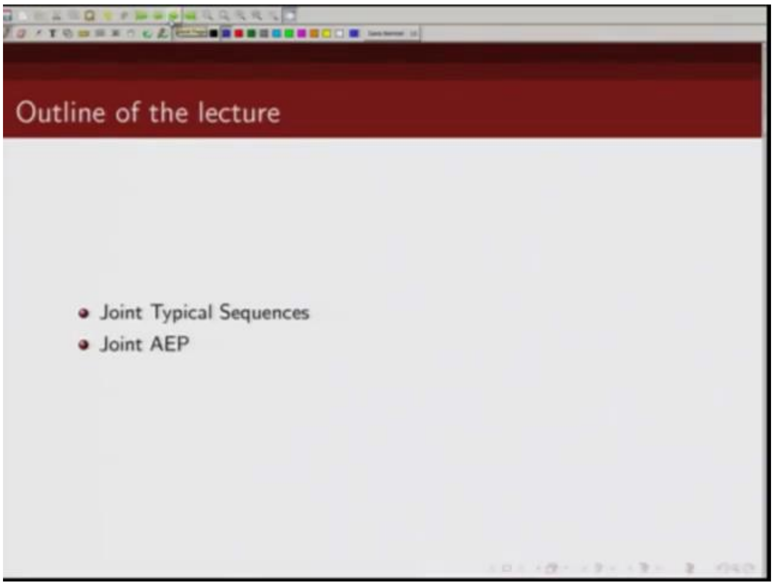
An Introduction to Information Theory

Prof. Adrish Banerjee
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kanpur

Lecture - 10A
Joint Typical Sequence

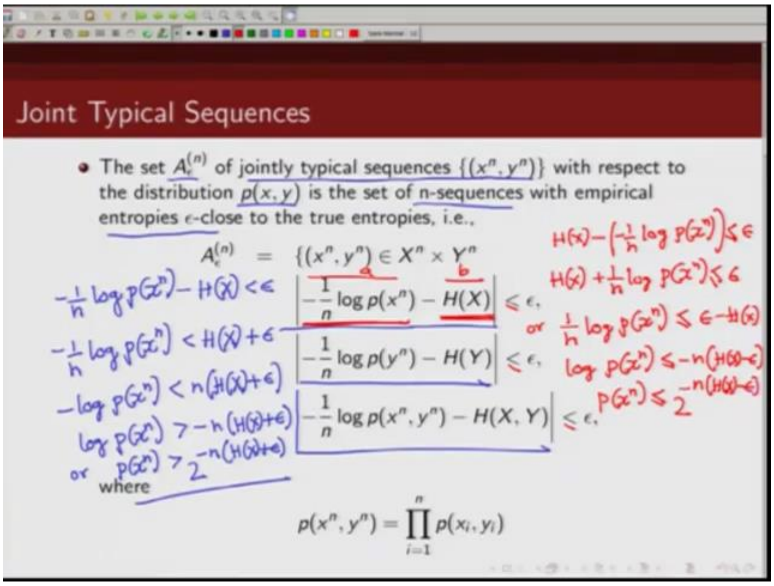
Welcome to the course on An Introduction to Information Theory. In today's lecture, we are going to first talk about, Joint Typical Sequence. Now to prove Shannon's noisy channel coding theorem, we are going to make use of typical set decoding and to explain set decoding, we first need to explain, what do we mean by a Joint Typical Sequence, and what are the properties of Joint Typical Sequence.

(Refer Slide Time: 00:51)



So, this lecture is about Joint Typical Sequence and, it is properties which are joint, which are known as Joint Asymptotic Equipartition Property.

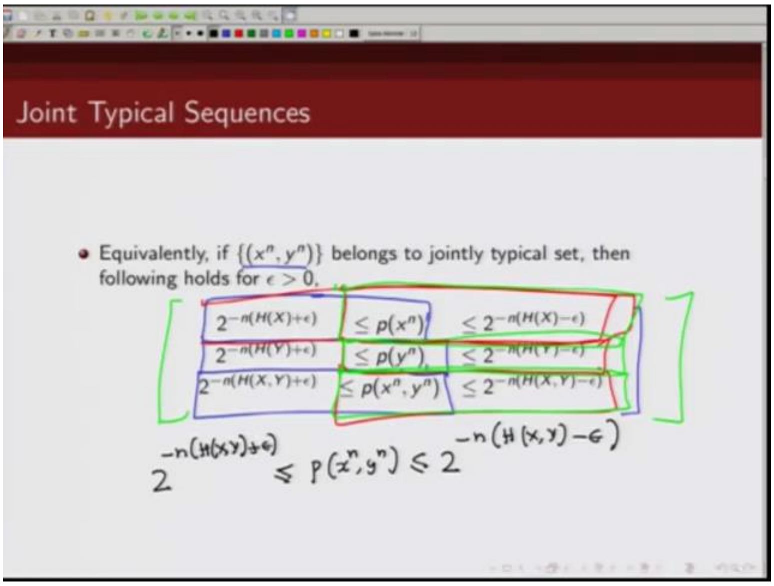
(Refer Slide Time: 01:02)



So, a set of Jointly Typical Sequence of this pair X, XN and YN, with respect to this joint distribution, PFXY is a set of n-sequences, whose empirical entropies is epsilon close to the true entropy. So, the Joint Typical Set is defined as follow. So, it is a pair of (Refer Time: 01:36) variables, XN and YN, XN belonging to XN and YN belonging to YN, which satisfies these properties.

So, this is the true entropy of X, and this is the empirical entropy of X. And what we are saying is, the absolute difference of this, is less than epsilon. And that is what we meant when you said, the set of typical joint, Jointly Typical Sequence which respect to distribution, joint distribution of P of X and Y, is the set of n- sequences, whose empirical entropy is, is close to the true entropies. Similarly this is the true entropy of Y, and this is the empirical entropy of Y. This is the true joint entropy of X and Y, and this is the empirical entropy of the joint distribution, P of X, XN and YN. Now note that, for X and Y to be jointly typical, we need all these 3 conditions to be satisfied. So, if X and Y is a N (Refer Time: 3:06) and it is a jointly typical sequence, then it will satisfy these 3 properties.

(Refer Slide Time: 03:13)



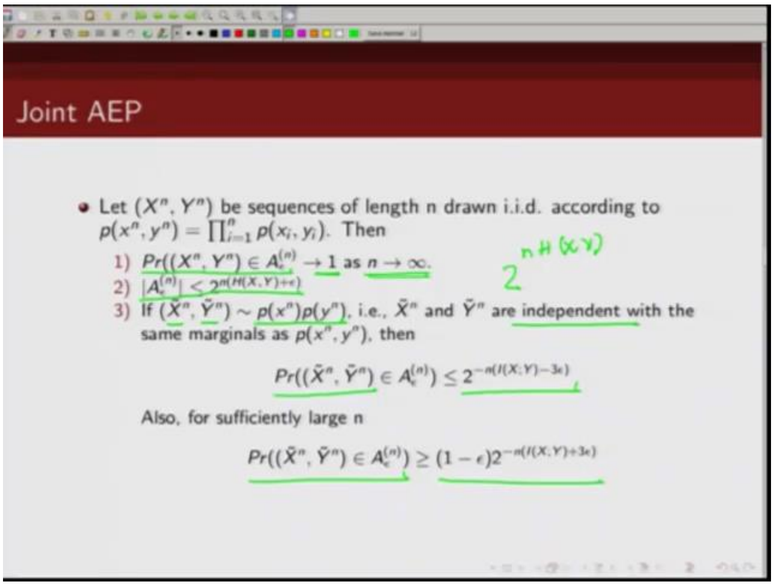
Now equivalently, we can also write that, if X N and Y N are joint they belong to the Jointly Typical Set, then these 3 relations hold, and this come directly from the conditions that we have written here. So, you can see from this condition. So, 1 minus, 1 by N minus, log of P of X. N, when this is greater than true entropy H of X, then this has to be less than epsilon.

So, from here, we can write 1 minus N, log of P of X N to be less than H of X, plus epsilon, and minus of log P of X N is less than N times H of X, plus epsilon, and we multiply, if we multiply both side by minus 1, we get log of P of X N to be greater than minus N of H of X plus epsilon, or P of X N, is greater than 2 the raise to power minus N H of X plus epsilon. And this is what, I have written that P of X N is greater than 2 raise to the power minus N H of X plus epsilon. Similarly, we can see from this condition we will get this one, and from this condition following the same procedure, we will get this condition.

Similarly we can prove, that P of X N is less, than equal to 2 raise to power N H of X minus epsilon, and we can also prove this, as well as this. So, let us go back and look at this. So, if you have, let us say this is my A, and this is B. Now if B is better than A, then I will write it is as B minus A, is less than equal to epsilon. So, H of X, minus of minus one by N, log of P of X N, that has to be less than equal to epsilon. So, this becomes H of X, plus one by N, log of P of X N. This is less the equal to epsilon, or I can write, 1 by N log of P of X N, to be less than equal to epsilon minus H of X, or log of P of X N, can be written as less than equal to minus N, H of X minus epsilon, or P of X N, is less than equal to 2 raise to the power, minus N, H of X minus epsilon.

So, this, this condition can be written equivalently like this. So, we just now showed that P of X is less than equal to this. Similarly we can show, that P of Y N is less than equal to this, and joint probability of X N Y, this is less than equal to this. So, these will follow from this condition, and this condition. So, in other words, this follows from the definition of Joint Typical Set. So, if X N, X N and Y N, they belong to this Joint Typical Set, than we know these conditions should be satisfied.

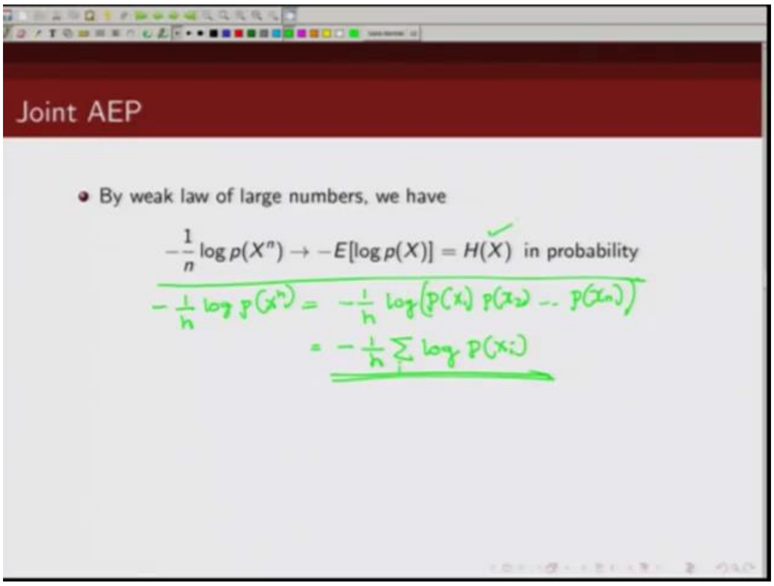
(Refer Slide Time: 08:08)



Now, let us prove some properties of these Joint Typical Sequences. So, if X N, Y N are sequences of length N and they are drawn independent, identically distributed, according to this distribution, then these following condition hold. What is a first property? Probability that X N and Y N belongs to Joint Typical Set, that probability approaches one when these length of these sequences is very, very large. So, probability that, X N and Y N are jointly typical, that probability goes to 1 as N increases, N is very large, and how many such typical, jointly typical sequences of there. That is basically upper bounded by, 2 raise the power N, joint entropy of X and Y, plus epsilon. So, is roughly 2 raise the power N, H of X Y these many numbers of jointly typical sequences are there.

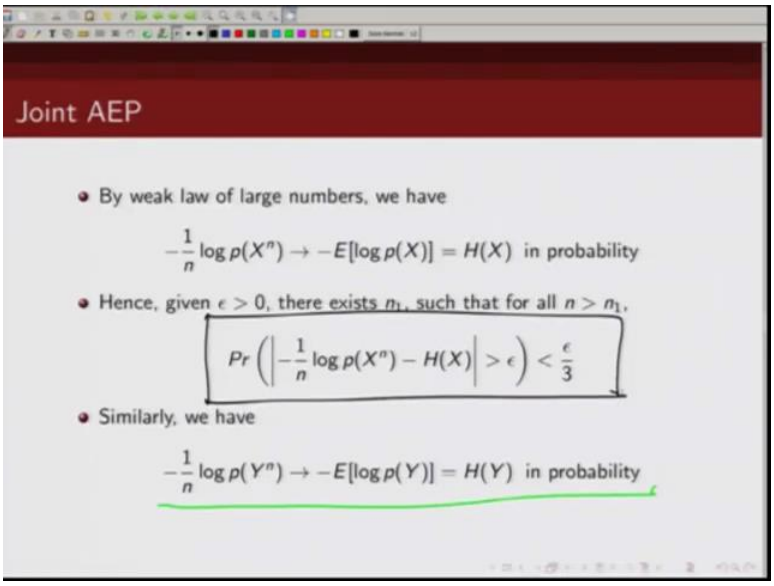
And what is the third property? If X N tilde, Y N tilde, these are independent endless sequences, with same marginals, given by this, than probability that, X N tilde and Y N tilde, that they belong to this Jointly Typical Set, this probability is upper bounded by this, 2 raise to the power, minus N, mutually information between X and Y, minus 3 epsilon. Or you can also show that, probability if X N tilde, Y N tildes are independently drawn, than probability, that they are jointly typical, that is upper bounded by this property.

(Refer Slide Time: 10:33)



So, we are going to prove one by one, each of these properties of Jointly Typical Sequence. So, we know from law of large numbers, that minus 1 by N, log of P X N, that converges to entropy, and this converges is in probability. This can be shown easily. So, this is 1 minus 1 by N, log of P of X N, so P of X N is IIT distributed. So, I can write it has is equal to minus 1 by N log of P of X 1, P of X 2, like that P of X N. Now this can be written as minus 1 by N. So, this is log of products that can be written as sum of log term. So, this can be written as summation log of P X I where I, is 1 to N, you can write it like this and this is expected value of log of PFX, and this converges is in probability, this flows from the weak law of large numbers.

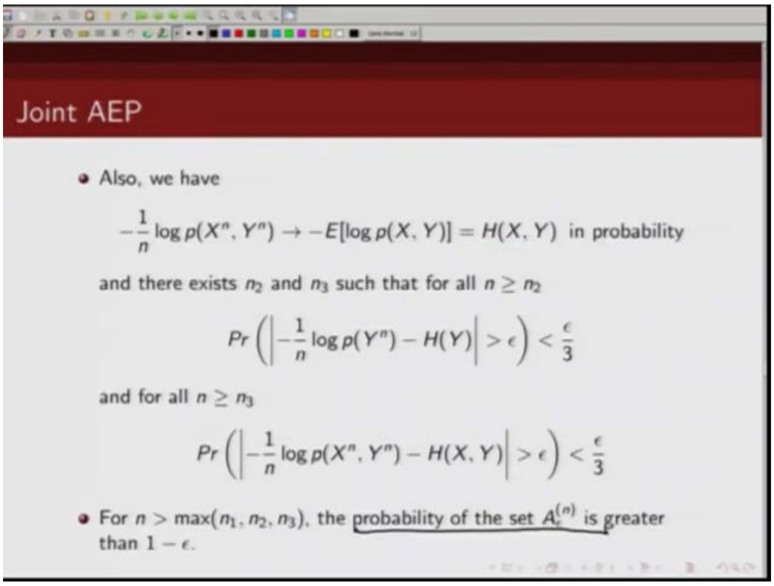
(Refer Slide Time: 11:49)



Equivalently, I can write this same condition, like this. So, probability that, expected value of minus log of P X, and absolute difference between that, and the true entropy, that is absolute difference greater than epsilon, is less than sense of small epsilon, this quality, epsilon dash 1 or something like that, which we are choosing epsilon by 3.

So, this is equivalent to this statement, that minus 1 by N, log of P of X N, converges to entropy, and this converges is in probability. So, this is an alternative way writing the same thing. Similarly we can write that, minus 1 by N, log of Y N, it converges to entropy of Y, and this convergence is in probability, and to joint distribution of X N, and Y N, that also converges through this joint entropy of X and Y, in probability.

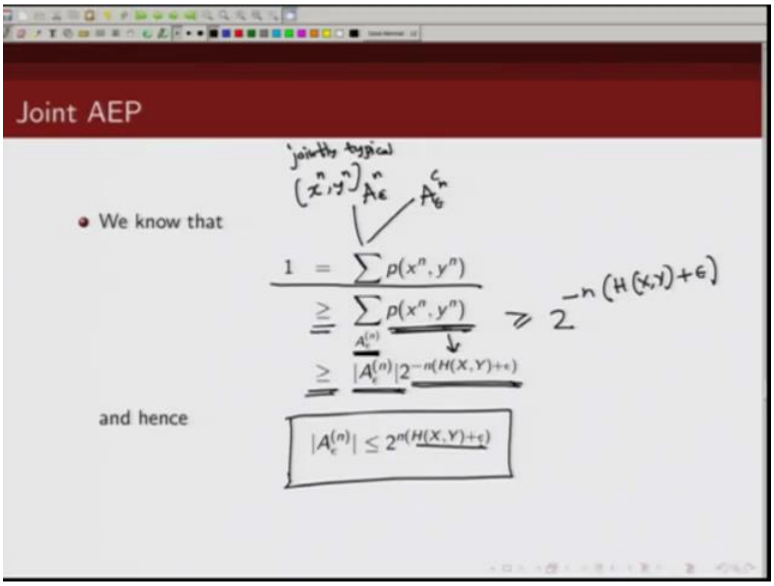
(Refer Slide Time: 12:56)



And as we said, we can equivalently write this conditions, as follows, that there exist, lets say, some N2 and N3, such that for all N greater than N 2, probability that the absolute difference between the empirical entropy, and the true entropy exceeding epsilon, is less than some small epsilon, we calling it an epsilon by 3 and similarly, probability that the absolute difference between the, true joint entropy and the empirical joint entropy, exceeding epsilon, is less than some epsilon by 3, and this holds for all N, greater than equal to N3.

So, then, we can combine these 3 conditions, one is this one, another is this one, and the third one, is this one. So, if we combine these 3 conditions, what we get is the probability of this set of Jointly Typical Set. This probability is greater than 1 minus epsilon, for all N, which is greater than, the maximum value of N1, N2, N3. As we know that, from the property of Joint Typicals, Joint Typical Set, we know that these 3 conditions have to be satisfied, right? These 3 conditions have to be satisfied if, XN, YN belongs to a Jointly Typical Set. So, what we have shown right now is, that, probability that X, XN ,YN belongs to the Joint Typical Set, which is satisfying these 3 conditions that happens for N greater maximum of N1, N2, N3, and we have shown that.

(Refer Slide Time: 15:18)



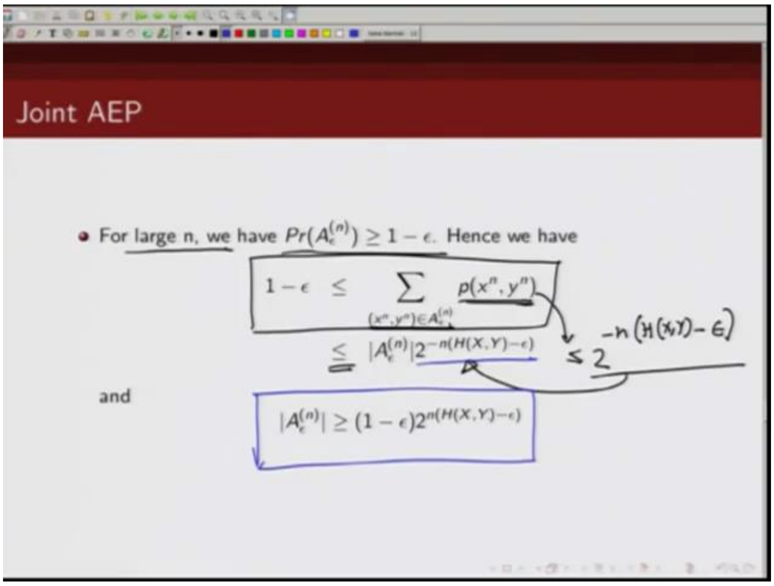
If you add all of these, we get the probability of typical set, to be greater than 1 minus epsilon, because probability of this is less than epsilon by 3, probability of this is less than epsilon by 3, and this probability is also less than epsilon by 3. So, if we add these probabilities, we get overall probability less than epsilon.

So, in other words, probability that the, it belongs to the typical set is then, 1 minus epsilon, which is the compliment of this probability. So, we have proved the first property that, if N is large, X N and Y N probability that, they belong to Joint Typical Set. That probability is 1 minus epsilon, and epsilon is small quantity so this probability goes to 1 as N increases. Next property is regarding the size of the Jointly Typical Set. So, how large is the size of the Jointly Typical Set. So, to prove this, we know that, when distribution X N, Y N sum overall, X and Y, that probability should be 1.

Now if we sum them over only, Jointly Typical Set. So, you can think of it, this summation is over, 1 set where X, X N and Y N they are jointly typical, jointly typical, which is this, and then there is a set, which is a compliment of this, right? Compliment of this, Jointly Typical Set, if I sum it over only, Jointly Typical Set, I get the sign that, this summation should be, less than equal to 1, and invoking this property, which follows from the definition of Jointly Typical Sequence, I know from here, P of X N, Y N this is less than equal 2 this, and greater than equal to, this quantity.

So, using this relation, if we plug-in the value of P of X N, Y N here, we know that, this is greater than, equal to, 2 raise power minus N, H of X Y, plus epsilon, right? And, so, we are replacing this by smaller quantity, so that is why, we again get, greater than equal to, and so, number of elements in this Jointly Typical Set multiplied by lower bound, on this joint probability. So, this has to be less than equal to 1, and from here then, we get the condition that, this number of elements in this Jointly Typical Set, is less than equal to, 2 raise power N times, joint entropy of X and Y, plus epsilon.

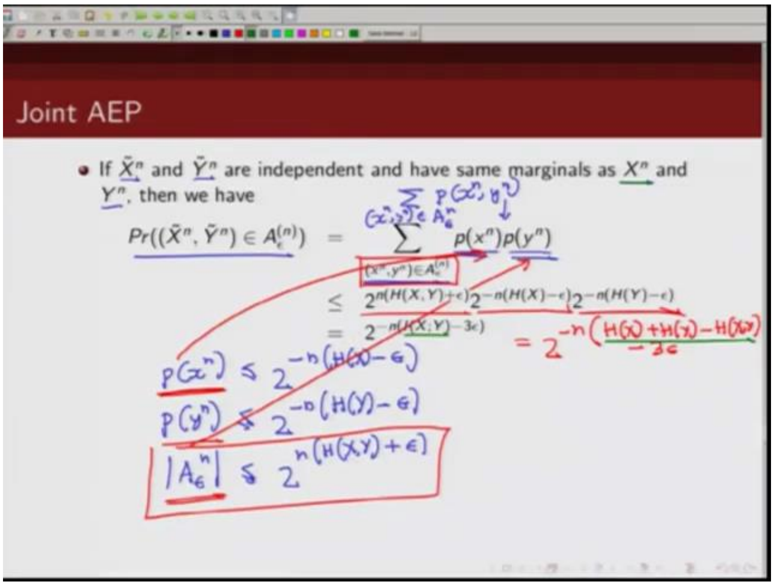
(Refer Slide Time: 19:45)



Next, we have shown that, as N goes to infinity or N is large, we have shown probability that, X N and Y N are jointly typical, that probability is greater than, equal to 1 minus epsilon. So, then, we can write that, 1 minus epsilon is less than equal to, this probability, why? Because when N is large, probability that X N and Y N are jointly typical, that probability is greater than 1 minus epsilon. So, when you are summing this over, this Jointly Typical Set, this summation is greater than 1 minus epsilon. And we know, that this joint this is upper bounded by 2 raise to power minus N, H of X Y, minus epsilon.

This follows from the properties of Jointly Typical Set. So, if we upper bound this, by this quantity, and we plug this in here, we get this less than equal to sign here. So, what we get is 1 minus epsilon is less than equal to, number of elements in this Jointly Typical Set, multiplied by this, and from here, we can get a lower bound or number of elements in this Jointly Typical Set, which is basically greater than equal to 1 minus epsilon, 2 raise to the power N times, joint entropy of X and Y minus epsilon. And epsilon is very small, this is roughly you can see that, roughly 2 raise to power N, H of X Y elements in this Jointly Typical Set.

(Refer Slide Time: 21:49)

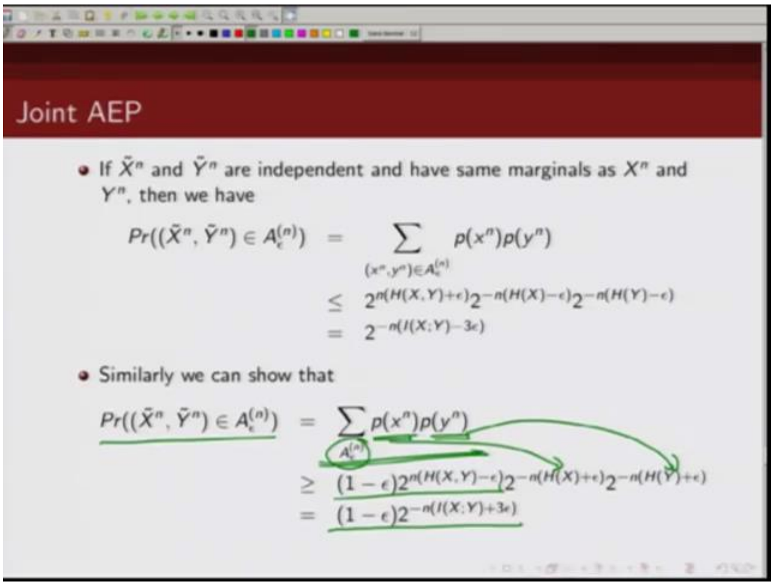


The third property that we are going to prove is as follows. So, X N tilde and Y N tilde are independent, and they have the same marginals as X of N and Y of N. Now we want to find out, if we independently take X N and Y N, what is the probability that they belong to that Jointly Typical Set.

Now this probability can be valid (Refer Time: 22:24) X, X N tilde and Y N tilde are independent. So, we can write this summation of P of X N, Y N sum over, this, Jointly Typical Set. We can write this as, product of these. And we are summing up over, this Jointly Typical Set. Now, we know that P of X N is upper bounded by 2 raise power minus N, H of X minus epsilon. We know that, P of Y N is less than equal to 2 raise to the power minus N, H of Y minus epsilon. Again these follow from the properties of the Jointly Typical Set. And we know that, number of elements in here, is upper bounded by 2 raise to power minus, H of X Y plus epsilon, right? So, if we plug-in these values of P of X N, P of Y N here, right and some over, all X, X N, Y N belonging to this Jointly Typical Set, which is number of such element is upper bounded by this, if you do that, what we get is this term, which we are getting from upper bound on number of elements in this Jointly Typical Set, this term, which we get by upper bounding on probability of X N, and this we will be getting on P of Y N, upper bound on this.

So, we combine all of this, what we are getting is 2 raise to power, minus N, H of X, plus H of Y, minus H of X Y, and we get minus 3 epsilon. Now this particular term, H of X plus H of Y minus H of X Y, this is nothing, but mutual information between X and Y. So, we plug that in here, we get this third property, which says probability, that if X N and Y N tilde are independent, and have same marginals as X of N and Y of N, then probability that, they are jointly typical, is less than 2 raise to power minus N, times mutual information (Refer Time: 25:31) X and Y minus 3 epsilon.

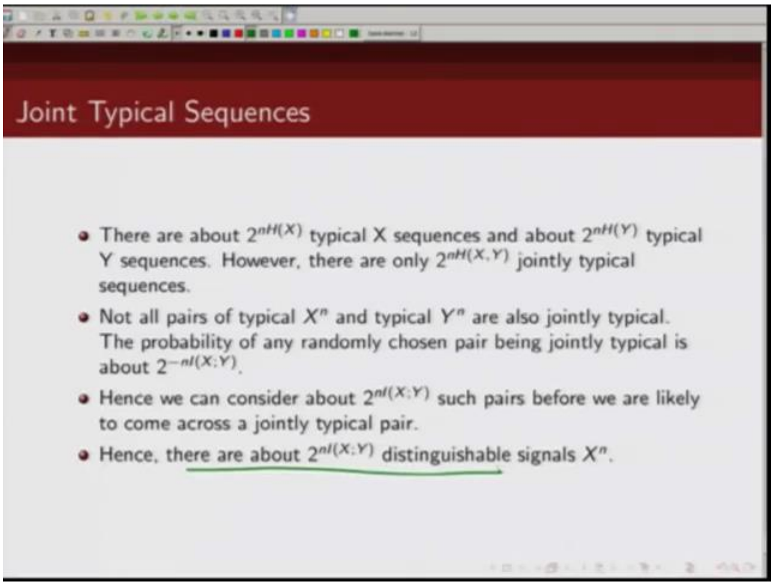
(Refer Slide Time: 25:30)



We can similarly, also get a lower bound on, the probability that X N tilde and Y N tilde, belongs to a Jointly Typical Set. And how do we get this? So, again we start with summation of P of X N, Y N over this typical set. Now we are going to lower bound these probabilities. So, this is lower bounded, P of X N is lower bounded by this, and P of Y N is lower bounded by this, and similarly number of elements in this Jointly Typical Set, that is lower bounded by this quantity.

So, if you plug, plug-in these values, what we get is, probability that X N tilde Y N tilde which are independent, are jointly typical, is given by, is, is, is lower bounded this probability.

(Refer Slide Time: 26:36)



So, what we have seen so far is, so, we have total these many numbers of typical X sequences, and these many numbers of typical Y sequences; However, they are only 2 raise power N times, joint entropy of X and Y, number of jointly typical sequences. So, if you pickup independently let us say, typical sequence from X, and typical sequence Y, they will not be jointly typical. So, not all pairs of typical X N and Y N sequences are jointly typical, because we only have so many numbers of Jointly Typical Sequence. In fact, this probability is given by this, divided by this, into this, which comes out to be this. So, probability of any randomly pair, chosen pair, of X N and Y N which are typical, probability that they are jointly typical, this probability is given by this.

So, in other words, what we can say is, we can consider around 2 raise to the power N times mutual information of X N. We are going to encounters this different number of pairs, before we are likely to get one jointly typical pair. So, as I said, because the probability of randomly chosen pair, being jointly typical is this. So, we can roughly encounter so many such pairs, before we are likely to encounter a, jointly typical pair. So, in other words, we can conclude that, we have total of 2 raise to power, N times, mutual information of X and Y, those many numbers of distinguishers signals.

So, what we have done so far is, we have described what is Jointly Typical Sequence, and we have described some of the properties of Jointly Typical Sequence, namely what is the probability of occurrence of Jointly Typical Sequence. How many such Jointly Typical Sequence exists. And if we randomly pick 2 typical sequences, what is the probability that, jointly they are also typical. That we have shown. Now these properties we are going to use, to prove our Noisy Channel Coding Theorem, which we will talk about in the next lecture.

Thank you.