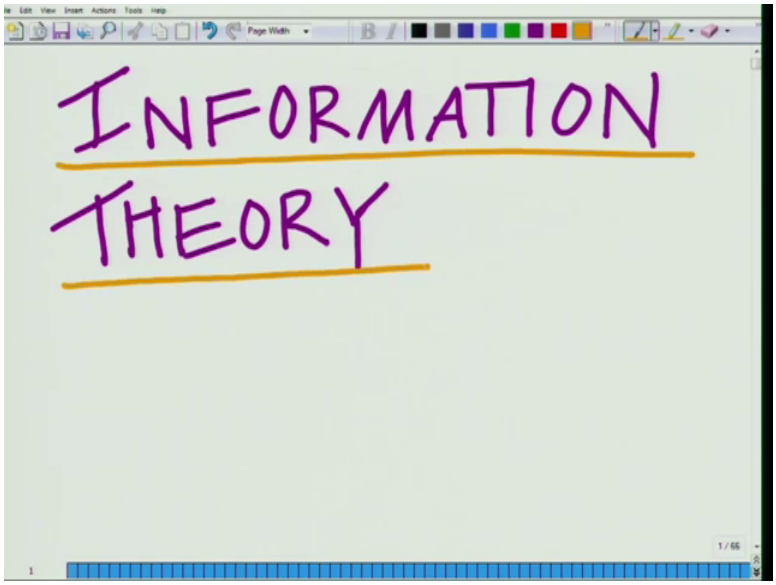
Principles of Communication Systems – Part II

Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 26
Introduction of Information Theory, Relevance of Information Theory, Characterization of Information

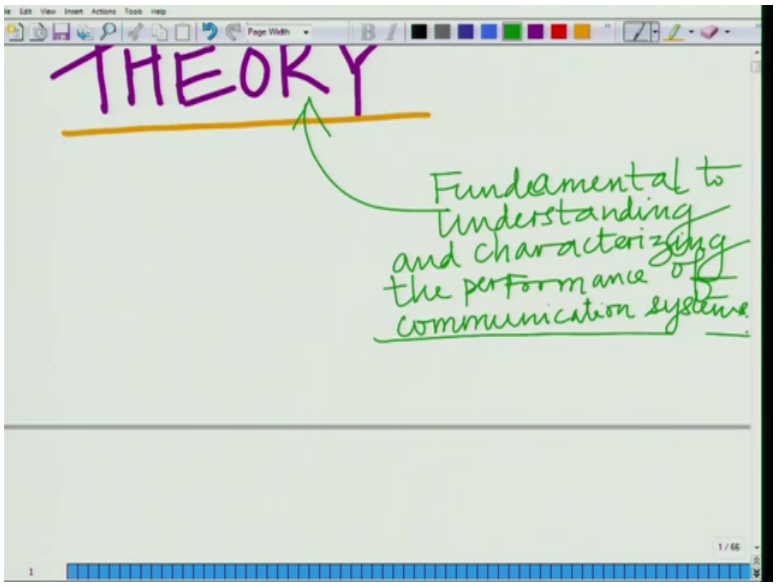
Hello, welcome to another module in this massive open online course and in this module we will change tracks and start looking at an extremely different topic and an extremely important topic in understanding the performance and behavior of communication systems which is termed as information theory.

(Refer Slide Time: 00:35)



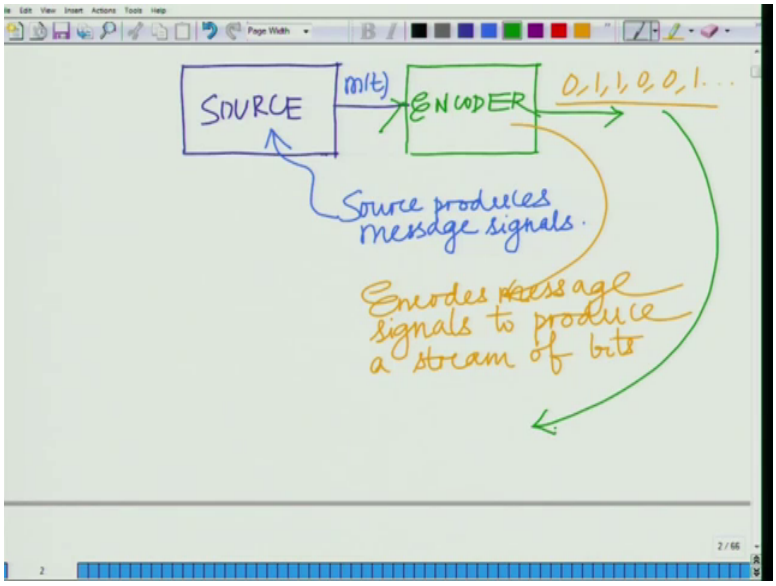
So, as I have already eluded this is going to be very fundamental and its one of the fundamental milestones in understanding the behavior and then also designing better and more efficient communication system. So, this is termed as information theory. Many of you must have heard of this at several, in several different settings, correct.

(Refer Slide Time: 01:18)



So, information theory and as I have said, information theory is fundamental it is something that is very fundamental to understanding and characterizing the performance of. And in fact, information theory is not only although originally intended to study communication systems it has evolved it to encompass a variety of applications such as economics, probability, stock markets, investing and so on. So, are applications of information theory or a vast and ubiquitous all right; however, in this course as part of this course we will restrict ourselves to the fundamental concepts in information theory and the relevant applications in the context of communication systems.

(Refer Slide Time: 02:49)

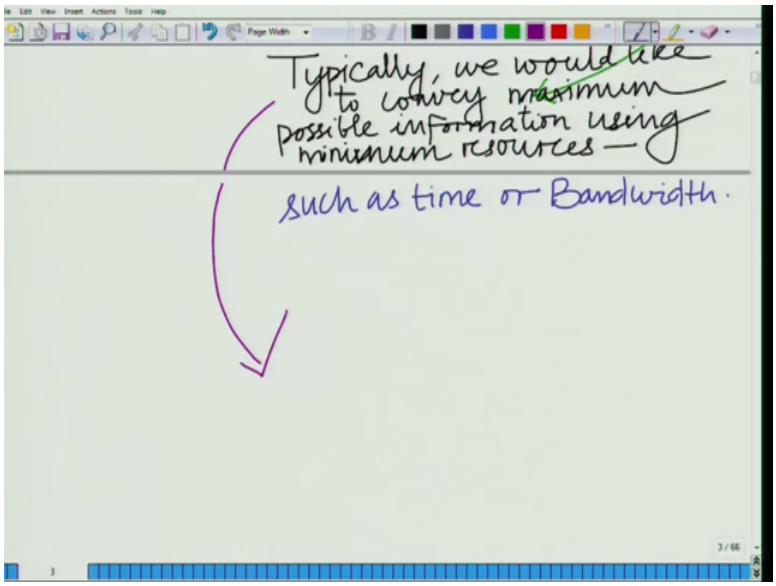


So, we are going to change tracks as I said and start studying information theory and basically to describe information theory or the relevance of information let us go back again look at the structure, the fundamental structure of a digital communication system which we have looked at in one of the very first lecture. So, we have in every digital communication system we have a source and the source produces messages which are led to an encoder correct. So, I have a source, I have an encoder the source produces message signals. So, the source produces message signals. So, the source produces message signal and the encoder encodes them into a stream of bits as we have seen before, encodes messages to (Refer Time: 04:00) encoder. Encodes message signals encodes message signal produce a stream of bits alright. So, we have shown here a typical stream of such bits.

Now the question is obviously, these bits are going to be transmitted or channel let us look at it let us take a simple example. So, the bits are going to be transmitted over the channel naturally that requires bandwidth right that requires usage of the channel alright, and technically one would like to minimize the usage of the channel. So, for a given amount of, so for a given amount of resources for transmission let us say in terms of the channel bandwidth or the time one would like to transmit well the maximum amount of information or what would I (Refer Time: 05:03) remember communication is all about conveying information over a vast distance right from point a to point b or even in a broadcast guide from point a to several points.

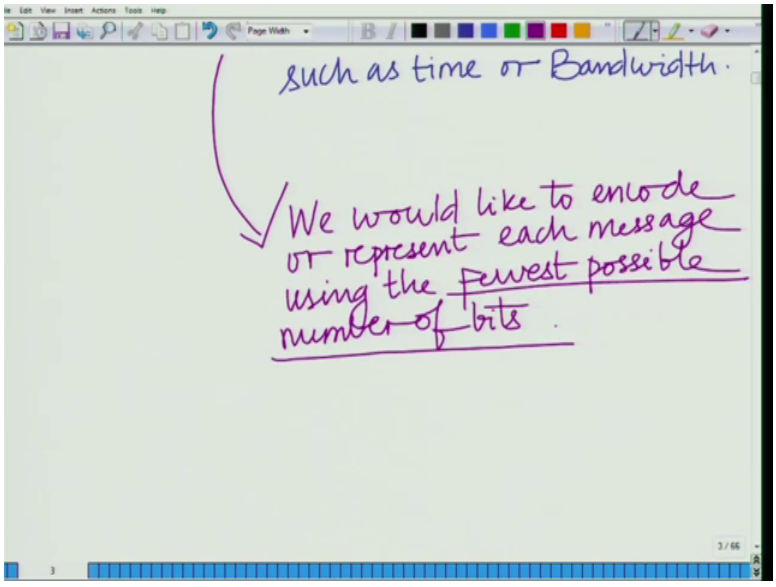
So, typically we are interested in conveying the maximum amount of information using the minimum number of resources or to put it in other way we would like to have a minimum number of bits which signifies a channel uses the number of times we have to use the channel. So, we would like to have a minimum number of bits in comparative or encapsulating a (Refer Time: 05:35) encapsulating the maximum amount of information.

(Refer Slide Time: 05:56)



So, that is the motivation. So, typically we would like. So, how many bits? So, typically we would like convey the maximum possible information using minimum resources such as time or bandwidth purposefully being a bit abstract which implies we would like to encode the messages which implies.

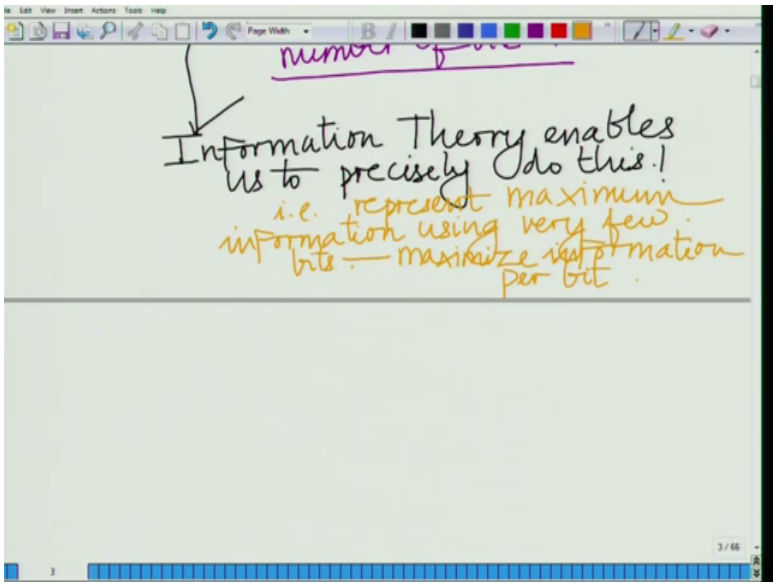
(Refer Slide Time: 07:02)



We would like to encode or represent each message using the fewest possible number, using the fewest possible number of bits yet in other words we would like to compress a lot of information into a very few number of bits. And information theory is nothing but

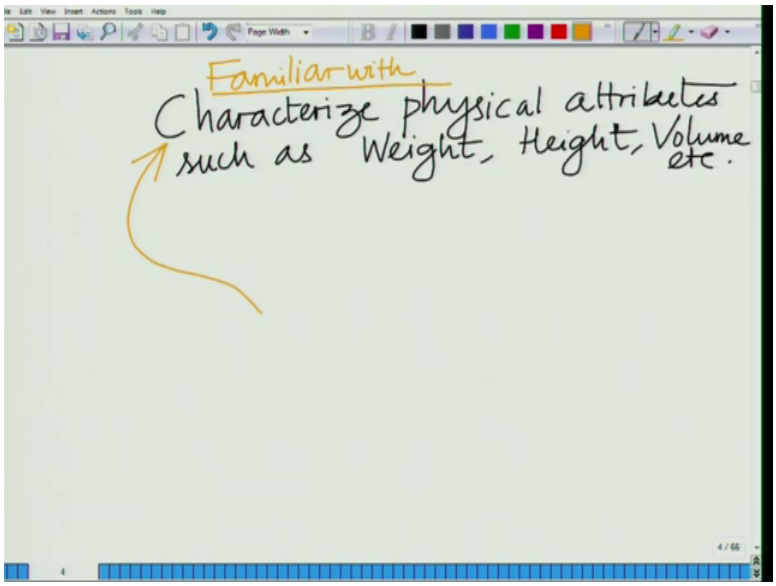
a systematic study to enable us to do this, enable us to compress or enable us to represent a large amount of information in very few bits or in very few which there by utilize very few channel resources they by using conveying information efficiently using as few resources as possible.

(Refer Slide Time: 08:18)



So, essentially, so for that, so this is what information theory enables us to information theory enables us to precisely do this, precisely do this that is represent maximum information using very few. Or basically maximize information per bit you can also say that maximize information, maximize the information content on a bit by basis, maximize the information content per bit.

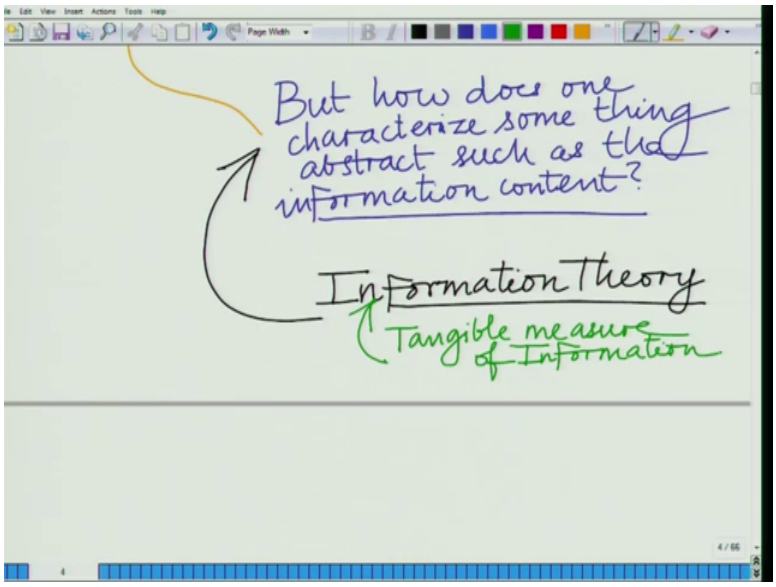
(Refer Slide Time: 09:45)



And now we all know we are very familiar with how to characterize aspects, physical aspects such as or let us say characterize that we will be able to explicit characterize physical attributes such as you know the weight of an object, the height for resistance, the volume etcetera several other aspects.

But how does one characterize something abstract as information. So, these we are familiar with.

(Refer Slide Time: 10:51)

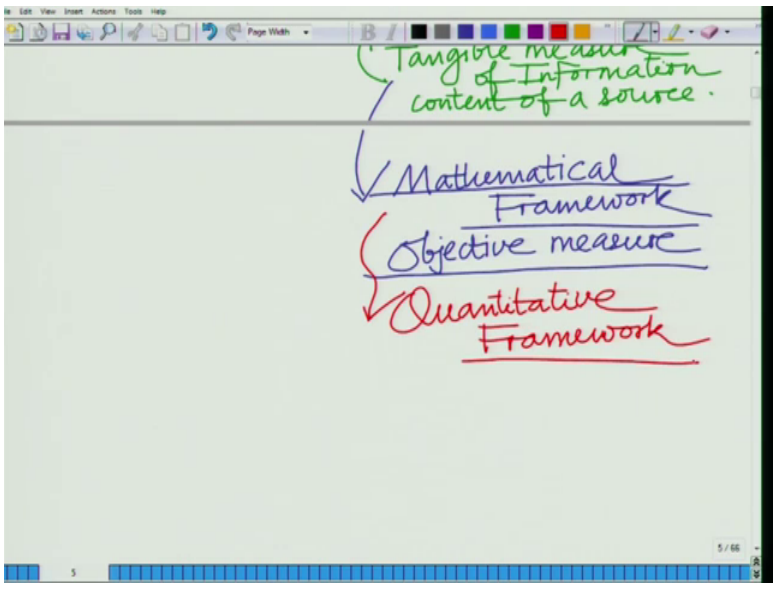


So, we are familiar with how to characterize, but how does one characterize something abstract such as information such as the information content. For instance every day we are exposed to several sources of information right, from personal communication, from watching for instance resources such as the TV. So, such as broadcast media such as the TV or newspaper or several other sources from our day to day communication or what we or through hear say or so on. So, through a variety of modes we access a lot of information, but we never really stop to think about how does one really capture the information content.

For instance when you see a piece of news we intuitively feel that some news right there is more information that other kinds of news or some kind of communication has more information or when a particular person is speaking he typically conveys more information than someone else is speaking. So, how do you characterize something that you cannot feel, somebody you characterize something that is as abstract as information which is tremendously useful, and information theory does nothing, but provide us with a means to the tangible means to characterize to provide us a tangible measure of this otherwise abstract notion of information. So, that is what information precisely does.

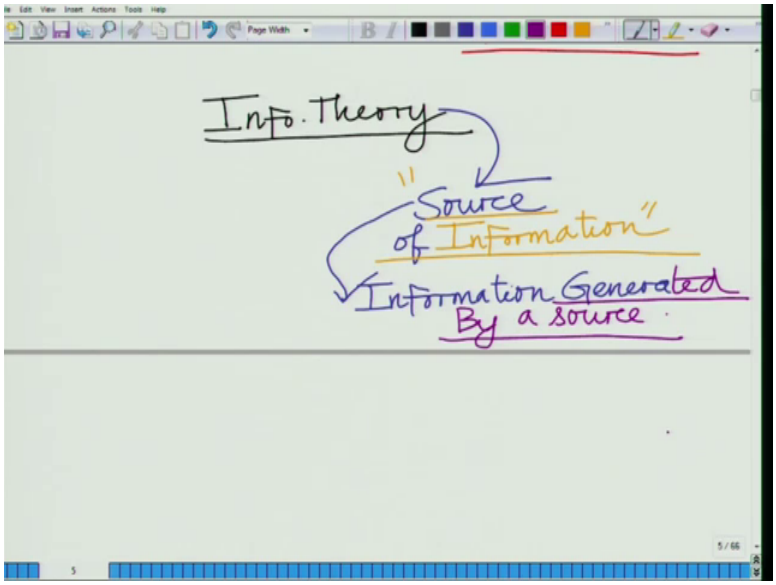
Information theory is nothing, but provides tangible measure, so that you can precisely characterize what is the information of one source versus the other.

(Refer Slide Time: 13:28)



So, naturally the information content of a source. So, naturally and this provide us a mathematical framework. So, information theory is a mathematical framework we are talking about not a qualitative, but a quantitative framework rather. We know qualitatively what constitutes more information. It is an objective measure or a mathematical framework or you can also say as an objective measure removed from personal perception that is the important aspect. Because different people might feel that the information content is different and more importantly this is mathematical framework or a numerical framework or it is a quantitative framework. Not just simply a qualitative framework, but a quantitative framework.

(Refer Slide Time: 14:37)

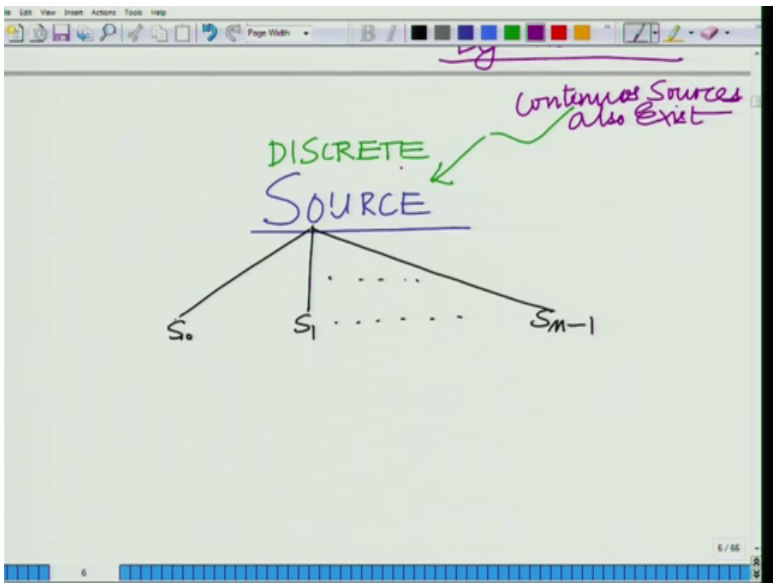


At central to the notion of information theory if you look at information theory central to the notion of information theory is the concept of a source, a source of information. So, think of information as being generated as by source this source can be several things for instance can be people who generate the information or things such as a next higher layer such as a source, such as for instance the internet or so, such as for instance the TV from where you find information right which basically conveys information in some sense to you, alright.

So, there are several sources we generate the information for instance, when you are talking to your mobile phone it is a human being who was generating the information which is later encoded which is later the voice which is digitally sampled and encoded

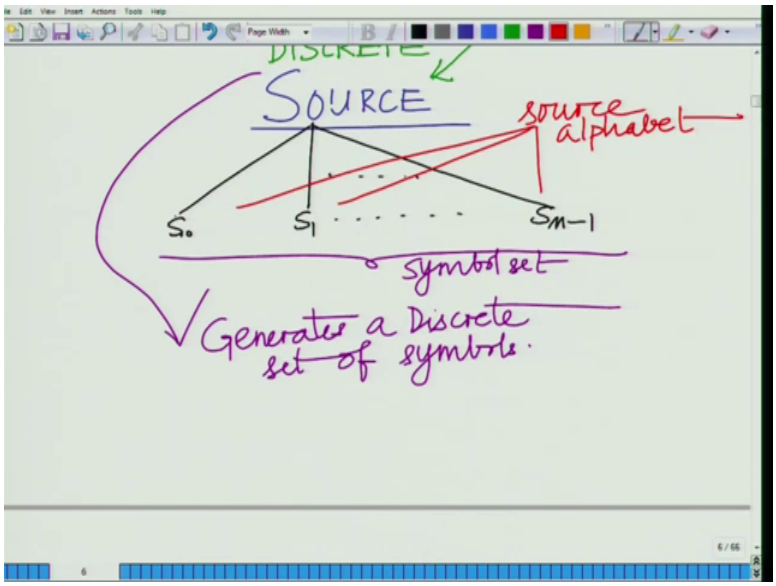
and converted into bits. So, there are several sources of information. And now information theory as we have said is a mathematical framework to characterize the information generated average what is the information of a source. So, we also talked about what is the information it can be a physical, it can be a living of (Refer Time: 16:14) information or information generated what is the information generated we talk about this, information generated by a source what is the information generated by a source. So, we would like to characterize the information generated by a source.

(Refer Slide Time: 16:41)



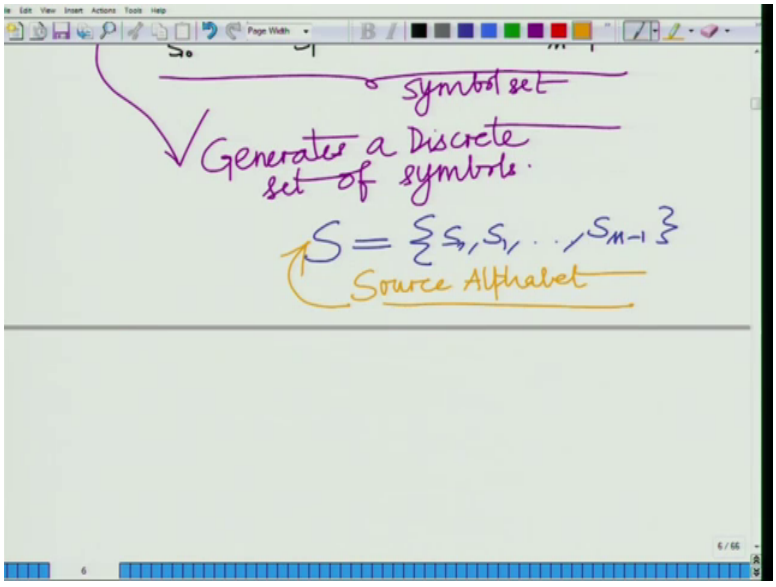
So, let us typically consider a source as we have said a source generates messages. So, the source let us say generates source let us say generates symbols S 0, S 1, so on up to S m minus 1. In fact, there is a discrete set of symbols. So, you can also call this as a discrete source they can also be continuous source, continuous sources can also exist which we will deal with later.

(Refer Slide Time: 17:40)



So, discrete source implies generates the set of discrete symbols, generates a set of discrete set of symbols. The way generates a discrete set of symbols for instance this S 0, S 1 up to S m minus 1 this is the symbol set, this is the and this is also termed as the source alphabet this set is also termed as the source alphabet.

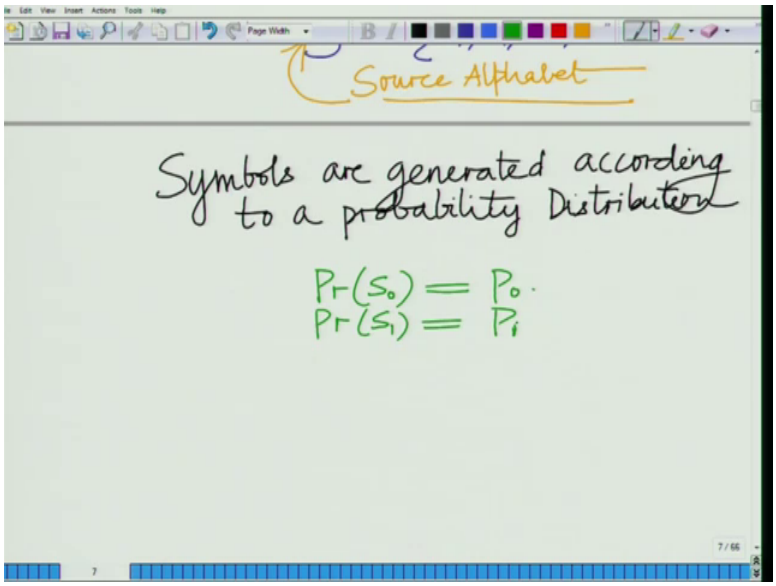
(Refer Slide Time: 18:35)



So, we have a set S consisting of symbols S 0, S 1, S m minus 1 generated by the source this is also termed as the, this is also termed as the, also termed as the source alphabet. So, this is also termed as the source alphabet and further the source is generating this

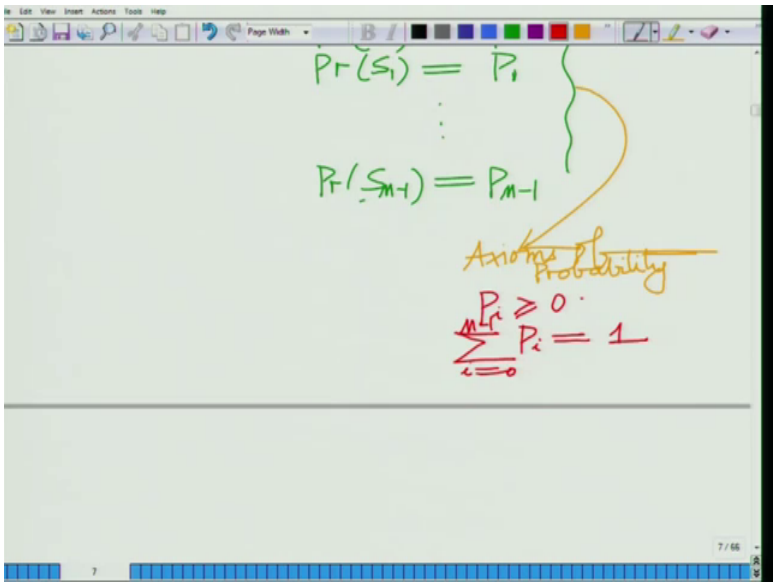
alphabets this various alphabets at various time instants according to a certain probability distribution right according to certain probability. So, it is a continuous source it generates continuous it takes values from a continuous range with a probability density function.

(Refer Slide Time: 19:36)



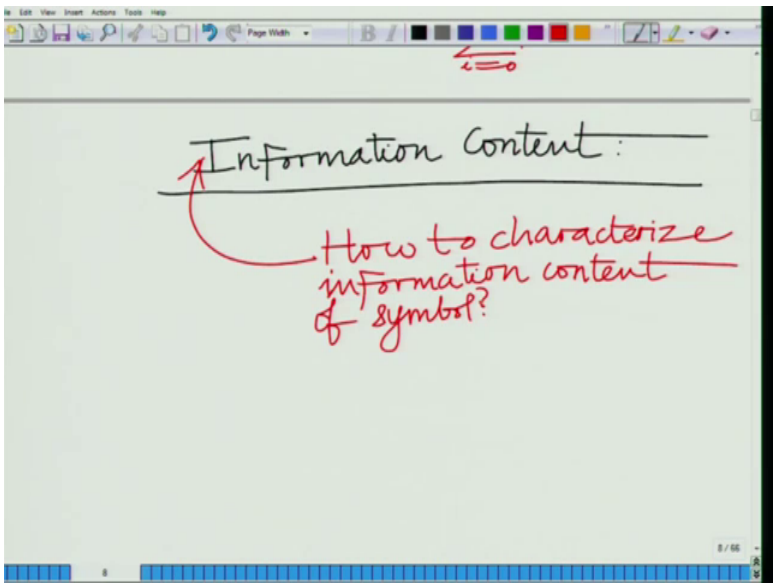
Now, we have a discrete source which is generating discrete symbols according to a set of probabilities. So, the symbols this is the system model or this the model of our source which generated according to a symbols are generated according to a probability distribution we have probability of the symbol S naught equals P naught, probability of the symbol S 1 equals P 1 so on and so forth.

(Refer Slide Time: 20:31)



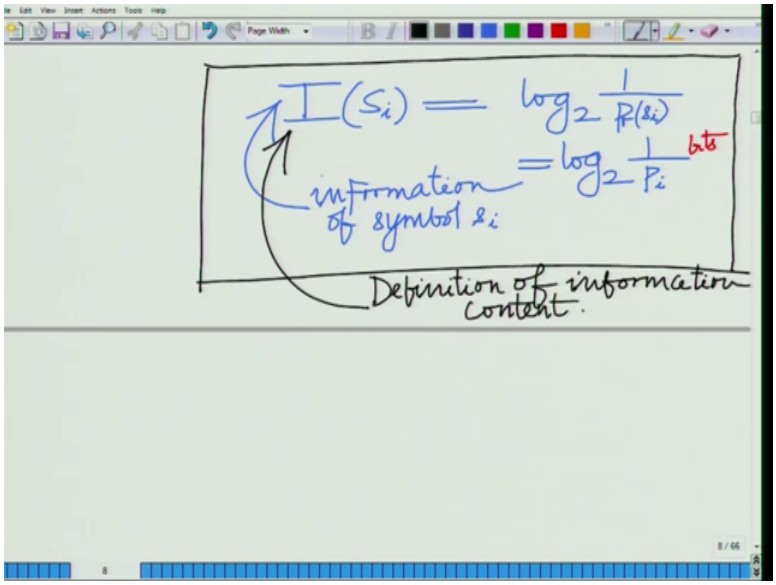
Probability of the symbol S m minus 1 equals P m minus 1 and naturally it goes without saying that from the axioms of probability; obviously, this satisfies the axioms of probability from the axioms of probability. Again I urge everyone to take revise their fundamental the knowledge or revise the concepts of probability and random variables we must have each P i greater than or equal to 0, every each probability has to be greater or equal to 0 and the summation of all the probabilities belonging to the source alphabet has to add up to 1. So, all the probabilities have to add up to 1.

(Refer Slide Time: 21:26)



Now, let us come to the concept of the information content of a symbol S i the (Refer Time: 21:31) this come through this concept of how to characterize the information content of a symbol how to characterize information content of a symbol or basically how to characterize the information content of a symbol and for that we are going to define the information of symbol S i, I S i this is basically your information of symbol S i; as log to the base 2, 1 over the probability S i equals log to the base 2, 1 over P i; 1 over P i and there is a unit for information the unit is bits.

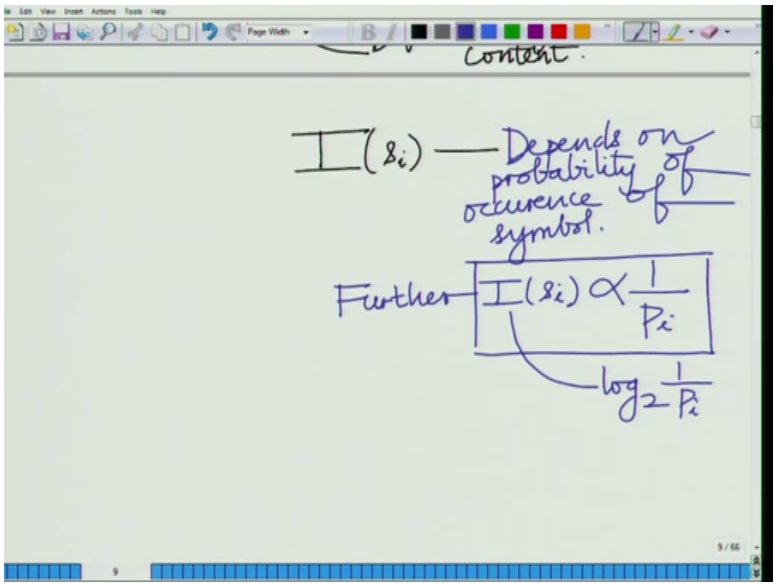
(Refer Slide Time: 22:24)



So, we are defining this is a very interesting and this is one of the fundamental definitions in information theory we are defining the information content which was, which we said is not very clear or which is the rather abstract quantity. So, the information content of a symbol. So, this is the definition of information content of.

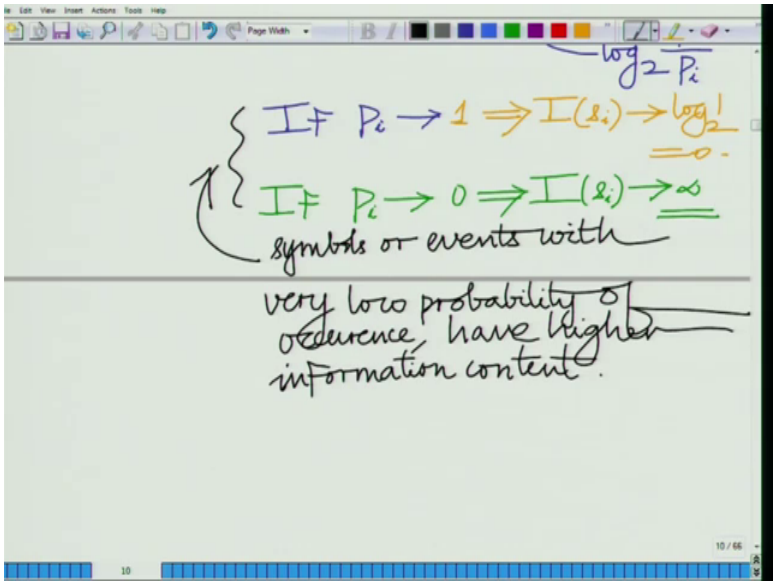
Now observe something very interesting, first of all I will talk more about this definition at definitely talk a lot more about this definition as we go through the various other aspects. But firstly, observe the most important thing is that the information of the symbol has nothing to do with how the symbol itself is represented S i rather it is everything through with the probability P i of the occurrence of the symbol.

(Refer Slide Time: 24:38)



For instance the symbol can be aware, the symbol can be a voltage level, the symbol can be for instance such as an event. So, it does not have anything to do with that particular symbol, but it basically deals with the probability of occurrence of that symbol all right and that is the important thing to observe. So, it depends I S i the information of the symbol depends on the probability of occurrence of the symbol, depends on the probability of occurrence of the symbol. Further you will observe, further I S i is proportional to 1 over P i, it is not proportional to P i it is rather proportional to 1 over P i, I S i is basically log to the base 2, 1 over P i.

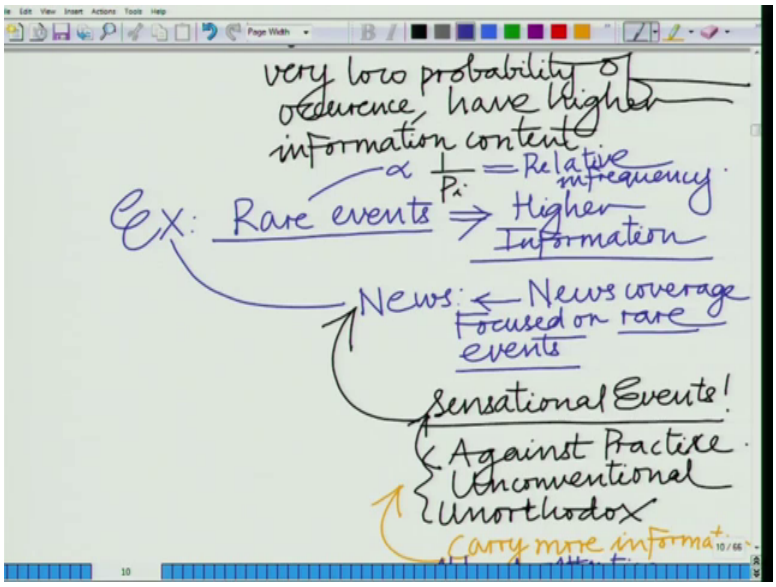
(Refer Slide Time: 25:37)



Which means if P i tends to 0, if P i tends to 1 that is it occurs very frequently I of S i tends to log 1 to the base 2 equals 0, if P i tends to 0 then I S i tends to infinite and this is very important aspect it shows that lowers that is symbols which have lower probability of occurrence have a much have much higher information associated with them.

So, symbols or events which are very low probability of occurrence have much higher information content associated with them. So, that is the important. So, that is the important takeaway point from this source. Symbols or events with very low probability of occurrence have higher information content. So, symbols which have a very low probability of occurrence have much higher information content and so therefore, rarer events, events which are rarer which are very low probability have higher information, correct.

(Refer Slide Time: 27:38)



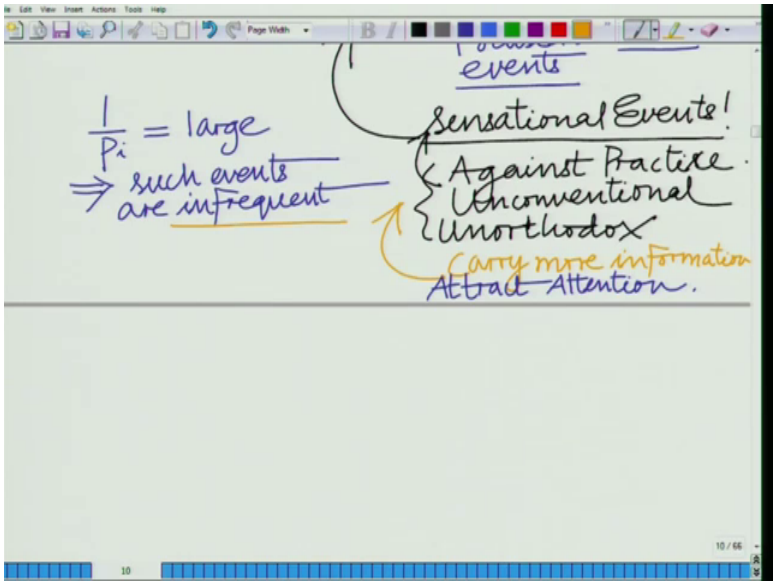
So, for example, right rare events, rare events have higher information. Think about for example news, a news coverage is focused on such rare events, for instance even event occurs on a daily basis or instance such as the local happening which occurs on a daily basis that is not things such as that are not covered in the news because did not constitute typically we know intuitively that they do not constitute much information. So, events which occur very rarely such as happenings on a national scale or let us say such as happenings on a national scale or calamities right more prominently calamities deserve a lot; so calamities attract a lot of attention, all right. So, coverage of such events, such rare

events which do not happen very frequently attracts. So, the rarer the event the more information, the more information it contains.

So, events which events which occur very frequently contain less information in fact, you can think of even sensational events such as events which do not occur events which are against the norm, events which are against the convention such events are assumed or associated with a large information a lot which deserve which is said to deserve a lot of attention because they have a lot large amount of information.

So, for instance the other aspect is you can think of sensational events. The very definition means that, means that against these, such events are against the practice, these are unprecedented or unconventional or unorthodox. All these basically carry more information and attract a lot of attention and therefore, attract lot of attention or warrant or attention. So, as to speak colloquially (Refer Time: 30:33) heads turn because they seem to carry a lot of information. So, the most important aspect, correct.

(Refer Slide Time: 30:23)



So, I would like to summarize this by saying once again that these rare events or sensational event events which are rare right, so these rare events which means rare events which means and we are using this 1 over P i we are saying that the rareness of an event or the relative infrequency with which an event occurs is proportional to 1 over P i. This is a measure of, 1 over P i is a measure of P i is a measure of the frequency 1 over P i can be said as a measure of the relative infrequency, such or basically 1 over P i is large

implies frequency or events are infrequent implies such events; such events are infrequent and therefore, carry a large amount of information.

So, what we have seen in this module is basically we have started with a very a very high level introduction to information theory we have motivated. This feed, this area of information theory, what is the relevance of information theory and we have basically started with a brief description of the system model and the set up of information theory, also characterized a fundamental definition of the information content of a particular alphabet belonging to the set, belonging to the source alphabet of a source.

We will start with this definition and further keep refining it progressively and look at various other norms, various other measures to characterize the information, content of a source and the relevance of all these aspects and the relevance in fact of this whole area of information theory in communications in our future modules.

Thank you very much.