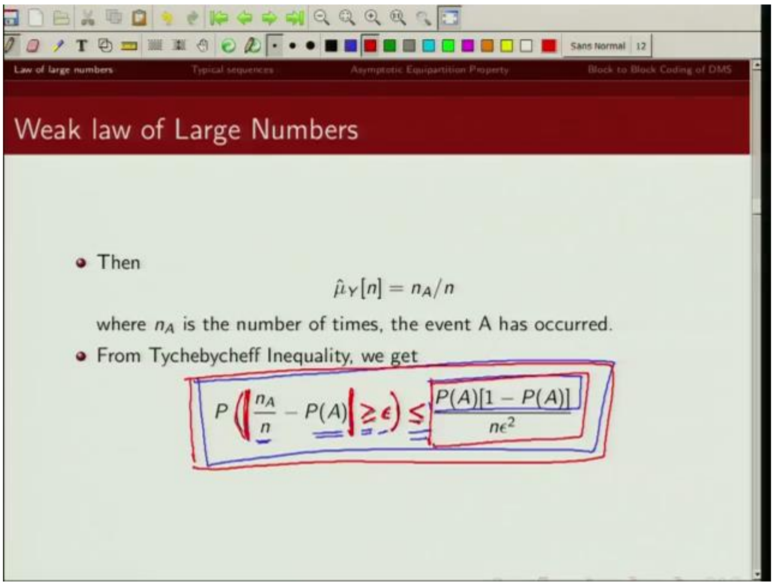
(Refer Slide Time: 14:20)



.

Now, let us define a random variable y. So, y is a indicator random variable for this event a, and what is this event a. we call this was the event that x minus mu x is greater than equal to epsilon. So, what does that indicator random variable do. So, it will indicate one, when event a has occurred; otherwise, it will be zero. So, we can say that expected value of y is, given by probability of occurrences of this event, because y takes two values; zero and one right. It takes value one. One happens when the event a happens, and that is happens with probability p of a, and event a does not happen that is probability of this complimentary event which is nothing, but 1 minus p of a. So, expected value of y is zero times this probability, plus one times probability of a. So, that is why expected value of y is nothing, but probability of a. similarly we can define another random variable we can see that y square will be y. So, when y is 0 y square will also 0, when y is 1 y square is also 1. So, then since y square is equal to y, we can write expected value of y square is expected value of y, which is given by this expression. Now what is the variance of y, this can be written as expected value of y square minus expected value of y whole square. So, that will be given by probability of a minus probability of a whole square

(Refer Slide Time: 16:49)



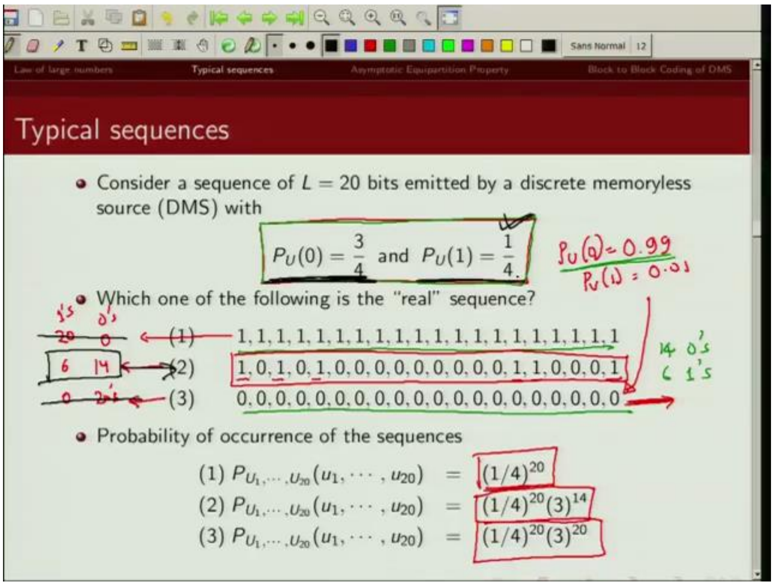
.

. So, that is nothing, but probability of a into 1 minus probability of a, where a is that event x minus mu of x is greater than equal to epsilon. Next we define mu of y as n a by n, where n a is the number of times that event a has occurred, out of n times v e. So, n a times event a has occurred at n minus n a time event a has not occurred. So, then from the Tychebycheff inequality we can write probability of n a by n minus mean of that, which is basically probability of a, this absolute difference is greater than equal to epsilon is upper bounded by variance of, variance by epsilon square, and variance we have computed, this is basically sigma y square by n, sigma y square is given by this quantity. So, this whole thing is then upper bounded by probability of a multiplied by 1 minus probability of a.

So, here we have used this Tychebycheff inequality form which we have derived earlier.

We used this, this form right, instead of mu of mu hat of x we had mu hat of y and here variance of y. So, we have used this form to get our result, which is this one. Now this result we are going to use later on to prove some properties of typical sequence, and we will define what is a typical sequence. you just remember this expression that probability that the difference between n a minus n n a by n minus probability of this event a, the absolute difference of that, being greater than equal to epsilon, probability of that is upper bounded by this quantity. So, now, we have developed the expressions that we need to prove some of our results. So, now let us talk about what is a typical sequence, and then we will prove some properties of typical sequence, and subsequently we will see what is the consequence of typical sequence for block to block length coding.

(Refer Slide Time: 19:45)

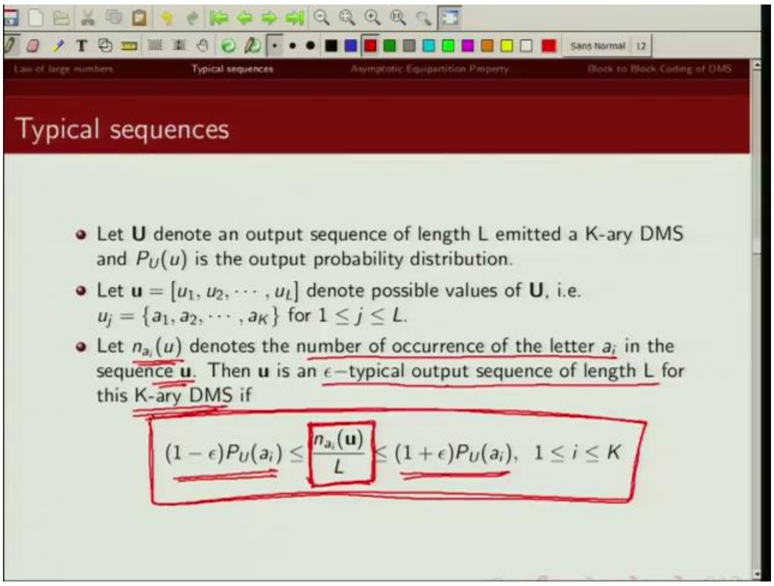


So, as an example let us consider a sequence of l bit's, and this speaking l to be small number twenty, because I just want to just illustrate and this particular with an simple example. So, let us consider discrete memoryless source, which is emitting binary sequence zeros and ones with probability given by this. So, probability of 0 is 3 by 4, and probability of 1 is 1 by 4. Now my question to you is, which of these three sequences are more likely to have come out from this particular source, this discrete memoryless source which is, we can make zero with probability 3 by 4 and one with probability 1 by 4. So, what is this sequence; the first sequence that we see has all ones, the second sequence that we see has zeros and ones, and third sequence that we see is all zeros.

Now, if we look at probability of occurrence of each of these sequences, because one happens with probability 1 by 4. So, this will happen with probability 1 by 4 raise power 20. In the second example we have 1 2 3 4 5 6 ones right and 14 zeros. So, this will be 3 by 4 raise to the power 14 into 1 by 4 raise to the power 6. So, that is this quantity. And similarly the all zero sequence, this would be the probability of occurrence. Now the point we noted is, even though this has have probability concedes 1 by 4 raise to the power 20 into 3 by 20, even though this is higher property of these three, sequences that we have shown here. This is unlikely to come out of this particular source. Why is this unlikely to come out of this particular source? Because I have been given that probability of occurrence of 0 is 3 by 4, and probability of occurrence of 1 is 1 by 4.

If I take large enough block size it is likely that I should get zero three fourth of time, and I should get 1 one fourth of time. So, and all zero sequence or all one sequences is very unlikely to come out of this particular source. If I would have lectures told you this is a probability of 0 is 0.99, and probability of 1 is 0.01. Then you would have believed if I have said this particular sequence would have come out from this particular source, because in this particular source probability of occurrence of zero is very high right, but for the source which I have given here, which is this probability of 0 being 3 by 4, and probability of one being 1 by 4. It is unlikely that this particular sequence would have come out from this particular source. Similarly all ones sequence is very unlikely to come out of this source. Whereas here if you see, I just mentioned it has 14 zeros and 6 ones, and after if you take large enough block size I expect three fourth of those bit's to be zeros, and one fourth of them to be one, which is close to what I am getting in sequence number two. So, this is a more likely sequence to come out of this particular discrete memoryless source. So, what I say is, this is a typical sequence which is expected out of this particular source. Now we will formally define what is a typical sequence, and then we will show some properties of typical sequence.

(Refer Slide Time: 24:06)

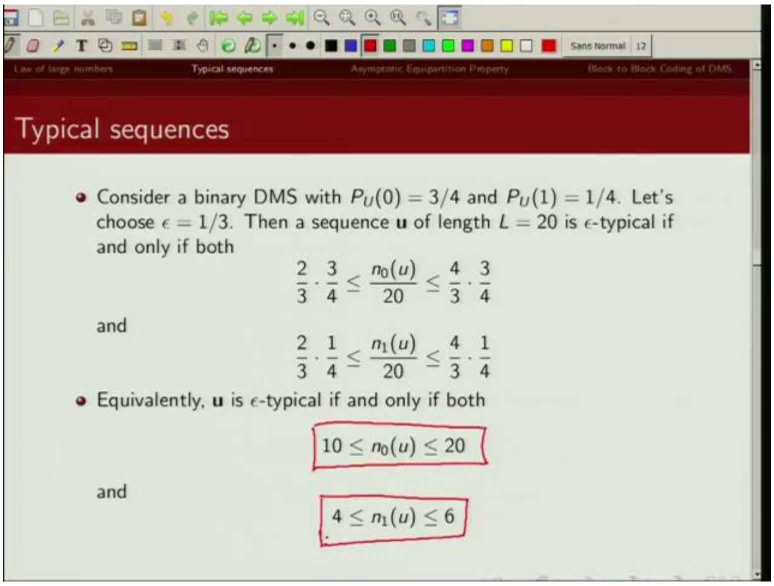


.

So, let u be the output sequence of length l, which is emitted by a k ary discrete memeoryless source, and let p u be the output probability distribution of that. So, since u is of length l, we are denoting u by u 1 u 2 u 3 u of l. So, that is my output sequence of length l. now each of these u is, since the source is a k ary source. So, this u i could be a 1 a 2 a 3 or a k. now let n a i u denotes the number of occurrence of this letter a i in the sequence u, then u is an epsilon typical output sequence of length l for a k ary discrete memoryless source if this condition is satisfied. So, n a i u by l, this is a fraction of the output. So, I have a output of length l ,out of them l sequence n a i are the number of basically bit's you can say which have a is. So, n a i denotes the number of occurrence of this letter a I, out of those l input u of length l. So, this is the fraction of a i in my output sequence. now what I am saying is, this fraction of a is in my output sequence of length l, this should be within 1 plus epsilon times p of u a i.

It should be within 1 minus epsilon p u a i and 1 plus epsilon p u a i, where epsilon is a very small number you can say whether epsilon is zero essentially what I am saying is this fraction of a is in my sequence of length l, should be equal to probability of occurrence of a i. So, if you take, if l is large enough then in a typical sequence what you would expect is, the number of occurrence of this letter a i. The fraction of the time a i is appearing in this sequence of length l, that should be close to probability of occurrence of a i. And this should hold for all the k ary alphabet. So, it should hold for a 1 a 2 a 3 a k. So, we say u is a the typical epsilon typical sequence of length l, if this condition is satisfied. Please note epsilon is a small number.

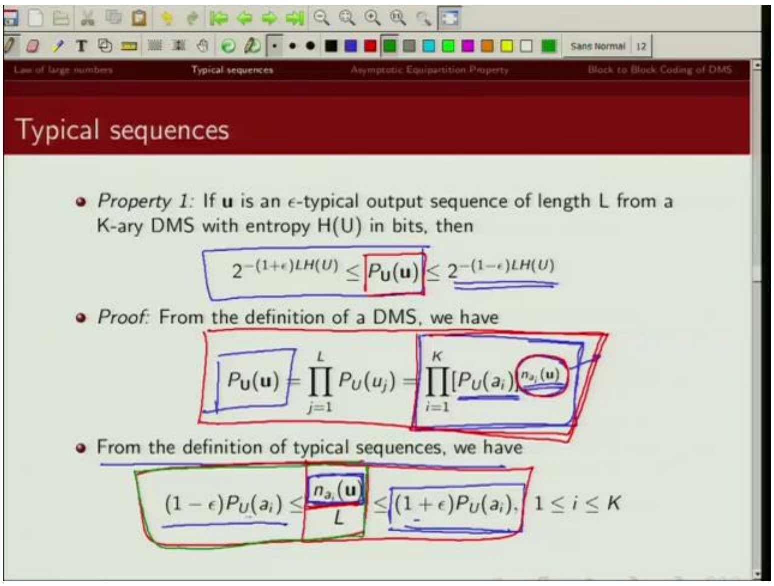
(Refer Slide Time: 27:39)



So, let us say illustrate with a help of an example. So, we consider the same example that we have considered earlier. So, we consider probability of u to be 3 by 4 by u being 0 to be 3 by 4, probability of u being 1 is 1 by 4. Let us take epsilon to be 1 by 3, then according to this number of occurrence of zero should satisfy this relationship, and what is that. So, number of occurrence of zero which I am denoting by n 0 u divided by length l, which is 20 here, this should be less than equal to 1 plus epsilon. 1 plus epsilon is 1 plus 3, 1 by 3 that is 4 by 3 times probability of occurrence of zero. What is the probability of occurrence of zero. That is given by 3 by 4. So, this is this, and this is u r bounded by 1 minus epsilon. So, 1 minus 1 by 3 will be 2 by 3 into probability of occurrence of zero which is 3 by 4. Similarly we can write the relationship for number of occurrence of one, which I am denoting by n 1 u. So, n 1 u by 20 should be less than equal to 1 plus epsilon which is 4 by 3 in this case into probability of occurrence of u equal to 1 which is 1 by 4. And similarly the lower bound will be 1 minus epsilon which is 2 by 3 into 1 by 4. So, these are the two, because we had binary source, so number of occurrence of zero should satisfy this relationship, number of occurrence of one should satisfy this relationship.

Now, simplifying this we get this condition that number of zeros in the sequence of length 20 should be between 10 and 20, and number of ones should be between 4 and 6. I will go back and see the example that we did earlier. Here the number of ones are 20 ones, and number of zeros are zero. Here number of ones are 6, and number of zeros are 14. And in this particular example number of ones are zero, and number of zeros are twenties. Now what did we said we said number of zero should be between 10 to 20. So, this cannot be a valid sequence coming out of this particular source, because number of zeros are not between 10 to 20, and what condition which we say about number of ones. You said number of ones should be between 4 and 6, so this is also ruled out. So, this is the likely sequence which have 6 ones and 14 zeros. This is a likely sequence to have come out of this particular source, which has probability of occurrence of zero, given by 3 by 4, probability of occurrence of one given by 1 by 4.

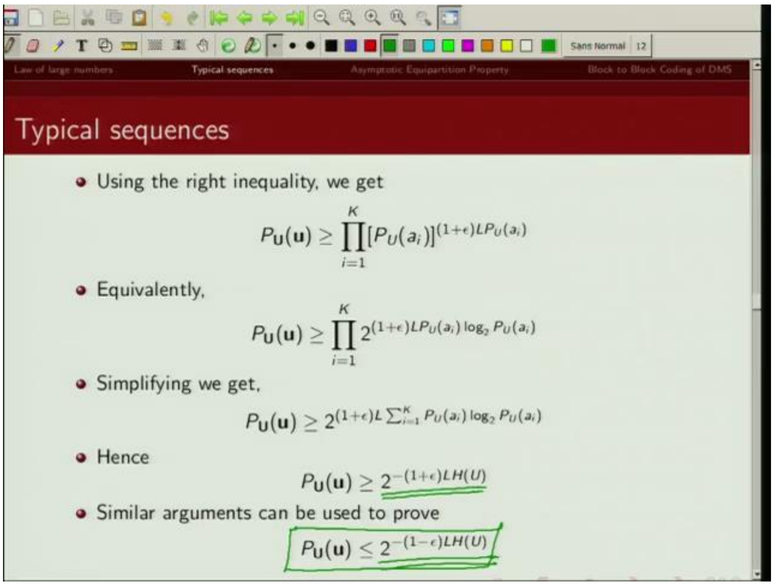
(Refer Slide Time: 31:19)



.

Now, let us prove three properties of these typical sequence. Now that we have formally defined what a typical sequence is, and we are also illustrated with the help of one example what a typical sequence is. Now let us try to prove some properties of typical sequence. So, we are going to prove three properties of typical sequence. The first property that we are going to prove is, as follows. So, if u is an epsilon output typical sequence of length l, which comes out from a k ary discrete memoryless source, which has entropy given by h of u, then probability of occurrence of u, is lower bounded by this quantity and upper bounded by this quantity. So, let us see how we can prove this. So, from the definition of the discrete memoryless channel, probability of occurrence of u can be given as probability of occurrence of u 1 into probability of occurrence of u 2 into probability of occurrence of u l. So, that is this quantity. Next we can also write it in terms of the alphabets emitted by the discrete memoryless source. So, since this is a k ary source, it will emit a 1 a 2 a k. So, let us say a 1 is emitted n a i times, a 2 is n a 2 u times, similarly a k is n a k u times. So, then I can write the same probability in this particular fashion. So, it is a product from I going from one to k probability of a i raise to power n a I, where n a i is the number of times this a i is appearing in this output sequence of length l. Now, from the definition of typical sequence, we know that n a i should satisfy this lower bound and this upper bound. So, n a i will be less than equal to l times, this quantity and it should be greater than l times this quantity.

(Refer Slide Time: 33:50)



Now, let us take the right inequality. So, we have shown that probability of u, this is equal to this quantity. Now pay close attention to terms here, this is probability of a i, what is it. It is the number between zero and one. Now if you try to raise it toward power which is greater than n i. So, you have a number between zero and one, if you raise it to a power which is greater than n a i then this term will become less. For example, half x power 3 is more than the half x power 10. So, if I replace this by a larger number this probability, this term will become less than equal to probability of u. So, that is what I am going to do first. I am going to replace this n a i u by l times this quantity. So, I am going to use this upper bound first.

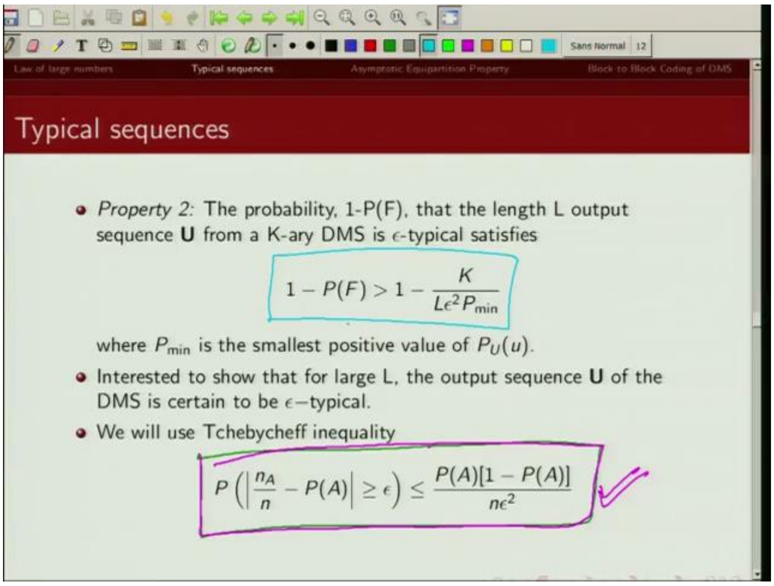
Now, if I do that. So, I replace n a i by l times 1 plus epsilon p of a I, and since this term was greater than n a i u. So, this whole term. So, I am raising a number between zero and one to a larger power. So, the equality which was there will be replaced by greater than equal to. next I will do some simplification I can write this thing, the same thing in this particular fashion, that 2 raise power 1 plus epsilon l probability of a i log of probability of a i. So, the same thing I can write it in this particular form. This is not very difficult to show, you can just take this equal to x take a log of it and then you can write it.

So, this same quantity can be written in this particular form. So, let us look at the terms here. So, I have terms of the form like this; two times let us say a 1 into two times a 2 into two times a 3 something I have a terms of the forms this now this can be written as two, and then summation of these a is. So, that is what I am doing in the next step. This quantity 1 plus epsilon does not depend on I, so I can take it out, and the remaining terms that I have is this one of course, l also does not depend on choice of i. So, what I am left with is something like this. now this minus of summation from 1 to k of p a i log p a I; that is basically given by the entropy function. So, entropy function we know is minus summation of p a i log p a i.

So, then this term becomes 1 plus epsilon l minus of entropy function. So, what I have shown you is, this probability of u is lower bounded by 2 raise to the power minus 1 plus epsilon l of l times h of u. Now following the same procedure we can prove that probability of u is less than equal to this. So, we will proceed exactly same fashion. We are going to first write the expression of probability of u, which will be given by this expression. Now instead of taking the upper bound and n n i u we are this time going to take a lower, I am going to take that a lower bound on n a i. So, if I am replacing this n a i by a smaller term, then this whole thing would be greater than equal to p of u i. So, I would get a upper bound on probability of occurrence of u.

So, exactly same procedure. Now instead of using this, I am going to use now this. So, again I will follow the same procedure I will replace n a i u by l times 1 minus epsilon p u a I, and since it is smaller than n a i. So, this whole thing would be greater than equal to p of u I, and subsequently I will write it in the form of 2 raise power 1 minus epsilon twelve times p u a i log p u a I, and following exactly the same procedure we will get the upper bound on p of u. So, I just verbally spoke out the proof, the proof is exactly similar to the way we prove the lower bound on probability of u. So, following those procedure we can easily show that, probability of u is lower bounded by this and upper bounded by this quantity. So, that is the first property of the typical sequence it tells us what is your probability of occurrence of one epsilon typical sequence.

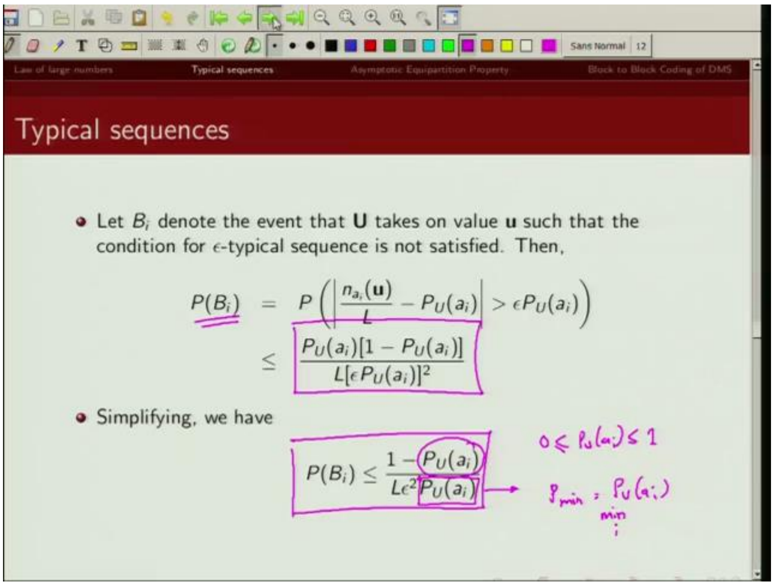
(Refer Slide Time: 40:24)



.

The next property that we are going to show for a typical sequence is as follows. the probability 1 minus p f that a length l output sequence coming out from a k ary discrete memoryless source is an epsilon typical sequence, is lower bounded by this quantity; 1 minus k divide by l epsilon square p minimum, where p minimum is the smallest positive value of p of u. So, let us prove this. So, again the consequence of this property would be, when n is very large this quantity will essentially go to zero then what we are essentially trying to say is, when l is very large probability of occurrence of a non typical sequence is very small. So, we will use this form of Tychebycheff inequality that we have proved earlier in this lecture, which is probability of the absolute difference between n a by n by minus p of a. it is absolute difference is greater than equal to epsilon, is upper bounded by probability of a into 1 minus probability of a divided by n epsilon square

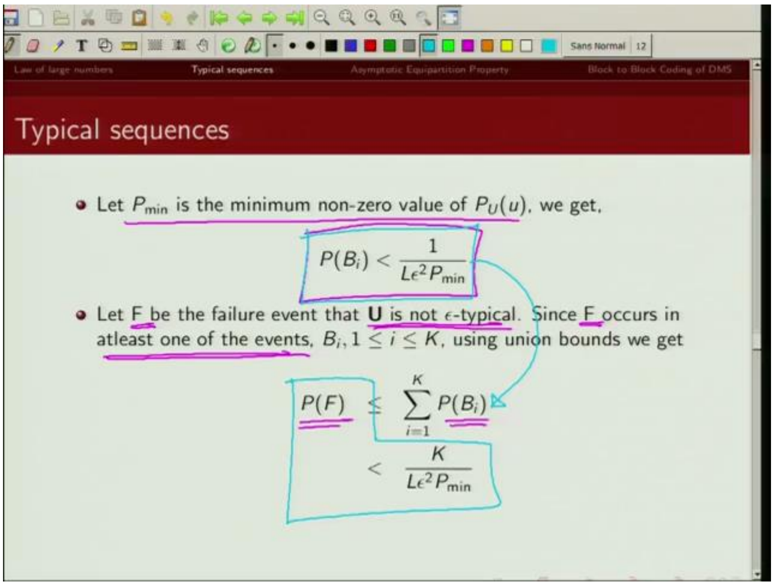
(Refer Slide Time: 41:53)



So, let b I denotes the event that u takes value such that the condition for epsilon typical sequence is not satisfied. Now go back to condition for epsilon typical sequence. This was a condition for epsilon typical sequence, this was a condition. We want number of occurrence of this a i divided by l, we want this fraction to be within 1 minus epsilon and into p of u i and 1 plus epsilon p of u a i. So, clearly when the absolute difference between n a i u by l minus epsilon p of u is greater than some term, then it is not going to be a epsilon typical sequence. So, then when n a minus n n a i u by l minus p of a i when the absolute difference is greater than epsilon times p of a I, then this condition happens. We have seen from the definition of typical sequence when this happens for any of these a is, then our sequence is not a typical sequence. So, what is your probability of occurrence of this. So, we can invoke tychebycheff inequality, and the result that we have shown here, we can invoke this result to get an upper bound on this probability. So, by doing this what we get is this expression. So, the probability that u takes a value which is not typical sequence is, this is. Now remember b I is the event that this is not satisfied for a i.

Similarly b 2 will be for condition when a 2 does not satisfy the condition for typical sequence, like that for all the case k ary inputs k ary which are I mean, k ary different symbols emitted by a discrete memoryless source, this condition if this condition holds for any one of these a is then it is not a typical sequence, because from the definition of typical sequence we said that this condition that we have mentioned, this should hold for all a is, this condition should hold for all a is. If any of the a is does not satisfy this condition, then u is not going to be an epsilon typical sequence. So, now, this probability when we plug in into this particular expression that we have derived, the probability that we get is as follows. So, probability of this event b I is given by probability of a i into 1 minus probability of a i divide by l into epsilon square n p of a i square. Now after simplifying we can write this p u a i will cancel one of this p u a is. So, probability of this b i is upper bounded by this quantity. Now this is the quantity which lie between zero and one. So, I can further simplify my expression as in this follows. So, this is a quantity between zero and one, this will be, this one minus this will be less than equal to one. And if I replace this p u a i by the probability of the p min, which is basically probability of u even a i minimum over all is

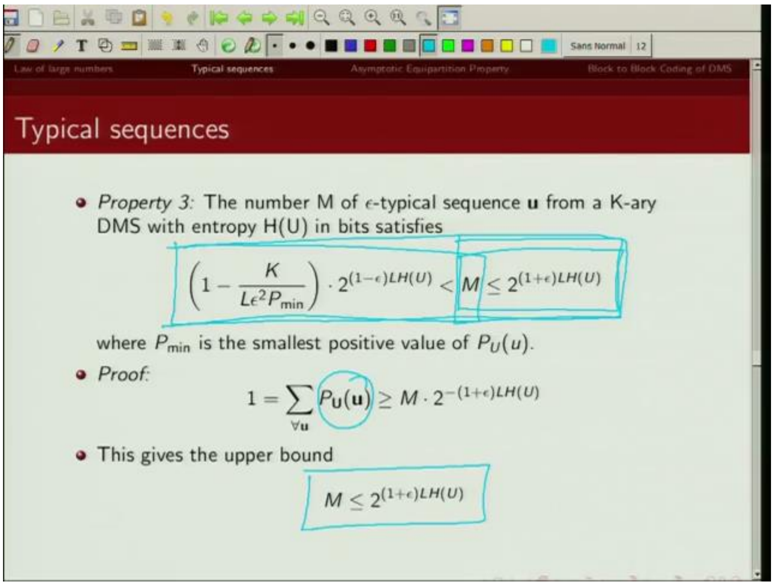
(Refer Slide Time: 46:38)



.

. So, if I replace it by that I get an upper bound, which is given by this expression. So, p min is the minimum non zero value of p of u i. Now we define f to be the failure event that u is not a typical sequence, and when will u not be a typical sequence, then any of this event b is occur. If a 1 does not satisfy the condition were a typical sequence and other a is satisfy, is still a not a typical sequence. So, if any of these a is do not satisfy the condition for typical sequence, the overall sequence u is not a typical sequence. So, f occurs with at least one of these events b is occur. Now using union bound I can write this probability of failure as sum of this union of these error. So, this can be using bound, I can write this as upper bounded by probability of b I sum over all is from 1 to k. And we know this probability of b I which is given by this expression. So, if you plug this value in here, what we get is, this expression. So, probability of failure is upper bounded by k divide by l epsilon square into p min

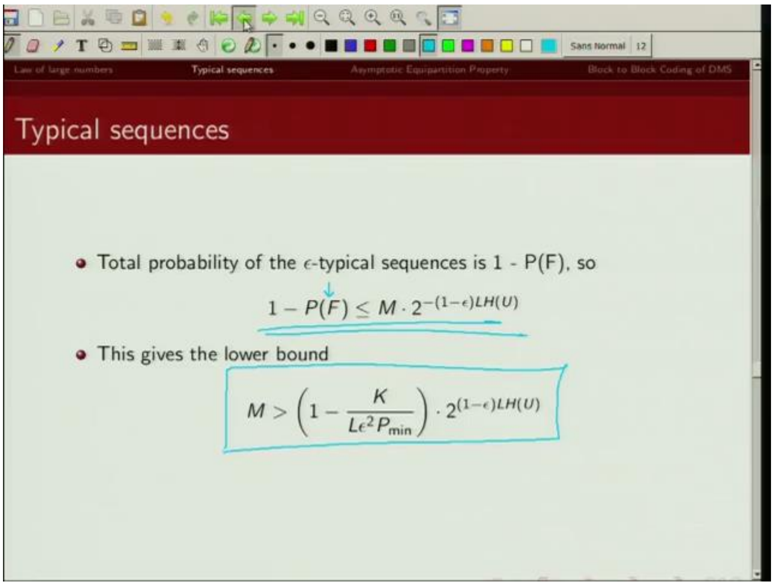
(Refer Slide Time: 48:12)



.

So, if we plug in this value and this expression, you can see we get the second property of a typical sequence that 1 minus p of f is greater than equal to 1 minus k times divided by k divided by l epsilon square p minimum. So, this is the second property of typical sequence, and what does it says is if l is very large you can see this probability of p f is a failure, so 1 minus p f that probability is almost equal to one. So, there is no probability of failure. So, if l is very large more of the sequences that come out of this source are going to be typical sequence. Now let us prove the third property of typical sequence which is related to how many such typical sequences exists. So, the number m of number of typical sequence, from a k ary discrete memoryless source with entropy given by each of u it satisfies this relationship. Now to prove this let us proceed. So, we know that p u sum over all u's this is given by one. Now using property one, we have computed lower bound and upper bound of p of u. So, if you replace p u by it is lower bound which is this quantity. If I replace p of u by it is lower bound and sum over all typical sequences, then this would be, this summation would be less than one. So, because I am replacing this by a lower bound on the probability of u for typical sequence, and number of sequences will have typical sequence as well as non typical sequences. So, then I can write that m times this quantity is less than equal to one. And from here I get this condition that number of typical sequences, is upper bounded by 2 times 1 plus epsilon into l times entropy of u. So, this proves this condition. Now we are going to prove this.

(Refer Slide Time: 51:04)



Now the total probability of epsilon typical sequence is 1 minus p f which if i. Now, if I replace p of u, but upper bound on p of u which is given by this quantity. So, then the total probability of epsilon typical sequences is upper bounded by m times, m is the number of typical sequence into probability of typical sequence, this is the upper bound on probability of occurrence of typical sequence. So, then this total probability of this typical sequence can be upper bounded by m times upper bound on the probability of occurrence of typical sequence. This we get from property one. And what is this quantity, we have already derived in property two. So, we plug in that value. So, from this expression and plugging in the value of p of f which we defined in property two, which is this. Probability of f is upper bounded by k divided by l epsilon square p minimum.

So, if we plug this one in here. So, we are putting an upper bound on this just particular larger number, then from here we can get m is greater than equal to this quantity. So, this is a lower bound on m, and earlier we have computed an upper bound on m. So, combining both of them we get expression upper bound and lower bound on how many such typical sequence exist. Now let us try to look at these results, when n goes very large that is asymptotic result, when l is very large. So, we have proved, we have shown three properties of typical sequence and we have proved them.

So, now let us look at what happens when l is very large and epsilon is small. When l is very large and epsilon is very small look at property number three, what does property number three says. Number of typical sequences is basically given by this, if epsilon is small and l is very large l is much larger than epsilon square that this term is almost going to zero, when epsilon is a small number, we can say that number of typical sequence, are roughly 2 raise power l times h of u. So, from property one, for a large l and small epsilon we can see that total number of typical sequences 2 raise power l of h of u. now what does property one says; property ones give us the probability of occurrence of these typical sequence. Again just refresh our memory and see what was property one.

(Refer Slide Time: 54:14)

.

If you go back and see what is property one or property one was this. It talks about probability of occurrence of this typical sequence. Now if you let epsilon to be very small and l to be very large, you can see this probability of occurrence of typical sequence is roughly 2 raise power minus l of h u. So, from the property one, the result that we get is as follows. So, from property one says that each of these typical sequences. So, that total this many epsilon typical sequence, and each of these typical sequence occur with probability 2 raise power minus l h of u. and finally, from property two we get this result, that total probabilities of these epsilon sequence is nearly one. Go back, what is this total probability; that is 1 minus p f. and what is 1 minus p f. 1 minus p f is upper bounded by this quantity. Now again when epsilon is small, and a l is much larger than epsilon square this term will go to towards zero. So, then this total probability of epsilon sequence is nearly one. So, these three properties collectively are known as asymptotic equipartition property of this discrete memoryless source.