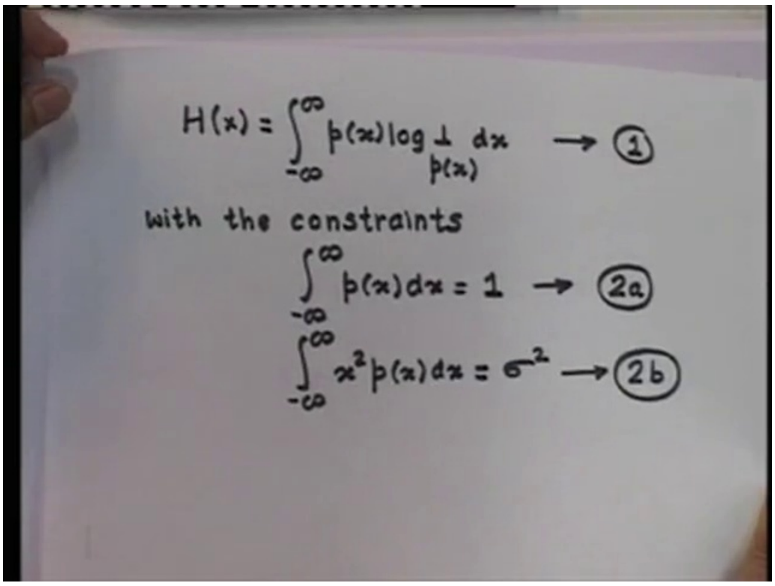
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 29
Differential Entropy and Evaluation of Mutual Information for Continuous Sources and Channels

In the last class we initiated the discussion on information transmission from a continuous source over a continuous channel. We define differential or relative entropy of the continuous random variable. During the course of our study we have also seen that for a discrete random variable the entropy is maximum, when all the outcomes are equally likely. In this class, we will investigate the problem of calculating the probability density function for a continuous random variable; that maximizes the differential entropy.

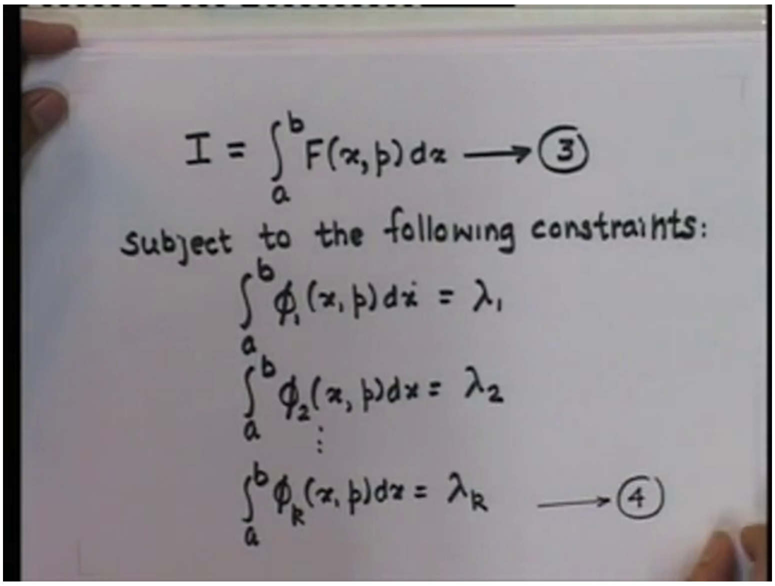
For a continuous distribution, however we may have additional constraint on the continuous random variable. For example, the constraint could be on the maximum value of the random variable or on the mean square value of the continuous random variable. Let us investigate the calculation of the PDF that is probability density function for the continuous random variable, that maximizes the differential entropy when the mean square value of the continuous random variable is constrained to be some constant.

(Refer Slide time: 02:54)



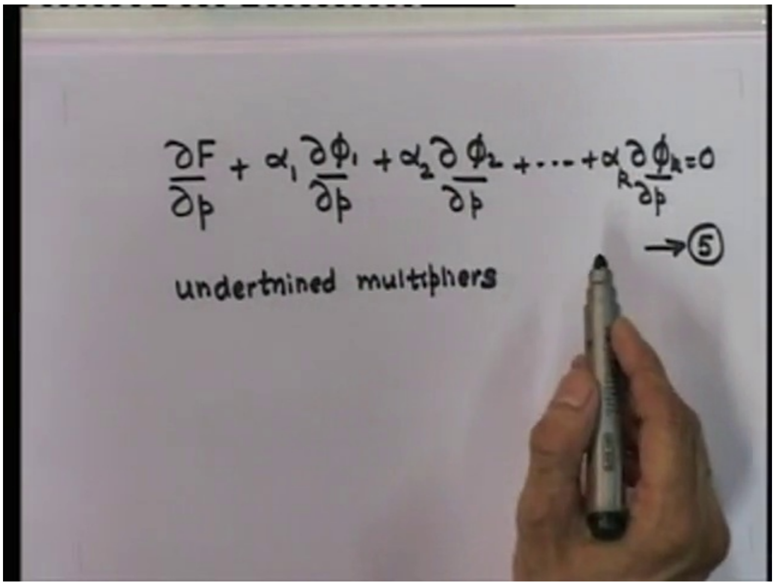
So, the problem is to maximize the differential entropy for the continuous random variable given as follows. With the constraints that the probability distributive function integrated over minus infinity to plus infinity is equal to 1, and the mean square value of the continuous random variable is constraint to be some constant sigma squared. To solve this problem, we use a theorem from the calculus of variation.

(Refer Slide time: 04:36)



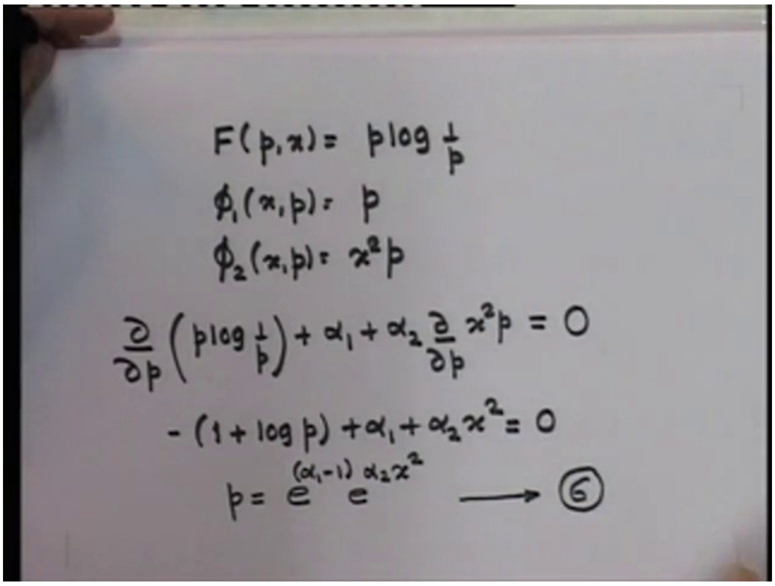
Given, the integral I equal to F function of x p dx subject to the following constraints, phi 2 x p dx integral is equal to lambda 2, where lambda 1, lambda 2, and lambda k are given constants. Now, the result from the calculus of variation states that the form of p x that maximizes I in equation number 3, with the constraints in the equation number 4 is found from the solution of the equation given as follows.

(Refer Slide time: 06:38)



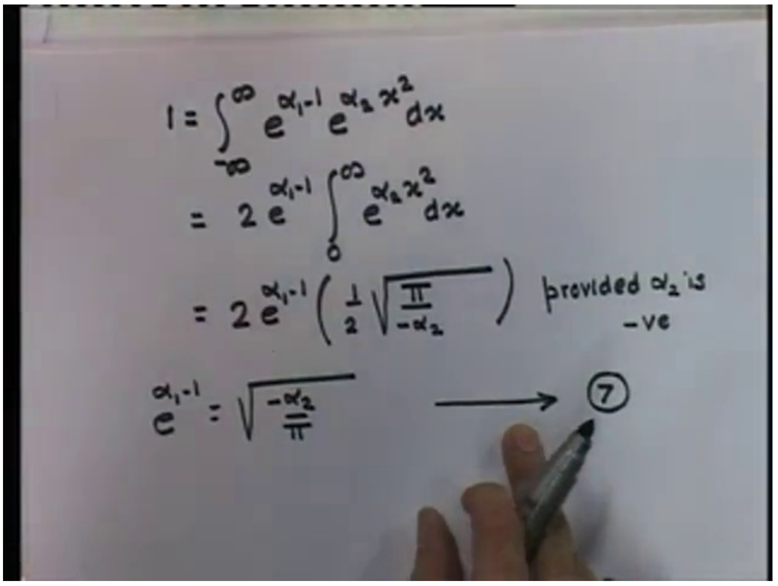
The quantities alpha 1, alpha 2, alpha k are adjustable constants called undetermined multipliers. Now, these multipliers can be found by substituting the solution of p x obtained from this equation in, equation number 4, that is all these equations.

(Refer Slide time: 08:08)



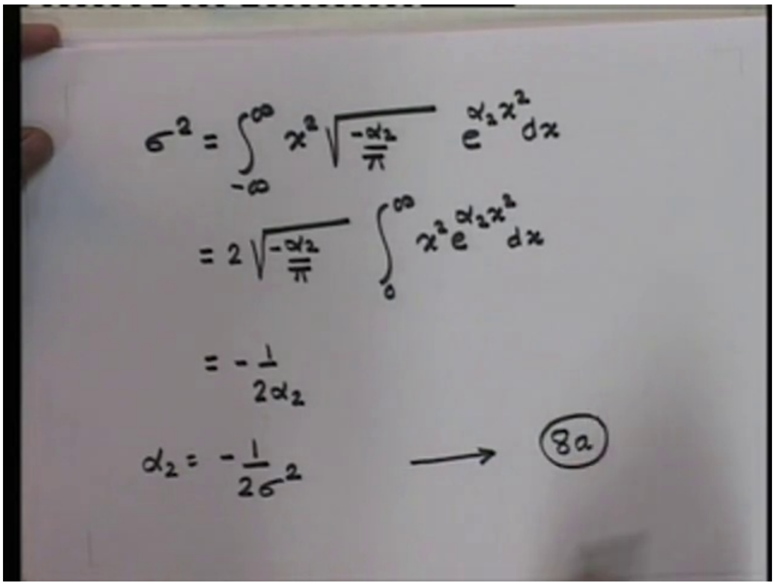
So, in the present case, we have F of p x is equal to P log 1 by p, phi 1 function is equal to p and phi 2 function is equal to x squared p. Hence the solution for p is given by partial derivative of p log 1 by p with respect to, p plus alpha 1 plus alpha 2 times the partial derivative of x squared p equation equated to 0. So, if we solve this we get the condition 1 minus 1 plus log p plus alpha 1 plus alpha 2 x squared is equal to 0. Solving for p from this equation we get, p is equal to e raised to alpha minus 1 e raised to alpha 2 x squared. Let us call this equation number 6, substituting equation number 6 in equation number 2 a.

(Refer Slide time: 10:14)



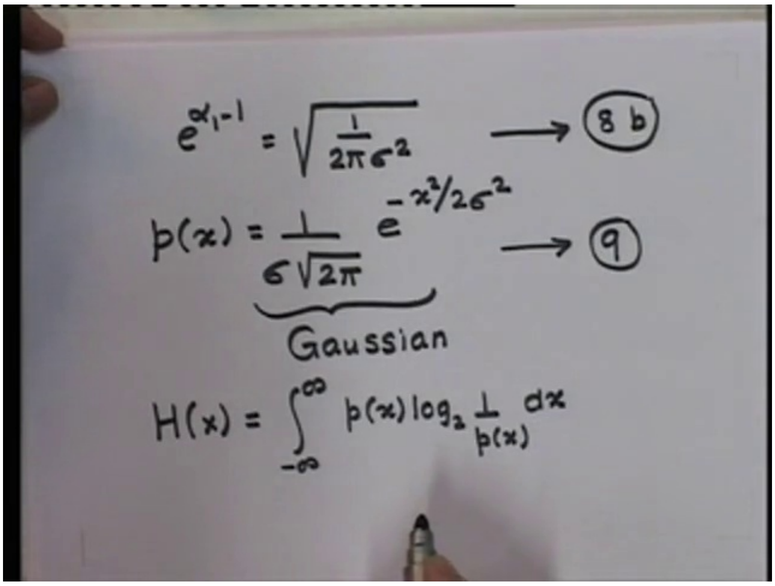
We have 1 is equal to the quantity on the right hand side and this can be simplified as follows, which can be further simplified as 2 into e raised to alpha minus 1 into bracket half root of pi by minus alpha 2 provided alpha 2 is negative. So from this we get e raised to alpha minus 1 is equal to root of minus alpha 2 by pi. This is equation number 7. Now, we substitute the equation number 6 and equation number 7 in equation 2b.

(Refer Slide time: 12:06)



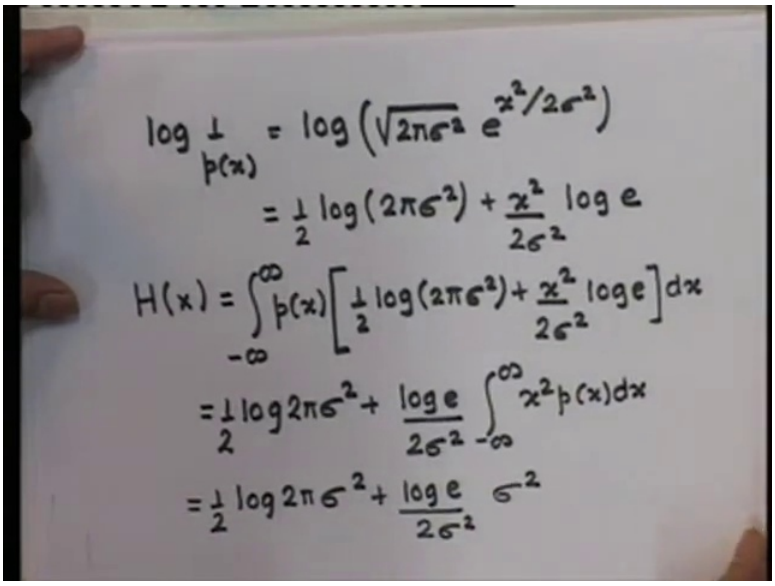
If we do that we get, the result as follows sigma square is equal to integral minus infinity to plus infinity x squared root of minus alpha 2 by pi e raised to alpha 2 x squared dx, which can be simplified further as follows and this is equal to minus 1 by 2 alpha 2 or alpha 2 is equal to minus 1 by 2 sigma squared. Let us call this result number 8 a and e raised to alpha minus 1 is equal to root of 1 by 2 pi sigma squared.

(Refer Slide time: 13:20)



So, if we substitute equation 8a and 8 b into the earlier equation, so if we substitute equation number 8 a and 8 bin equation number 6 we finally, get the result, which we are looking for PDF for the continuous random variable is equal to 1 by sigma root 2 pi exponential minus x squared by 2 stigma squared. This is the final result which you wanted. We therefore, conclude that for a given mean square value the maximum entropy or maximum uncertainty is obtained when the distribution of x is Gaussian. This maximum entropy on uncertainty is given by evaluating the differential entropy that is H x is equal to minus infinity to infinity p x log of 1 by p x dx.

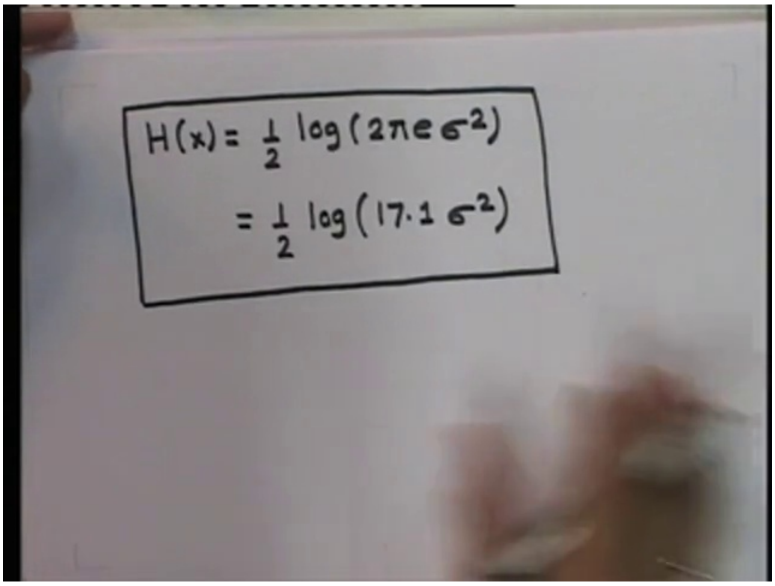
(Refer Slide time: 15:35)



Now, note that log of 1 by p x is equal to log of the quantity in the bracket that is 2 pi sigma squared root multiplied by exponential x squared divided by 2 sigma squared, which is further simplified as half log 2 pi sigma squared plus x squared by 2 sigma squared log e. If we use this relationship and plug it into the equation for the differential entropy then we can evaluate the quantity as follows, p x multiplied by half log 2 pi sigma squared plus x squared by 2 sigma squared log e integrate this quantity.

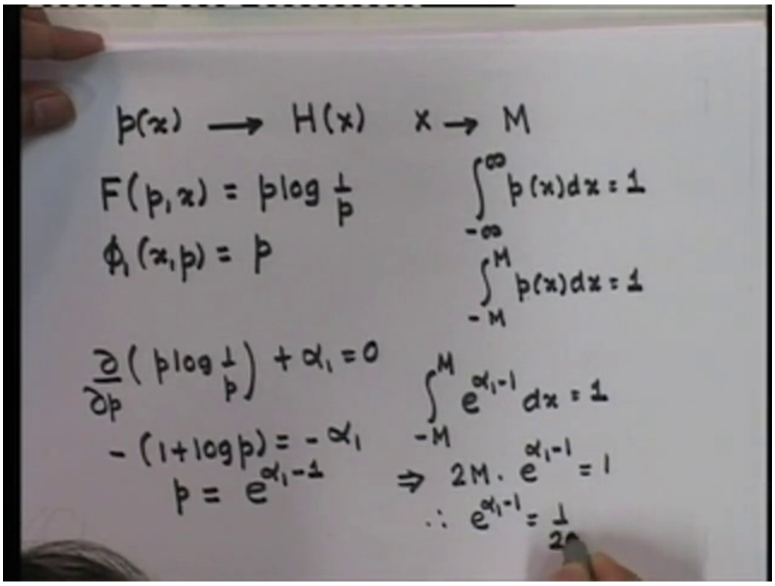
And this can be simplified as half log of 2 pi sigma squared because integration of p x dx minus infinity to plus infinity is equal to 1 plus log e divided by 2 sigma squared, multiplied by integral of minus infinity of x squared p x dx. Now, this can be further simplified as half log 2 pi sigma squared plus log e divided by 2 sigma squared and this quantity out here is sigma squared.

(Refer Slide time: 18:13)



So, the final result which we get as differential entropy the maximum value is equal to half log of 2 pi e sigma squared, or it can also be written as half log of 17.1 sigma squared. This is the maximum differential entropy which we will get when the continuous random variable is Gaussian distributed with, the mean square value equal to a constant sigma squared. Now, let us calculate the PDF for a continuous random variable, which will maximize the differential entropy when there is a bound on the maximum value of the continuous random variable.

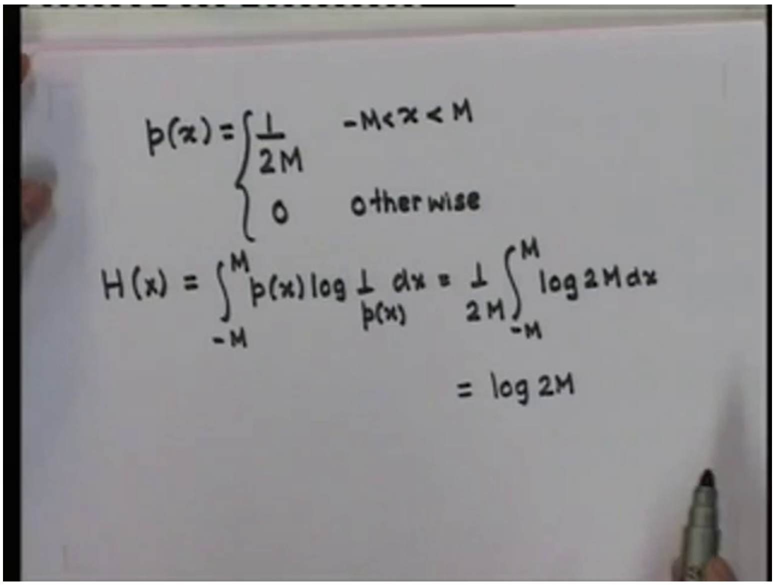
(Refer Slide time: 19:26)



So, the problem is to find out that p x for which the differential entropy is maximum given that x the random variable, the continuous random variable is constraint to some peak value M. So, in this case F of p x is equal to again p log 1 by p phi 1 x p is equal to p because the only constrain we have is integral minus infinity to infinity p x dx is equal to 1 that is, integral of p x dx over minus m to plus m is equal to 1. Therefore, if we use this equation we will get the following results.

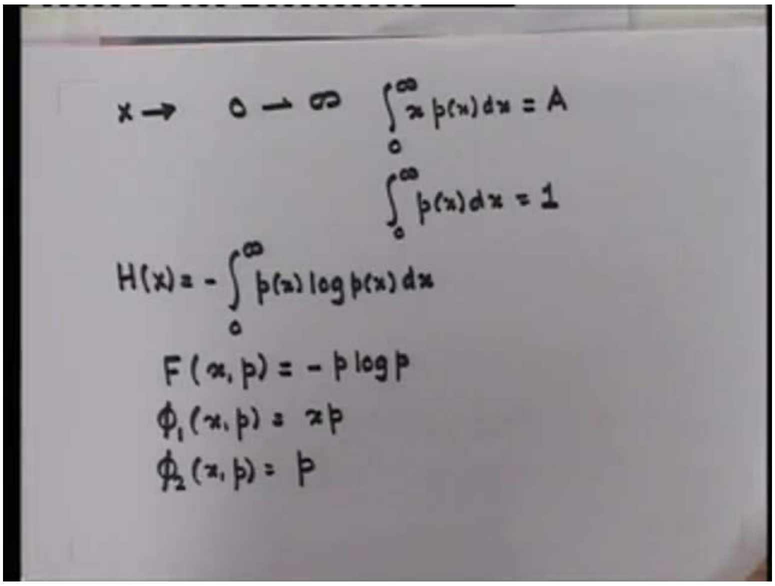
Derivative of p log 1 by p with respect to p plus alpha 1 is equal to 0, which simplifies to minus 1 plus log p is equal to minus alpha 1, which implies p is equal to e raised to alpha 1 minus 1. Now, we know that p x dx over minus m to plus m is equal to 1. So, if we substitute the value of p x in this equation we get the result as follows. Implies 2 M multiplied by e raised to alpha minus 1 is equal to 1.

(Refer Slide time: 22:21)



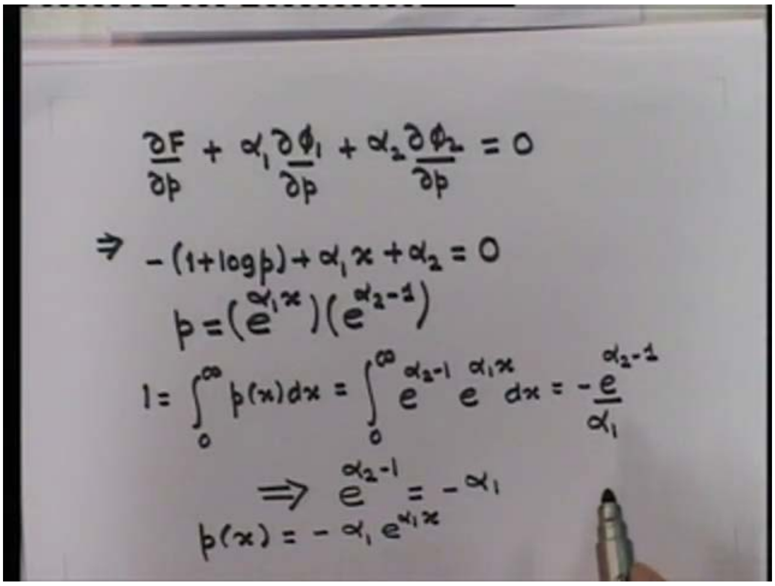
Therefore, e raised to alpha 1 minus 1, minus 1, is equal to 1 by 2 M, which in turn implies that, the probability distributive function which is uniform over minus M to plus M and is equal to 0. Otherwise, will provide the maximum differential entropy and the value of that maximum differential entropy can be evaluated as follows, which results into 1 by 2 M integral minus M to plus M log of 2 M dx, which simplifies to log of 2 M. So for the case, where we have uniform distribution, the differential entropy is maximum and the value is equal to log of 2 M. Before we go ahead let us look into another problem of calculating the PDF for the continuous random variable, that maximizes the differential entropy for a different constraints as follows.

(Refer Slide time: 24:06)



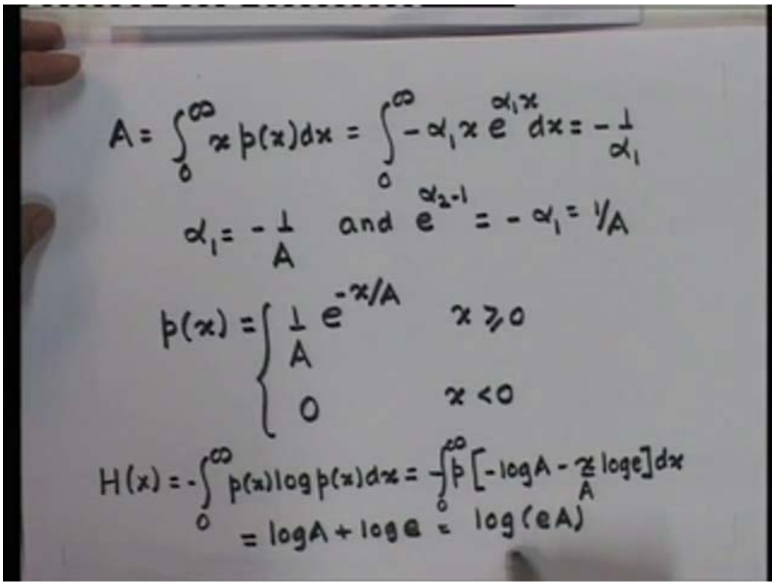
The random variable x is constraint to be positive that is, x lies between 0 and infinity and is also constraint that, the average value of the random variable is given to be equal to the constant A. So, we have the constraint from 0 to infinity x p x dx is equal to A and 0 to infinity p x dx is equal to 1 and we want to maximize the differential entropy given as minus 0 to infinity p x log p x dx. So, in this case F of x p is equal to minus p log p, which was there earlier too phi 1 x p is equal to x p, phi 2 x p is equal to p. Therefore, if we use this equation for determining p, we get the result as follows.

(Refer Slide time: 26:06)



Derivative of F plus alpha 1 d phi d p plus alpha 2 d phi 2 by d p is equal to 0, which implies minus 1 plus log p plus alpha 1 x plus alpha 2 is equal to 0. This can be solved as p is equal to e raised to alpha 1x, multiplied by e raised to alpha 2 minus 1. Now, substituting this relationship in the earlier constraint that is, this we can get a result as follows. 1 is equal to 0 to infinity integral of p x d x which simplifies as minus e raised to alpha 2 minus 1 over alpha 1, which implies that e raised to alpha 2 minus 1 is equal to minus alpha 1, which in turn implies that PDF is equal to minus alpha 1 times e raised to alpha 1 x. We use this, and substitute into this constrain which gives us the result as follows.

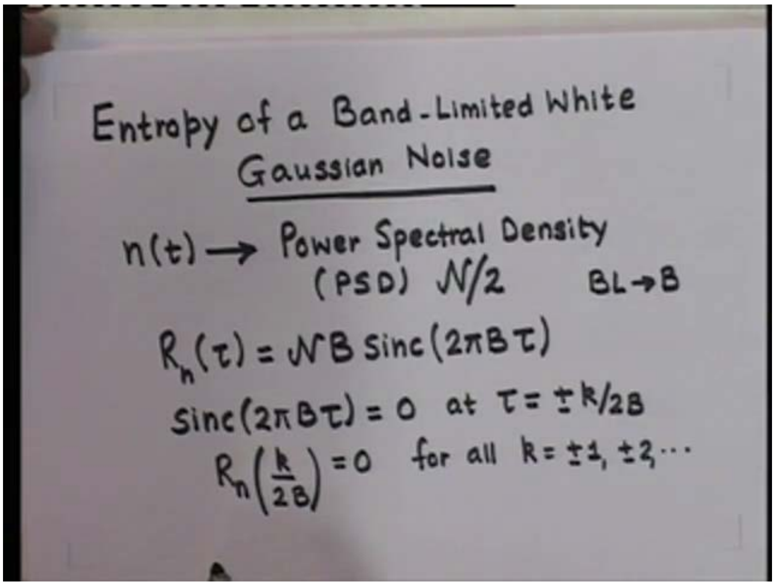
(Refer Slide time: 28:53)



A is equal to integral of x p x over 0 to A is equal to 0 to infinity minus alpha 1 x e raised to alpha 1 x dx, which can be simplified as minus alpha 1, which implies that alpha 1 is equals minus 1 by A and e raised to alpha 2 minus 1 is equal to minus alpha 1, is equal to 1 by A. Therefore, the PDF which will maximize the entropy is given as 1 by A e raised to minus x by A for x greater or equal to 0 and is equal to 0, for x less than 0. Now, for this PDF we can evaluate the differential entropy as follows, is equal to minus p minus log of A minus x by A log e.

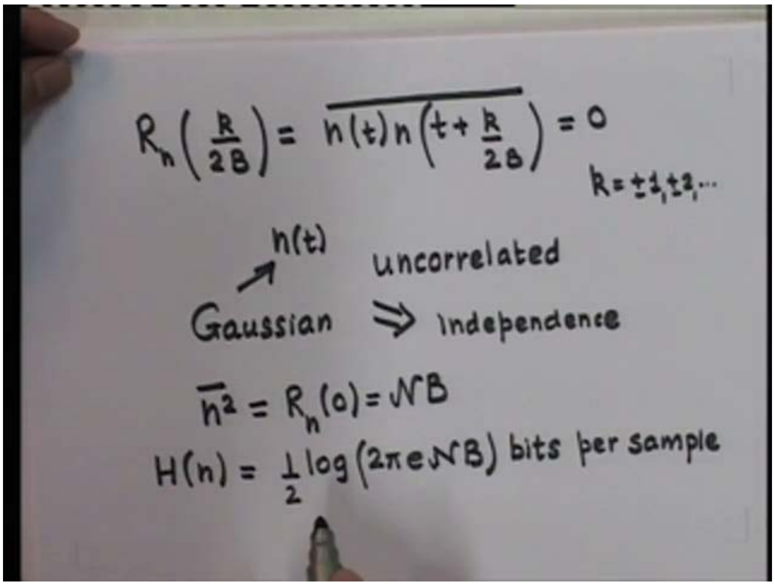
And this quantity can be shown to simplify as log of A plus log of e, which is equal to log of e times A. So, if you have the earlier constrains then the PDF which will maximize the differential entropy is given by this expression and the maximum differential entropy is given by this expression. Now, having defined the differential entropy, let us look at the calculation of entropy of a band limited wide Gaussian noise, which plays an important role in communication systems.

(Refer Slide time: 32:15)



So, let us evaluate entropy of a band limited white Gaussian noise. Let us consider a band limited white Gaussian noise, denoted by n t with power spectral density, that is PSD equal to N by 2. Now, we know that the power spectral density for a random process is the Fourier transform of the auto correlation of the process. So, in this case it implies that the auto correlation function for the noise can be written as N times the band width B multiplied by sine function 2 pi beta tau. This is band limited to B. Now, we know that sine 2 pi beta tau is equal to 0, at tau equal to plus minus k by 2 B where, k is an integer. Therefore, it what it implies that auto correlation function at k by 2 B is equal to 0, for all k equal plus minus 1, plus minus 2 and so on.

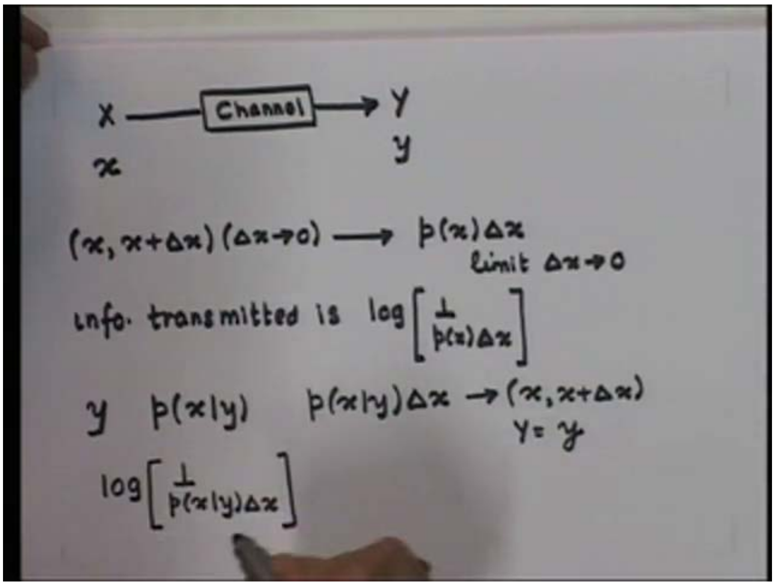
(Refer Slide time: 35:23)



This implies that auto correlation function at k by 2 B, which is equal to by definition expectation of n t and n t plus k by 2 B equal to 0, for k equal to plus minus 1, plus minus 2and so on. Now, because n t and n of t plus k 2 B for k is equal to plus 1 minus, plus minus 1, plus minus 2 nyquist samples of random process n t. It follows that the nyquis samples of n t are all uncorrelated. Now, because n t is Gaussian and it is uncorrelated, it implies that this R also independence.

Hence, all nyquist sample of n t are independent. Now the variance of the noise sample which is equal to auto correlation evaluated at 0 log is equal to N times the band width of noise process, that is B. Hence the variance of each nyquist sample of n t is N B. Now, we also know that the entropy of each nyquist sample of n t is equal tohalflog 2 pi e N B. Because each sample is a Gaussian and we have calculated earlier in the class the differential entropy for the PDF which is Gaussian.

(Refer Slide time: 38:36)



Now, because n t is completely specified by 2 B nyquist sample per second, the entropy per second of n t is, the entropy of 2 B nyquist sample. Now because all the samples are independent that is the knowledge about one sample gives no information about any other sample. Hence the entropy of 2 B nyquist sample is the sum of the entropy of the 2 B samples. And therefore, the entropy of the noise process evaluated per second is equal to B n B times log of 2 pi e N B bits per second.

From the result derived so far, we can draw one significant conclusion among all signals the band limited to B hertz and constrained to have a certain square value sigma squared, the wide Gaussian band limited signal has the largest entropy per second. The reason for this lies in the fact that, for a given mean square value Gaussian samples have the largest entropy moreover all the 2 B samples of a Gaussian band limited process are independent. Hence the entropy per second is the sum of entropies of all the 2 B samples. In processes that are not white, the nyquist sample are correlated, hence the entropy per second is less than the sum of the entropies of the 2 B samples.

Next, if the signal is not Gaussian, then its samples are not Gaussian, hence the entropy per sample is also less than the maximum possible entropy for a given mean square value. So, to reiterate for a class of band limited signals constrained to a certain mean square value, the white Gaussian signal has the largest entropy per second or the largest amount of uncertainty. This is also the reason why white Gaussian noise is the worst possible noise, in terms of interference with signal transmission.

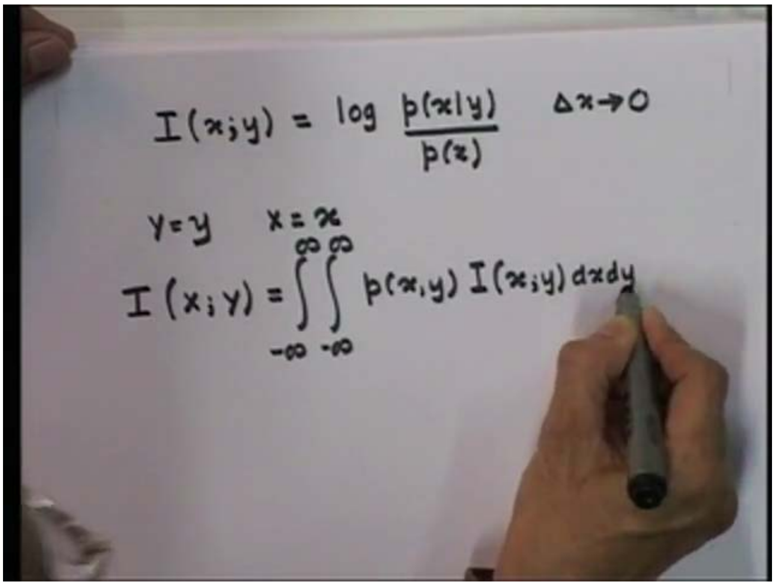
Now, the ultimate test of any concept is its usefulness. So after having defined relative entropy or differential entropy of the continuous random variable. Let us see how does this definition lead us to meaningful results, when we consider the mutual information between continuous random variable x and y. Let us assume that we wish to transmit a random variable x over a channel each value of the random variable x, in a given continuous range is now a message that may be transmitted. For example, as a pulse of height x, the message recovered by the receiver will be another continuous random variable.

Let us denote it by y, now if the channel were noise free the received value that is y of the continuous random variable, would uniquely determine the transmitted value of x of the variable. But because of channel noise there is certain uncertainty about the true value of the random variable x. Now consider event at the transmitter, a value of the random variable x in the interval x, x plus delta x has to be transmitted assuming delta x tends to 0. The probability of this event is given by p x delta x, in the limit delta x tends to 0. Hence, the amount of information transmitted is given by log of 1 by p x delta x.

Now, let the value of random variable y, at the receiver be denoted as y and let p of x given y that is, the conditional probability density of the random variable x, when the random variable y, at the receiver is equal to this small y. Then probability of x given y multiplied by delta x, is the probability that the random variable x will lie in the interval x, x plus delta x. When the random variable at the receiver is equal to y provided delta x tends to 0.

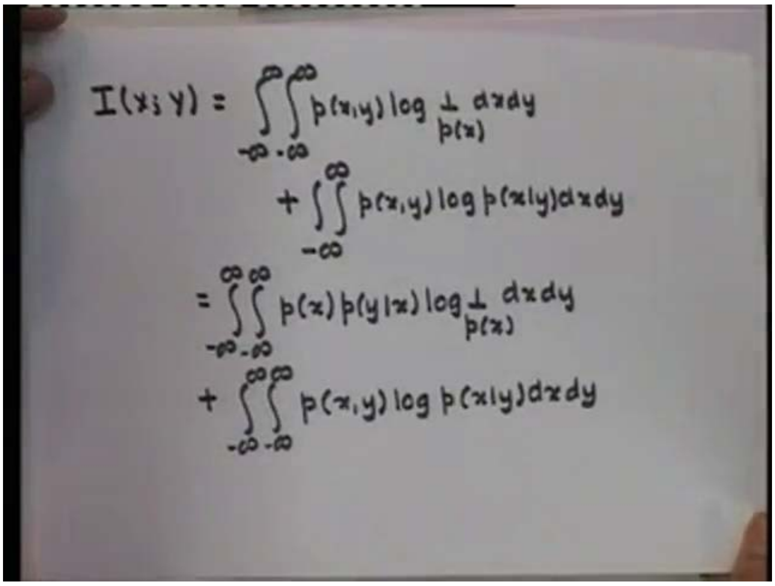
Obviously there is an uncertainty about the event that, the random variable x will lie in the x and x plus delta x. This uncertainty is given by log of 1 by probability of x given y multiplied by delta x. This uncertainty arises because of channel noise and therefore, represents a loss of information because log of 1 by p x delta x is the information transmitted and this is the information lost over the channel.

(Refer Slide time: 46:54)



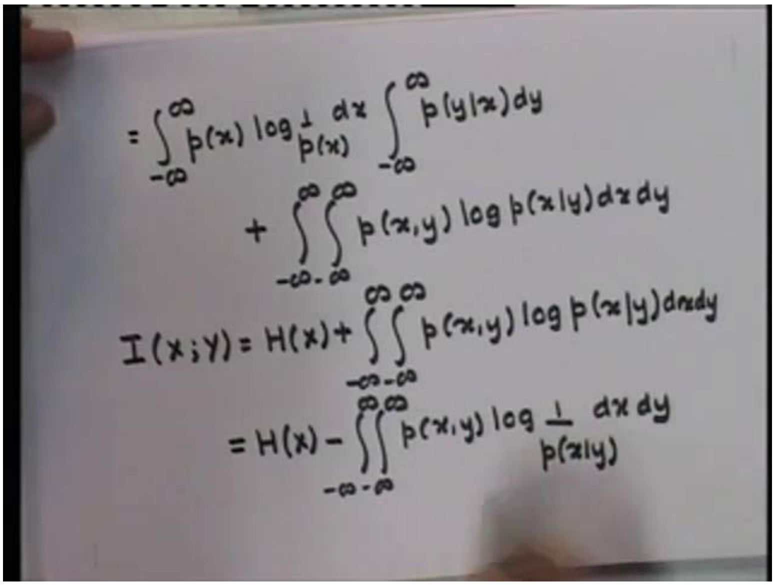
The net information received is denoted by I of xy equal to log of probability of x given y divided by probability of x. Note that this relation is true in the limit as delta x tends to 0.Therefore, I of x y represents the information transmitted over channel when we receive y when, x is transmitted. Now, we are interested in finding the average information transmitted over a channel when, some x is transmitted and a certain y is received. So, we must average I x y over all values of random variable x and y. So in this case the average information transmitted will be denoted by, I x y equals double integral minus infinity to infinity p joint probability of x and y multiplied by I, given by this expression.

(Refer Slide time: 49:00)



Now, we can simplify this expression as follows I of x y is equal to p of x y log of 1 by p x dx dy plus double integration of p x y log of p of x given, y dx dy. This can be further simplified as joint probability of random variable x and y is given as probability of x, PDF of x multiplied by conditional PDF of y given, x log of 1 by p x dx dy plus, joint probability distribution function p x y log of p of x given, y dx dy.

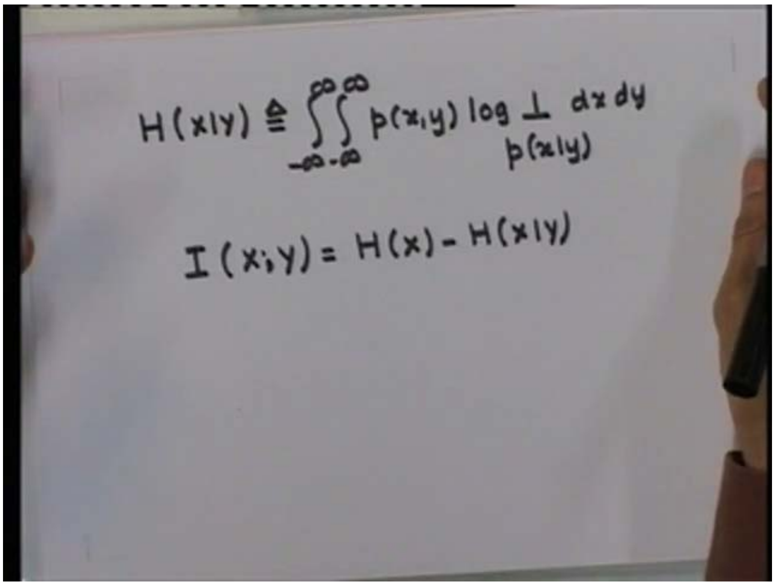
(Refer Slide Time: 50:51)



This can be further simplified as p x log 1 by p x dx multiplied by minus infinity to infinity joint PDF of y given, conditional PDF of y given, x plus joint PDF p x y log of conditional PDF x given, y dx dy. Now, this quantity is equal to 1 and this quantity out here by definition is the differential entropy. Therefore, we can simplify I of x y is equal to differential entropy plus, double integration of joint PDF p xy log conditional PDF of x given, y dx dy.

This can be written as now, the integral on the right hand side is the average over random variable x and y of the quantity log of 1 by p of x given, y. But log of 1 by p of x given, y represents the uncertainty about the random variable x when, the random variable y is received. This, as we have seen is the information lost over the channel so the average of this quantity is the average loss of information over the channel.

(Refer Slide Time: 53:46)



So, thus by definition H of x given y that is the equivocation of the random variable x with respect to y is equal to, double integral of joint PDF of random variable x y multiplied by log of 1 by conditional PDF of x given, y. And using this definitions we can write the mutual information between the random variables x and y is equal to differential entropy minus equivocation of the random variable x, with respect to y. Thus when some value of x is transmitted and some value of y is received, the average information transmitted over the channel is given by this quantity. In the next class, we will define the channel capacity for a continuous channel and derive the channel capacity for an additive wide Gaussian noise case.