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

Lecture – 29
Maximum Entropy of Source with M-ary Alphabets,
Concave/ Convex function, Jensen's Inequality

Hello, welcome to another module in this massive open online course. So, we are looking at entropy, entropy of a source and the various properties of entropy all right. Let us now look at maximum entropy that is when is the entropy maximized.

(Refer Slide Time: 00:28)
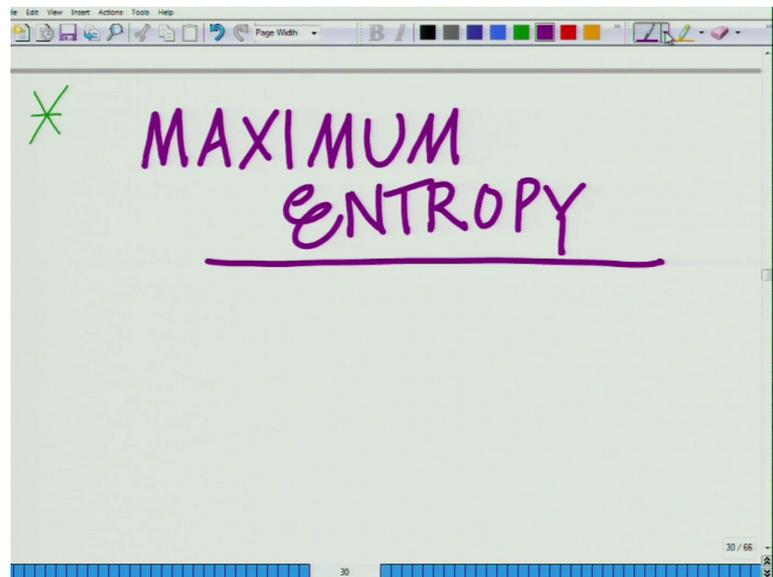
So, let us look at entropy maximization or maximum entropy basically into understand when is entropy maximized.

Now consider a source with alphabet of size 4 S 0, S 1, S 2, S 3 that is alphabet or size 4 that is 4 symbols in the source alphabet. Let the probabilities b P 0 equals 0.5, P 1 equals 0.25 and probability of S 2 equals P 2 equals probability S 3 equals P 3 equals it is a naturally these probability as we going to be 0.25 divided by 2 because the total probability has to add up to one. So, these are going to be 0.125 ok.

And the entropy of this source if you can see is H of s equals 0.5 log to the base 2 1 over by 0.5 plus 0.25 log to the base 2 1 over 0.25 plus 2 into 0.125 because there are 2

symbols of probability 0.125 each. So, this is equal to well half into log to the base 2 is 1 plus 1 by 4 into 2 plus twice 1 by 8 into 3, that is equal to 7 by 4 bits per symbol which is equal to well that is equal to well 1.07 correct 1.75 bits per second ok.

(Refer Slide Time: 03:21)

Now if you look at this and. In fact, if you look at this interestingly I can represent S o, S 1, S 2, S 3 as follow I can represent S o using 0 that is a single bit, I can represent S 1 using 2 bits one 0 I can represent S 2 and S 3 using 3 bits each 1 1 1 and 1 1 0. So, this as 2 bits. So, this is an encoding what we are trying to give here is an encoding of a 0, S 1, S 2, S 3 and you can see that basically we can represent S 0 using 1 bit, S 1 using 2 bits, S 2 and S 3 using 3 bits each and basically you are seen as the probability decreases the number of bits.

So, basically symbols which occur with smaller probability have a larger number of bits to present them that is the strategy. So, that the average number of bits per symbol can be reduced and if you look at this average number of bits per symbol.

Well this is equal to average number of bits per symbol is equal to 1 into 0.5 plus 2 bits occurring with probability 2.5, plus 2 into 2 symbols with 3 bits each occurring with probability 0.125 and you can this is the again 1.75 which is equal to H S. So, you can see there is a fundamental relation between this is not by co incidence we will justify this rigorously later, there is a fundamental relation between the number of bits per symbol required to represent a source only with the source alphabet and certain probability source with certain entropy correct.

So, there is fundamental relation between the entropy of the source and the number of bits the average number of bits per symbol represent required to represent that source required to encode that source. We will explore this I am just hinting at this relation, we will rigorously explore this of course, in the future models ok.

So, this is an important relation between it is sufficient to note at this point, that there is an important relation between entropy and number of bits or let us rather say average number of bits per symbol. Average number of bits per symbol required to encode the sources.

(Refer Slide Time: 06:49)

Now, let us look at another source at another source or rather the same source, but change probability let s a each let us a each symbol has probability has equal probability equal to 1 by 4 then H of s we can see. So, basically all symbols are equiprobable.

(Refer Slide Time: 07:18)

So, H of s we can see in this scenario is 4 into p into log to the base 2 1 over p that is 4. So, this is 1 into log to the base 2 4 this is equal 2. So, you can see H s equal to 2 which is greater than. So, it seems this is greater than the previous entropy; a parentally h s is larger when the symbols are equiprobable.

And previously even for the binary source we had seen that for either probabilities of for a either symbols are probability 0, and other symbols are probability 1 and the entropy is 0, and the entropy is maximized when both the symbols are equal probability that is half all right.

So, this is a conjecture that the entropy is a maximized when all the source symbols all the all the symbols in the source alphabet correct all the alphabets in the source alphabet have equal probability all right. So, let us rigorously justify this conjecture alright when is entropy maximized. For that we have to start by looking at the property of concave functions all right to understand this we have to start by looking at what are known as concave functions. So, I will start with the definition of a concave function to understand when is the entropy maximized.

(Refer Slide Time: 08:42)
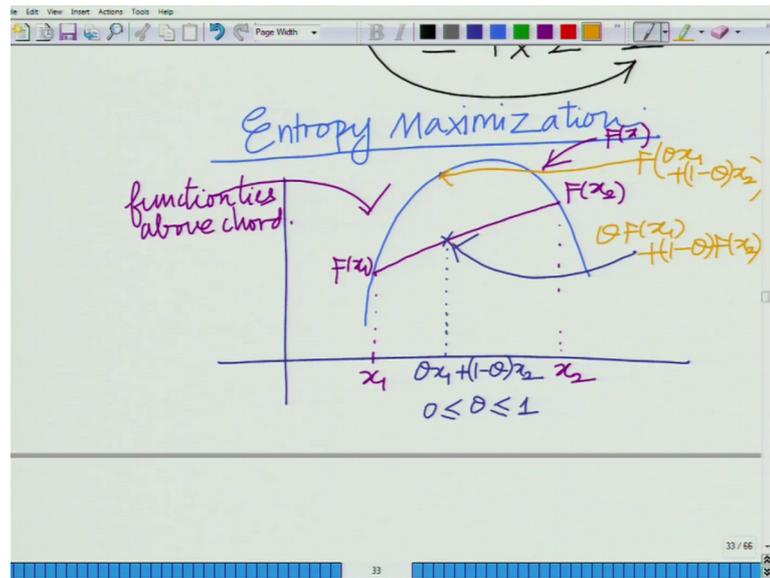
Let us start by defining a concave function; a concave function is simply a function which looks like an inverted bawl. So, basically if I draw a cord between 2 points x 1 and x 2 ok.

This is your F of x. So, this is F of x 1, this is F of x 2, you can see the function lies above the cord that is if you take a point which is here let say which can be represented as follows. This point to corresponds to lets a waiting by theta x 1 plus 1 minus theta x 2, where 0 is less than equal to theta less than equal to 1 and this is the value of the function at that point F of theta x 1 plus 1 minus theta x 2. And this is the value of the cord all right at the point and if at look at the function at the point, I am sorry this rather this is your F of theta x 1 plus 1 minus theta x 2 and this is basically theta times F of x 1 plus 1 minus theta times F of x 2.

And for a concave function this is the function always it looks like an inverted bowl. So, the function always lies above the cord therefore, it must be true that for any x 1.
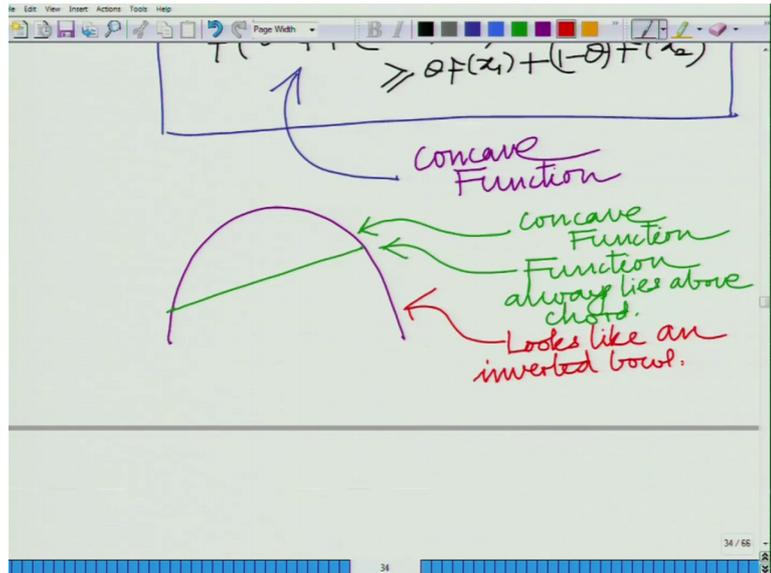
(Refer Slide Time: 11:17)

Any pair of points x 1 comma x 2, 0 less than equal to theta less than equal to 1 we must have F of theta x 1 plus 1 minus theta x 2 is greater than theta (Refer Time: 11:44) times x 1, F x 1 plus 1 minus theta times F x 2 ok.

Such a function which satisfies this property for any x 1 x 2 any pair of points x 1 x 2 and any theta which is 0 less than equal to theta less than equal to 1, such a function is term as a concave function there is a convex function this is a concave function such a function is termed as a this function is termed as a concave function, that is basically it looks like as an you can see a concave function looks like an inverted bowl that is the function always lies above the, this is a concave function ok.

 (Refer Slide Time: 12:35)

The function always lies above the cord. So, it is look like an inverted bowl, on the other hand if you have a function that actually looks like a bowl.

(Refer Slide Time: 13:29)
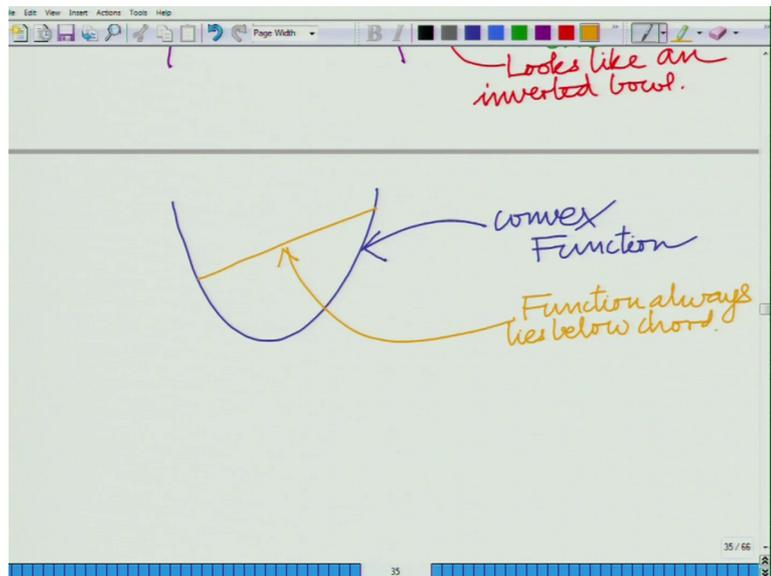
Although it is not relevant to a discussion this function is a convex function. The inequality is reversed for a convex function it looks like a bowl not like a inverted bowl. So, the function always lies below the cord for a and now, if you can see the log function.

(Refer Slide Time: 14:05)
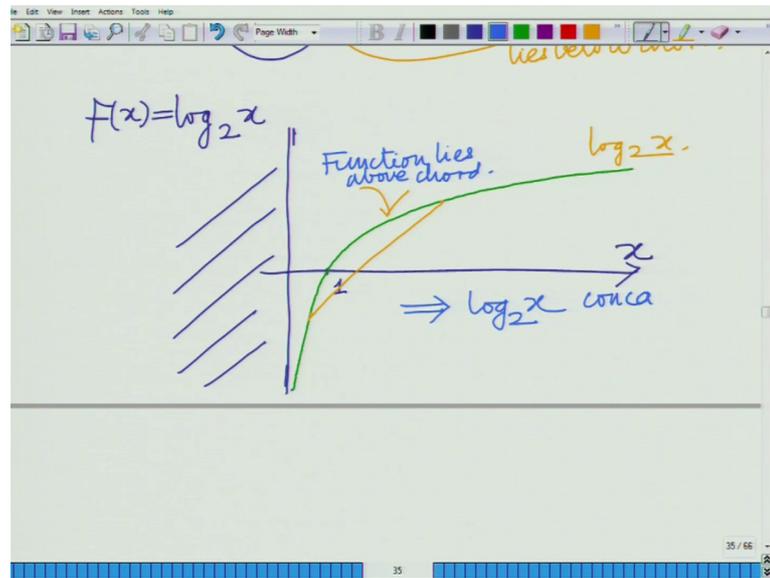
You will observe it is easy to see that the log of course, the log is only defined for x greater than 0, such not define for x less than 0 and for x equal to 0 it tends to minus infinity as x tends to 0, at x equal to one the log is 0 and then at x greater than 1 its greater than 0 and rises slowly ok.

And now if you can see the log function if I draw a cord. So, this is your log to 2 to the base x, log x to the base 2 and the function lies above the cord implies log x to the base 2 is concave. And this can also be seen as follows if you take this know as the derivative text because the log function is differentiable if a differential twice examine the second derivative, the second derivative is always less than or equal to 0 then the function is concave. So, let us differentiate this twice.

F prime of x is 1 over log to the base 2, 1 over x and differentiating log to the base 2 if double prime of x is 1 over log 2 times minus 1 over x square.

So, F prime of x remember x is its log is define only for x greater than equal to 0, F prime of x less than equal to 0 implies F of x is or F of x rather this is F of x equal to log of x is concave. This is term as the derivative test or the second derivative test and now what does this mean for us, well if you look at this now look at this.

Inequality F of theta times h 1 plus 1 minus theta times x 2 is greater than theta times F x 1, plus 1 minus theta times F x 2 I can always represent I can always replace it has theta equals P 1, 1 minus theta equals P 2, where P 1 plus P 2 equal to 1 implies P 1 comma P 2 are basically probabilities is not it.

So, we will have F of P 1 x 1 and P 2 x 2 greater than equal to P 1, F of x 1 plus P 2, F of x 2 all right where P 1 and P 2 are probabilities this P 1 and P 2 are. So, I can always represent theta and 1 minus theta since theta lies between 0 and 1, I can always represent theta by P 1 and one minus theta by P 2 because theta and 1 minus theta are both positive and they sum to 1.

So, I can always think of these as probabilities P 1 and P 2 greater than equal to 0 sum into 1. So, I have F of P 1 x 1 plus P 2 x 2 is always greater than equal to F of the P 1 times F of x 1 plus P 2 times x 2 for a concave function. Remember this is for a concave function ok.

(Refer Slide Time: 18:23)
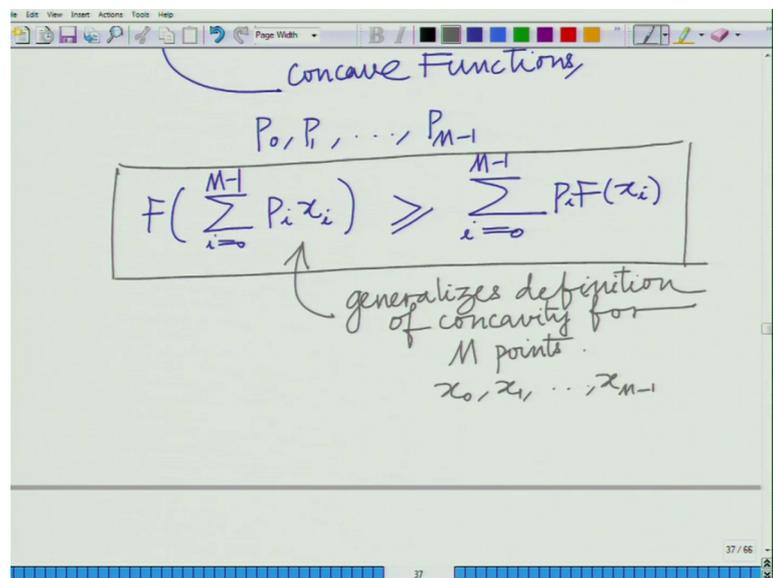
We are considering concave functions and now I can extended naturally no need to be limited by 2 probabilities I can consider for rather in the context of our work I can consider P 0, P 1, P m minus 1 and I can say F of summation I equal to 0 to m minus 1 P i, correct pi x i greater than equal to summation i equal to 0 m minus 1, but if you look at this ok.

So, this is a generalization of the definition of concavity for n point or m points rather and further if you look at this correct if you exam in this further, you can see that summation i equal to 0 m minus 1, P i x i is nothing, but expected value of x, where x is a random variable correct which takes the value x i with probability P i and further P i F I is nothing, but expected value of F of x. This is P I that is waiting each function at each point F of x i with a corresponding probability P i that is expected value of F x. So, what we have if you observe this. So, this is expected value F of expected value of x.

So, what we have for a concave function is if you replace look at it terms of expatiations we have F of expected value of x is greater than or equal to expected value of F of x, this is basically termed as very important property this is basically termed as Jensen's inequality for a concave function, which is naturally also valid for a log function remember we are interest in the log because entropy is built on the log function ok.

So, this is the very important property this is the Jensen's inequality for a concave function naturally for a convex function in the equalities then equality is reverse that is expected value of F of x would be greater than equal to F of expected value of x. So, the inequalities reversed for a convex function let us note that also because that might come handy.

(Refer Slide Time: 22:20)
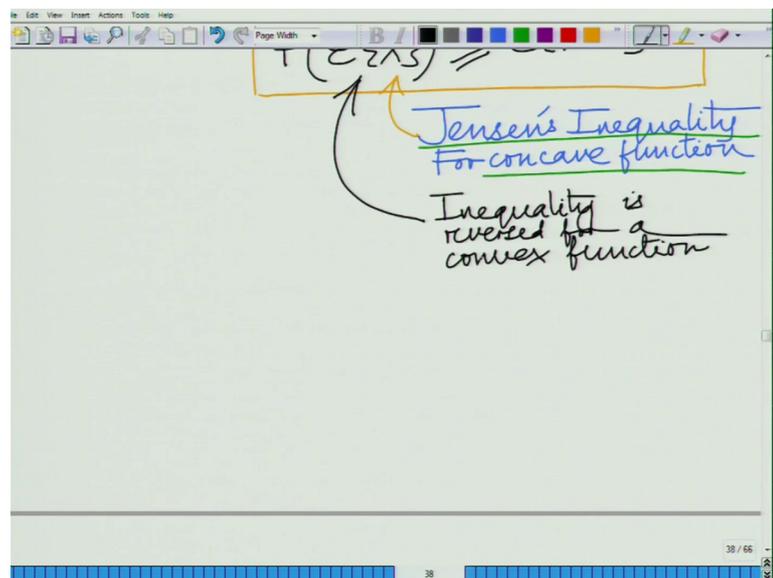
 The inequalities reversed for a convex function.

Now, coming to our problem if you look at entropy correct.

(Refer Slide Time: 22:47)

Remember the entropy H of s is summation I equal to 0 to m minus i, P i log to the base 2 1 over P i this is concave. So, this is P i I can treat 1 over P i as x i summation P i F of x i is less than or equal to F of less than or equal to F of that is log to the base 2 summation i equal to 0 to m minus 1, P i x i which is 1 over P i correct which is equal to log to the base 2.

(Refer Slide Time: 23:47)

Summation i equal to 0 to m minus 1 of one which is equal to log to the base 2 of m. So, H of s is always less than or equal to log to the base 2 of m and this is the maximum

entropy. So, the maximum entropy is log to the base 2 m where what is the m? M is the number of symbols in the source alphabet.

(Refer Slide Time: 24:35)

Ok and we approved this using Jensen's inequality and you can see that this maximum occurs, when the probability of each s of i equals 1 over m because under that scenario the entropy would be 1 over m log 2 to the base 1 over1 over m into m type because each symbol has the same per symbol of sapper symbol entropy. So, m times log m to the base 2 which is equal to log m to the base. So, occurs when probability of each s I is 1 over m that is one all symbols are occurs will all symbols are entropy is maximized when all symbols are equi probable I think that is.

(Refer Slide Time: 26:01)

 This is an important property of entropy let is the probability of each symbol P 0 equals p 1. So, on up to p m minus 1 equal to 1 over that is all the symbols are equiprobable ok.

So, whenever you have a source the uncertainty of the source right the information we have said the entropy is also characteristics also characterizes are also are also countifies the uncertainty of the source, the uncertainty of the source is maximum when all the symbols are equiprobable. If the probability distribution is secured that is sum of the symbols have higher probabilities in comparison to other symbols then entropy is lower than the maximal probable entropy, maximal possible entropy all right. So, the maximum uncertainty the maximum information per symbol or maximum information per alphabet occurs in the source when all the symbols of the source occur when the equal probability as we have seen in the binary source case that is an extreme example when we have 2 symbols correct, one occurs with probability 0 other occurs with probability naturally; that means, the other must occur with probability one because it is a binary source.

So, there is no uncertainty left correct there is no uncertainty left, which means that the entropy of the source is 0. So, maximum entropy occurs when both the symbols are equiprobable that is probability equal 2 half.

So, similarly this result generalizes this to source with any general that is m symbols in the source comprising of m symbols in the source alphabet, and we says that the entropy

is maximized when all the symbols have equal probability all right. So, we will stop here and look at other aspects (Refer Time: 28:14).

Thank you very much.