Fundamentals of MIMO Wireless Communication

Prof. Suvra Sekhar Das
Department of Electronics and Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture – 31
Fundamentals of Information Theory-I

Welcome to the course on fundamentals of MIMO Wireless Communications. Till now we have covered the diversity gains achievable in MIMO communication, in which we have seen received diversity, transmit diversity as well as diversity that can be obtained when there are multiple antennas both are the transmitted side as well as the receiver side with this it is times. Now that we move forward and try to understand the capacity gains that can be achieved by use multiple antennas and which is one of the amongst the several gains, one of the most important gains that a MIMO communications brings. Now in order to study such details we would require some revision of concepts from information theory.

So, this particular lecture would be dedicated towards revising some of the concepts which are fundamental towards building the expressions the way we would like to do it is instead of giving the direct expression of capacity for MIMO systems. It is otherwise to start with the basics and do a buildup of the capacity that is achieved for discrete random variable or discrete sources and then, move on to see how the capacity is achieved for the Gaussian channel, once you are able to see that then it is more or less straight forward that we will see what is the capacity in the case of multi varied it Gaussian distribution.

So, once we have that expression then, that would lead us directly to the case for MIMO communications where all we need to do is use this build up. Use the expression that we achieved when there is multi varied Gaussian distribution and see how this maps to the MIMO system, whether a multiple antennas at the transmitter multiple antenna at the receiver and with one to one mapping would be able to use expression that is derived for multi varied Gaussian directly into the case of MIMO. So, that would ease our way of doing things as well as it has, it will give us an opportunity to understand how the expressions are as it is from the basics. So, with this let us begin with the some of the fundamental definitions that are used in arriving at the capacity expression for typical a w g n channel which we will be exploiting later on.

So, when we are talking of such systems, we are essentially talking about digital communication system and when we talk about digital communication system we would like to begin with the discrete sources. So, of course when we go on finally, we would be reaching a situation, where we have wave form channel. So, in the beginning of this course we had described how you look at the channel.

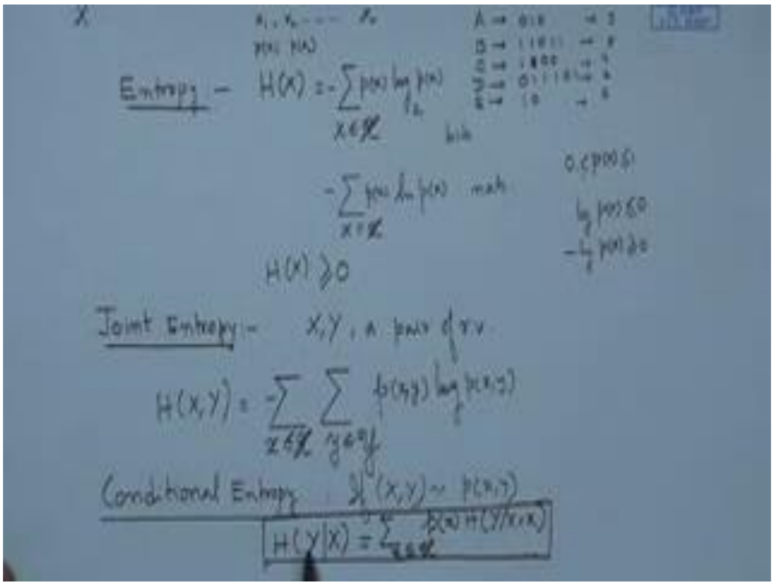
And at one point, we said that the channel could be looked upon as a discrete input discrete output as well as it could be wave form. So, it is a mathematically easier when you develop the discrete input discrete output configuration. And then that would be translated to the Gaussian or the wave form channel.

So, to begin with a let us take up discrete sources. So, when we look at a discrete source. So, we assume that there is finite alphabet at the source; that means, it has set of symbols one particular example could be the English alphabet, along with few punctuation signs. So, what we get in total let say there are 26 letters along with the few punctuation signs. So, which keep on coming the example, we could take is a if you take a file which has English characters and i would like to read them from the beginning. What i will be getting is the English letters from the English alphabet as well as punctuations one after another. So, we can say that the source is generating random symbols and because when, we look at the output of one file from the other or going look at the sequence there is no definite pattern unless it is premeditated. So, if it is not a premeditated there is you want to inject a pattern, if you look at any particular text at random the sequence of letters that would come out of it would random in nature.

So, when such things happen it is often important to measure the randomness of source, in this particular lecture we do not intend to describe the how do you measure randomness, why do you measure it in this way and all details i would like to directly give some definitions related to such sources which finally, build up towards measuring or towards defining how capacity is defined. And when we go down the course, we would also may be giving you a prelude to how the capacity expression is arrived at instead of giving the rigorous proof of the capacity expression, because here in this course our aim is to use the information theoretic capacity result where as detailed derivation of the exact capacity expression is rather more appropriate in a course on information theory.

So, with this background we would prefer to look into to the definition of entropy which is one of the fundamental measures of randomness of a source.

(Refer Slide Time: 05:27)



So, if we take that let x be the how to come of source, for instance i take this English letters which are like a, b, c, d and so on. So, basically x1, x2, basically you have x1, x2 dot, dot, dot as the as the outcomes we can always associate probability associated with each of the outcomes p of x1 p of x2. So, for instance that you look at the source we generates this English letters. So, if i take a file which has may be ten thousands of letters and then i try to make the histogram of the number of a s the number of bs and. So, on and i try to see a frequency plot.

Then what i can get is the probability of occurrence of a, probability of occurrence of b, probability of occurrence of c and so on and so forth. So, with the associated with each symbol there is a probability, with these there is a certain measure known as the entropy of the source. The entropy of the source gives you the inherent randomness in the source. So, we are again not going into details description of this randomness, but rather the exact definition that which is useful for us. The entropy as you said is also useful in describing the average number of bits 0s and ones, that could be required to encode the source in the most efficient way in other words it is the minimum number of bits that is required in order to encode. So, for instance we take these out comes that are coming out each with certain probabilities. So, given there is a symbol and there is certain probability and there is a certain probability and there is finite alphabets size let say x of n entropy which is usually given as h x we are going to define is the measure of number of bits that we would assign on an average to each of these. For example, i could assign to the bit sequence 010 i could assign to b the bit sequence 11011, i could assign to c for instance 100, i could assign to d for instance 011101, i could assign to e for instance one 0 and so on, right? So, this is the bit sequence. So, there on this as a length of 3, this has length of 1, 2, 3, 4, 5 this is the length of four this is the length of there and 3, 6 this is the length of 2.

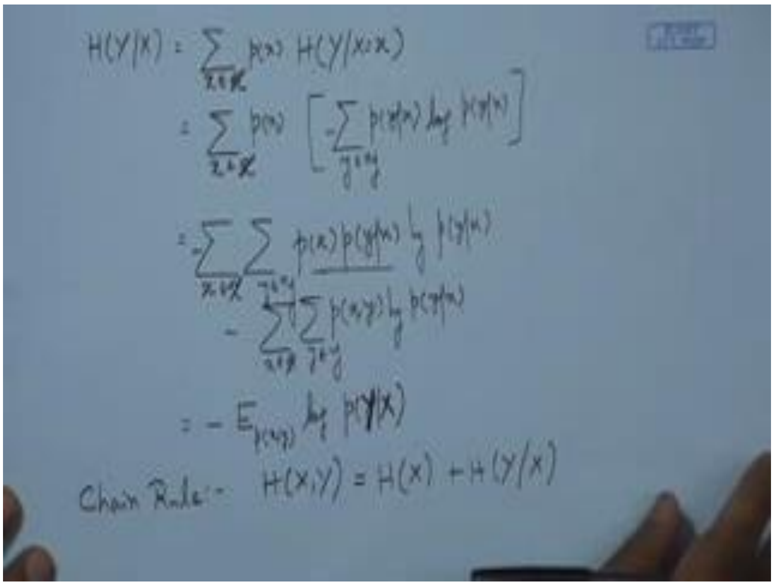
So, the average of these would be if i add them up and divide by 5, i am going to get the average bit length. So, what h x is going to give us is what would be the minimum value of the average number bits per such symbol that could be used. So, that is the measure of entropy and we all know what it is important for one direct example of this would be it would give us information about what is the maximum amount compression that could be done for the output of the sources. So, that is one way using it and of course, we are not going to use it in that sense our here just going to look at entropy because that. So, fundamental that will be use throughout until we arrive at the expression of capacity finally, So, h x is defined as minus sum of x element of this set which contains x p of x log base 2 p of x, when you define it in log base 2, we would give the units as bits. So, as many bits are required one could also define it as x element of big x p of x l n natural logarithm p of x n and in this case the definition of would go in nabs and in either of the case the we could see that we are taking p x, which is greater than or equal to 0 and rather we will be interested in the case of p is greater than 0 and since it is a probability this number is less than 1.

So, p x lies between one and 0 you could put an equal to, but you prefer not to do it because p x equals to 0 means that does not occur. So, it is not part of the alphabet. So, if you would look at this then the log of p x log of p x is definitely negative. So, this is less than or equal to 0, 0 equal to 0 when p x is equal to one and minus log p x x greater than or equal to 0. So, p x is positive and log p x is also positive so; that means, this sorry yeah minus log p x is also positive so; that means, this whole number is positive. So, we could say h of x is always greater than or equal to 0, this is one of the fundamental things that we should always remember that h x is never negative when, we have x as the output come of sources and p x as the probability of occurrence of such sources and remember we are talking about discrete random variable at this point of time. So, this is what you would call the definition of entropy ok.

So, with this we move forward and define a few more terms that would be necessary in due course of time and one of the next most important would be joint entropy the joint entropy of pair of random variables x, y which is a pair of random variables. So, we would have h of x y is equal to minus x element of big x which is whole set of that contains all the values of x y element of big y, which contains all y joint distribution or joint mass function of p of x y times log p x y in general we will use a log to indicate log base 2 or l n to indicate the natural log logarithm.

So, this is the definition of joint entropy the next important definition that, we need is conditional entropy. So, again i said we are not going to describe this much, but just keep the definition for joint entropy briefly, we could say that what is a probability what is a entropy that a two sources have certain entropy together when, there are two sources generating it. We could talk in that sense conditional entropy if given that; that means, if we say that x and y are distributed according to p x y then, what is the conditional entropy conditional entropy h is denoted as h of y given x this is a notation which would be some of x element of all of this set of xp of xh of y given x is equal to x. So, this is the basic definition of conditional entropy of y; that means, the entropy of y given x entropy of y given x.

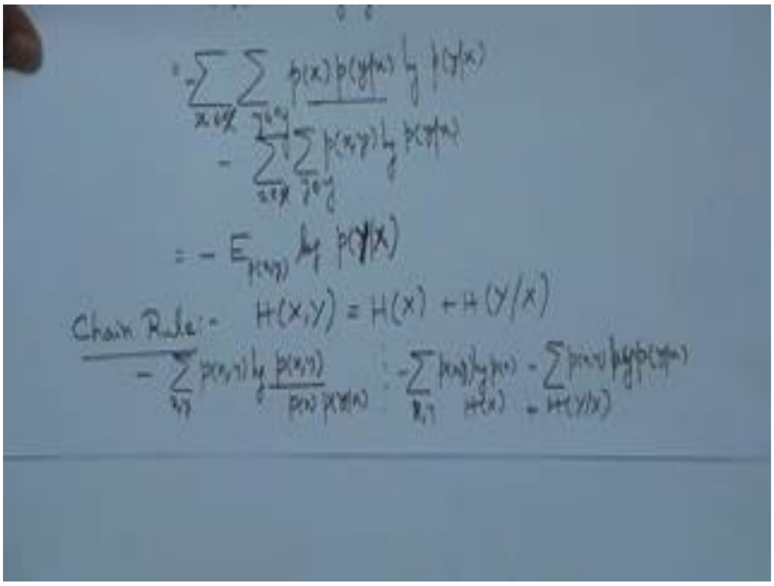
(Refer Slide Time: 13:45)



So, if we would proceed with this, what we get here is this particular expression or we can write it again h of y conditioned on x is equal to sum over x element of big of x p x h of y even x is equal to x. So, you are waying this with the probability and you could write this as x element of capital x p of x sum of this particular, one we could open it up saying it is y element of capital y minus p of y given x log of p y given x this is straight forward given the definition of entropy above. So, if we look at definition of entropy here following that we could write this description.

So, going by this we could write this as x sum over y element of all of x all of y p of x times p of x y given x there is a minus sign from here which comes out log of p y given x. So, this one is basically p of joint distribution of p of x and y sum over y sum over x log p y given x. So, this you could also write as minus expectation over the joint distribution of x and y log p y given x log p of y given x right.

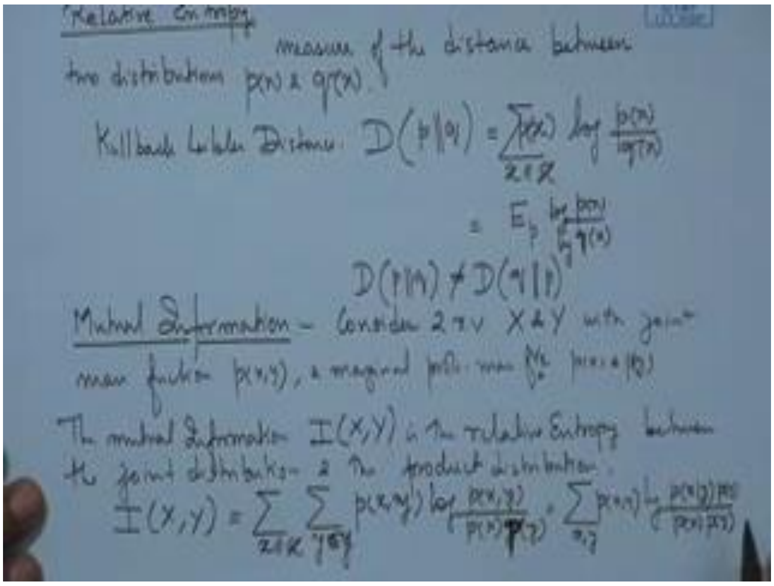
(Refer Slide Time: 15:58)



So, this is a what we could write we could also have the chain rule following this the chain rule says h of x y the joint entropy is equal to h ox plus h of y condition term of x. Now this could be done if you would write left hand side as p of x y over all of x and y log p x y now p x y you could break it into p of x by base theorem p of y given x, now this is straight forward you would have this multiplied by log p of x.

So, when this is added over y you are left with minus summation. So, we could rather do it you would be left with minus x y p x log p x right and another term minus p x y times p of log p y given x. So, this is p of x comma y. So, this is basically when, added over y y goes out adds up to 1. So, this would be h x and this whole term as we have just seen basically h of y given x. So, basically h of x y joint entropy is equal to entropy of one of them plus conditional entropy of the other with respect to the first. So, this is also what would be useful.

(Refer Slide Time: 17:29)



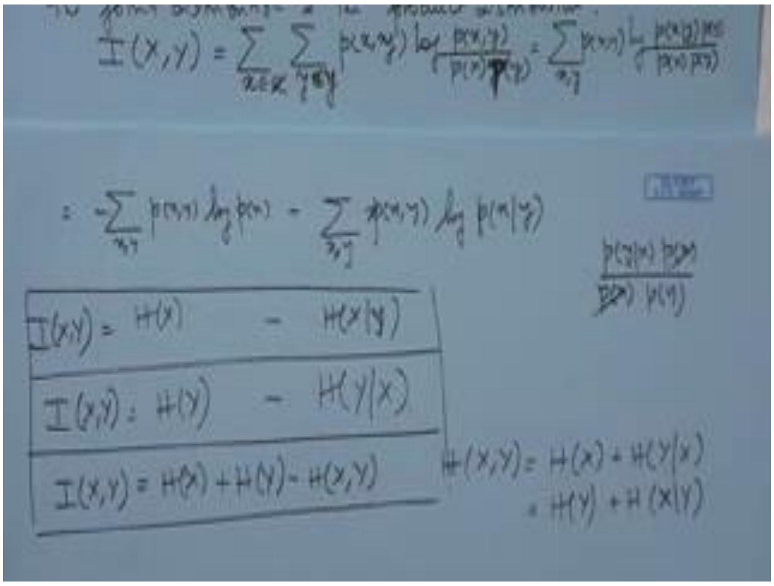
So, with that we move on further to talk about relative entropy. So, again we have given a list of definitions which are important relative entropy and we will talk about mutual information. So, we are talk of relative entropy, if we would write this as the measure of the distance between two distributions p of x and q of x it is also known as the Kullback Leibler distance and is denoted as p d of p with a parallel line between q; that means, we are indentifying the distance between the distributions.

Now, you can easily calculate distance between 2 points, but this is between 2 distributions is not very easy to find. So, there is a specific definition to it and that is particular definition that what we are looking at in this particular case. So, when we look at the relative entropy or the Kullback Leibler distance this is given as x element of p of x times log of p x over q x and this could be written as expectation over the density of p of the distribution of p times log of p x by log of q x and we would remember that, we should remember this distance of p from q the distribution of p from q as per this definition is not equal to the distance of the distribution of q from p this is also fundamental.

So, using these definitions that we have used, we would now give the definition of mutual information which is again fundamental in the expression of capacity. So, the expression of a mutual information mutual information, we look at it we would consider two random variables x and y with joint mass function p of x y see we are talking about discrete random variables. Therefore, we are talking about mass functions and marginal probability mass functions i would write in short functions in this way p of x and p of y.

So, this is the notation that we have been using the mutual information i of x y it is denoted as i x y is the relative entropy, see capacity is defined in terms of mutual information mutual information is defined in terms of relative entropy and relative entropy is defined in terms of entropy. So, all this things are following in suit and we have to develop one after the other, the relative entropy between the joint distribution and the product distribution. So, in other words i of x y is given as sum over x element of all of x times sum over y element member of all of y the joint over the p f x y. Say it is relive entropy between the joint distribution and the product distribution. So, here if you look at it distance between two distributions p x and q x. So, it is given as sum over p x log p x by q x same things follows over here, joint log p of q of x divided by p of x times q of y right. So, between these two distributions sorry; it is p of x times p of y. So, basically if you look at this it directly follows from the relative entropy and between p x y and x times p y. So, it is basically the distance between the joint distribution to the product distribution and some of the things you could get cleared that the this could be expanded easily as the we could write in short as sum over x y; that means, all of x y you are writing in short all of x y p x y times log of p x given y times p of y divided by p of x times o of y p of y p of y cancels.

(Refer Slide Time: 23:18)



So, with that we could write this further as this is equal to yeah sum over p x y over all x all x y times see log is there in the denominator p x in the denominator. So, minus log p x because we have p x in the denominator log of p x. So, minus log p x plus some over over all the x y p of x y that is p x y times log of p x given y, you can easily recognize this is basically x of x and this is basically minus of h x given y. So, this is what we are built earlier. So, basically i x y could be written in terms of h x and h of x given in the similar manner we could expand this part into p. If we would write p of y given x times p of x times p of x times p of y what do we get? These two we can cancel out if they are non 0 and then i of x y you can easily guess from this it will be h of y and from this minus it will be h of y given x. So, these are the definitions that we can follow and if you would go by the chain rule the chain rule tells us h of x y is equal to h of x plus h of y given x or you could also write it as h of y plus h of y x given y.

So, this you could write it as h x plus h y minus of h of x y. So, these are the 3 descriptions of mutual information which would be finally, using that we can write. So, here, we can see that h of y given x is h of x y minus h x. So, if we look at this one h of y given x. So, h of y given x is equal to h x y minus h x. So, h x y minus this a minus sign h x when comes to this side the minus, minus and minus plus.

So, we have h x h y and this term comes over here. So, we have h x plus h y minus h x y. So, we have reached important point in our description of these important terms in terms of the entropy. So, we started with entropy we talked about joint entropy conditional entropy relative distance between the two distributions. Then we define mutual information which is again very, very important term, when we talk about capacity. So, we conclude this lecture at this point and we would look at some more definitions in the following lecture.

Thank you.