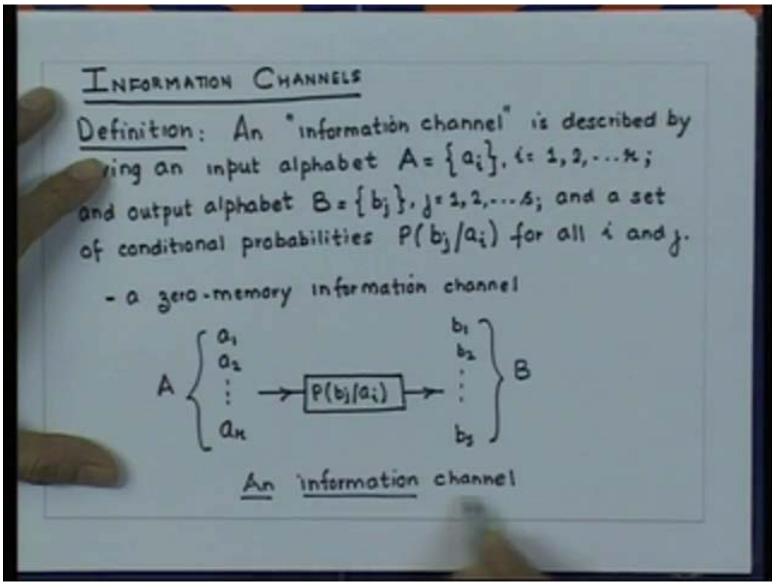
Information Theory and Coding
Prof S. N. Merchant
Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 20
Introduction to Information Channels

This will in turn lead to the possibility of coding in order to decrease the effect of errors caused by information channel. We will see that our information measure may be used to analyse, this type of coding as well as the type of coding already discussed. The central result of information theory and the most dramatic use of the concept of entropy will be discussed. This result in the form of Shannon's remarkable second theorem will use the entropy idea to describe, how we may utilise an unreliable information channel to transmit reliable information.

(Refer Slide Time: 01:46)

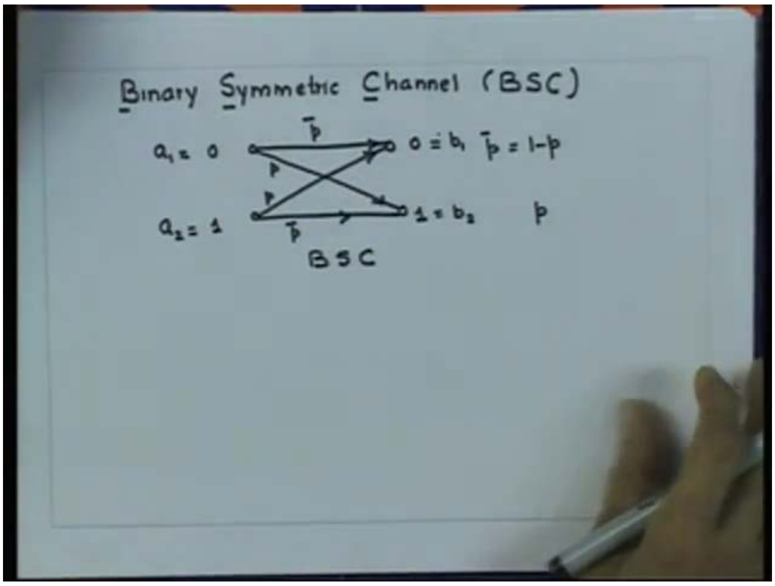


So, let us start with the definition of information channel. An information channel is described by giving an input alphabet. We will call that input alphabet as ‘A’consisting of input symbols denoted by a i where i takes a value from 1 to r which means the size of the input alphabet is r and output alphabet which we denote by B consisting of letters or symbols denoted by b j where j is equal to 1 to s. So, what it implies that it is not necessary for me to have the size of input alphabet and output alphabet to be the same.

With these alphabets, we have a set of conditional probabilities associated given by probability of bj given a i for all i and j. Probably the bj given ai is just the probability that the output symbol bj will be received if the input symbol ai is sent. The channel defined as given here is sometimes called a zero-memory information channel. A more general definition where the probability of a given output b j may depend upon several preceding input symbols or even output symbols is also possible. Such channels are referred to as channels with memory and we can represent this channel graphically as shown in the figures here.

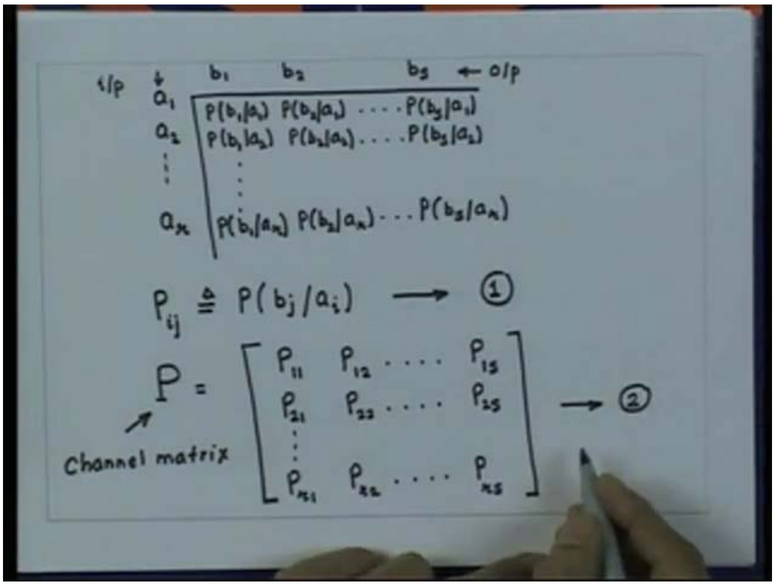
So, we have the inputs were channel given by a 1, a 2 up to a r and this passes through a channel where we have been given the conditional sets of probabilities and we have an output which is b 1, b 2 up to b s. This is the output alphabet b. This is our input alphabet a and this we can say is an information channel. So, a particular channel of great theoretical and practical importance is the binary symmetric channel.

(Refer Slide Time: 07:13)



Binary symmetric channel denoted as BSC, and the channel diagram for the binary symmetric channel is shown in the figure here. So, in this case I have the input alphabet consisting of binary letter 0 and 1 and output alphabet also consists of binary symbol 0 and 1. In this case, the input size and output alphabet size are the same. So, this is the channel diagram for the binary symmetric channel and as usual p bar is equal to 1 minus p. So, a 1 is equal to 0 and a 2 is equal to1, b 1 is equal to 0 and b2 is equal to 1. Now, this channel is symmetric because the probability of receiving a 1 if a0 is sent is equal to the probability of receiving a 0if a1 is sent and this probability is the probability that an error will occur is t. So, a convenient way of describing an information channel is to arrange the conditional probabilities of its output as shown in the figure.

(Refer Slide Time: 09:57)

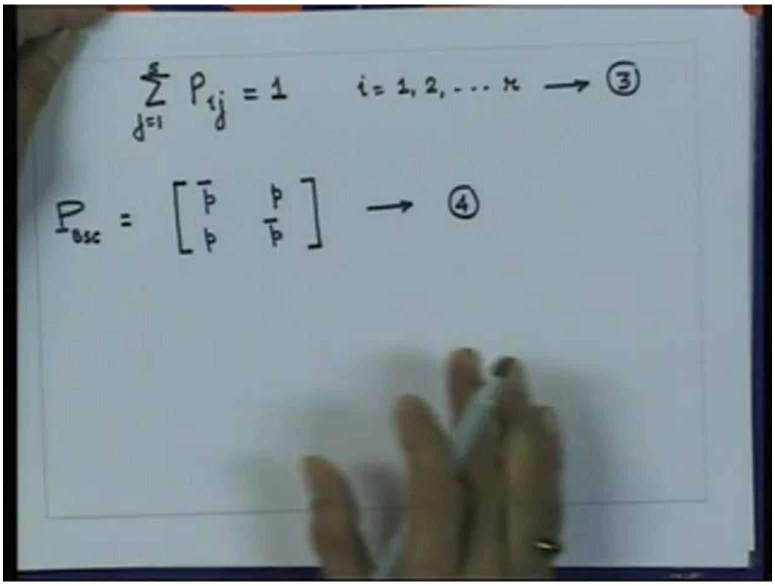


So, I have my inputs given by a 1,a 2 up to a r and outputs are b 1, b 2 up to b s. So, these are my outputs and this corresponds to my inputs and these are my conditional probabilities. Probability of b 2 given a 1 and finally, probability of b s given a 1, probability of b 1 given a 2, probability of b 2 given a 2,probability of ps given a 2,b 1 given a r probability of b2 given a r and finally, we have probability of bs given a r. So, channel can be described in a concise form as shown here. So, this will form a description of an information can. Now, note that each row of this array corresponds to a fix input and that the terms in this row are just the probabilities of obtaining the various b j at the output if the fix input is sent.

In order to simplify the description of information channel, it will be useful to have an aggregate notional notation for it. So, accordingly will define probability p i j is by definition probability of b j given a i. So, this is equation number1. So, if you use this notation then we can write a description shown here in a concise matrix form as follows. We have P 11, P 1 2upto P 1 s. P 2 1, P 2 2, P 2 s, and finally we have for the last row, P r 1, P r 2 and P r s. So, this becomes a channel matrix. So, an information channel is completely described by giving its channel matrix.

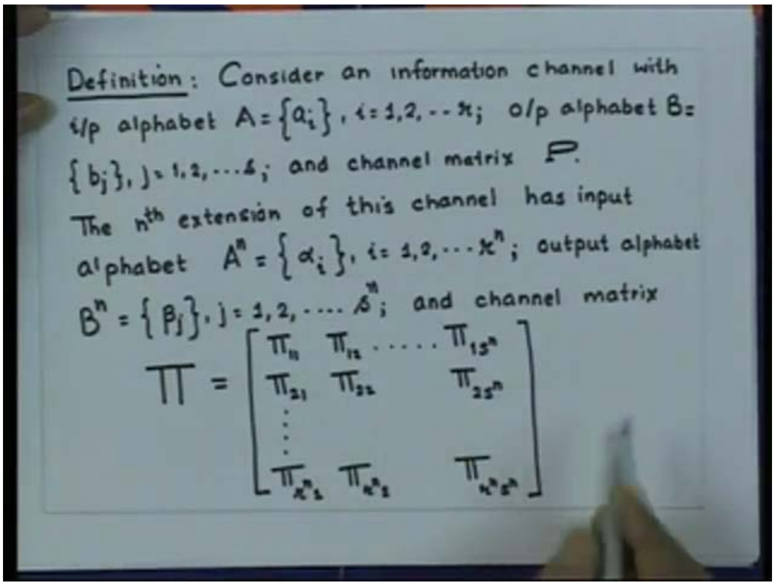
So, we will therefore use this matrix p interchangeably to represent both the channel matrix and the channel. Now, observe that each row of the channel matrix corresponds to an input of a channel and each column corresponds to a channel output. Also note, a fundamental property of this channel matrix, the terms in any row of the matrix must sum to 1. So, the summation out here will be 1 and the summation out here slimily will be 1 and such matrices are also called Markov matrices or Stokestic matrices. Now, this property is followed since if we send an input symbol ai, we must get some output symbol.

(Refer Slide Time: 14:52)



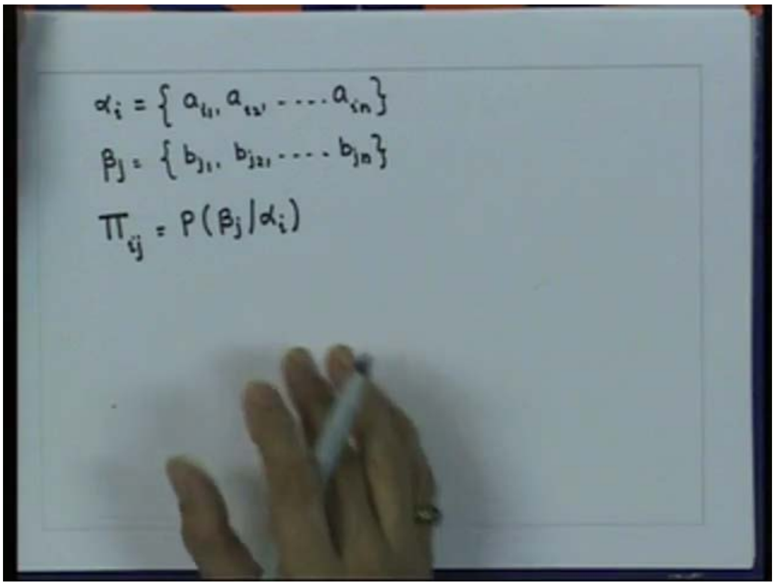
So mathematically what it means is that probability of p i j, j is equal to 1 to s should be equal to o1 for i equal to 1, 2 up to r. Now, the channel matrix of the binary symmetric channel will be simply a 2 x 2 matrix with the entries given as p bar, p, p and p bar. So, this will be the channel matrix for the binary symmetric channel. Now, just as we did in the case of information sources we may view the inputs and outputs of a channel in blocks of and symbols rather than individually.

(Refer Slide Time: 16:05)



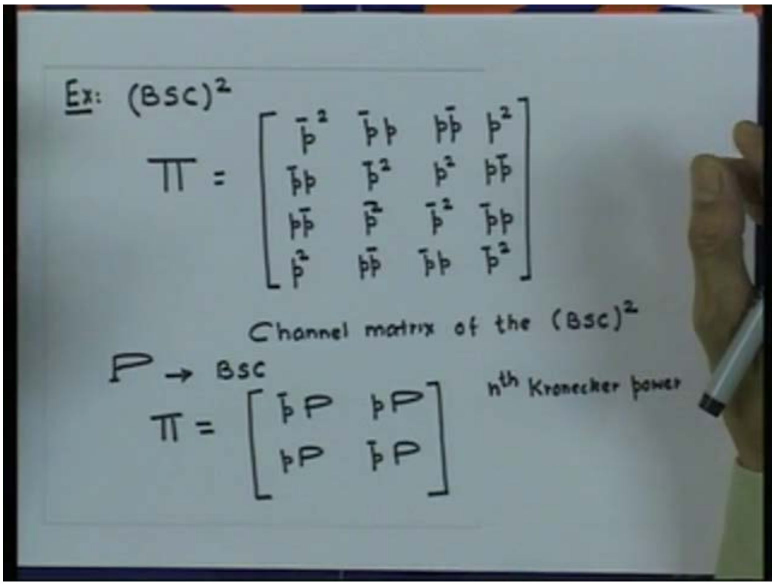
Thus, we defined the nth extension of a channel. So the definition will be as follows consider an information channel with input alphabet indicated by a, output alphabet indicated by band channel matrix p. Then the nth extension of this channel has input alphabet, which is indicated by a raise to n., with the letters of this alphabet indicated by alpha i where i ranges from 1, 2upto r raised to n and the output alphabet of this extension of the channel is indicated by b raised to n with the letters given by beta j where j is equal to 1, 2 up to s raise to n and for this channel, the channel matrix is indicated by capital pi with the entries in this matrix indicated by pi1 1,pi1 2 up to pi 1 s n. pi2 1,pi2 2 up to pi2 s n and finally, we have pi rn 1,pi r n 2,pi r raise to n s raise to n.

(Refer Slide Time: 20:14)



Now, each of the inputs alpha i consists of a sequence of n elementary input symbols. So, alpha i consists of ai 1, a i 2 up to ain and at each of the outputs beta j consists of a sequence of n elementary output symbols indicated by y j 1, b j 1, b j 2 up to b j n and the probabilities pi i j is equal to probability of beta j given alpha i. This probability consists of the product of the corresponding elementary symbol probabilities. So, just as was the case when we defined the extension of an information source the extension of an information channel is not really a new concept but just a new way of viewing an old concept. So, merely by looking at symbols of some channel in blocks of length n, we obtain the nth extension of that channel.

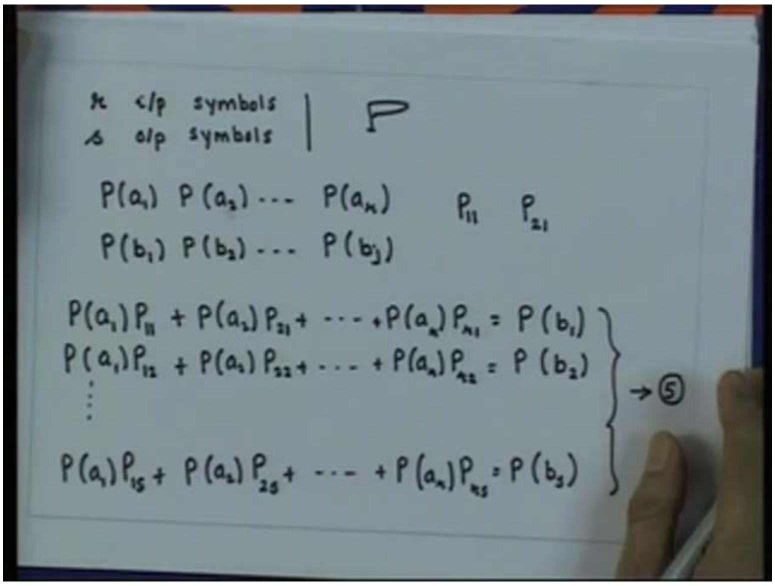
(Refer Slide Time: 22:46)



Let us look at 1 example let us look at a second extension of the binary symmetric channel. Second extension of the binary symmetric channel will have a channel with four input symbols and four output symbols and its channel matrix will be as shown here. So, we will indicate the second extension of binary symmetric channel by this notation and the channel matrix for this channel will be given as shown here. This is p squared, p square pp bar, p bar p, p bar square. So, this is channel matrix of the second extension of a binary symmetric channel. Now, we note that the channel matrix of the binary symmetric second extension of binary symmetric channel may be written as the matrix of matrices. So, let p as before be the channel matrix of the binary symmetric channel than the channel matrix of the second extension of the binary symmetric channel can be written as p bar p, p P matrix, p P, p bar P.

Now, this metric is known as the Kronecker square of the matrix P. In the more general case, the channel matrix of the nth extension of a channel is the nth Kr1cker power of the original channel matrix. Now, earlier in this course we use the information measure to measure of the average amount of information produced by a source. The function of an information channel, however, is not to produce information but to transmit information from the input to the output. We expect therefore, to use the information measure to measure the ability of a channel to transport information.

(Refer Slide Time: 26:48)

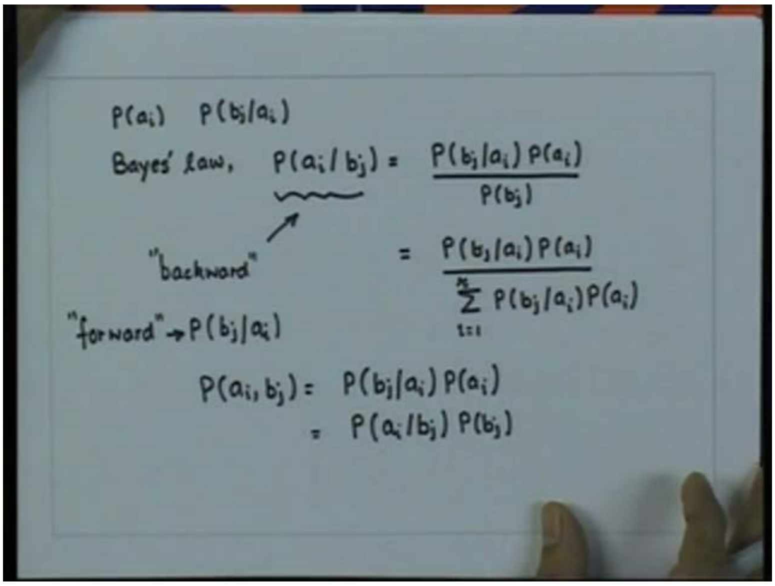


Now we proceed to investigate the amount of information a channel can transmit. So, let us look at probability relationships in a channel. So, let us consider an information channel with r input symbols and s output symbols and this channel is defined by the channel matrix P. Now, select the input symbols according to the probability P a 1, P a 2 up to P a r. We select the input symbols for transmission through this channel. Then the output symbols will appear according to some other set of probabilities and let us indicate that set as P b 1, P b 2 up to P b g.

Now, the relationships between the probabilities of the various input symbols and the probability of the various output signals can be easily derived. So, let us do that. Now, if you look at the output symbol y 1, for example, there are r ways in which we might receive output symbol b 1. So, if a 1 is sent, b 1 will occur with probability P 1 1. If a 2 is sent, b 1 will occur with probability P 2 1etc.Therefore, we can write probability of a 1, P 1 1 plus probability of a 2, P 21 plus probability of a r, probability of r1. This will be equal to probability of b1.Similarly, we can write probability a 1, P 12plus probability a 2, P 2 2, a r, probability r2 is equal to probability of b2, and finally probability of a 1, probability P 1 s plus probability a 2, probability 2 s, a r, probability r s is equal to probability b s.

These sets of relationship let us call as equation number 5. So, this equation provides us with the expression for the probabilities of the various output symbols if you are given the input probabilities P ai and the channel matrix P. That is the matrix of conditional probabilities. Probabilities of b j given a i. Now, for a further discussion we will assume that we are given the probabilities P a 1, P a 2 of P a r and the condition probabilities of b j given ai. So, that probability of b j s may be calculated from this equation. Note however that if you are given the output probabilities that is probability b 1 up to b s and probabilities of b j given ai. It may not be possible to invert the system of linear equations in order to obtain the probabilities P a r.

(Refer Slide Time: 32:19)

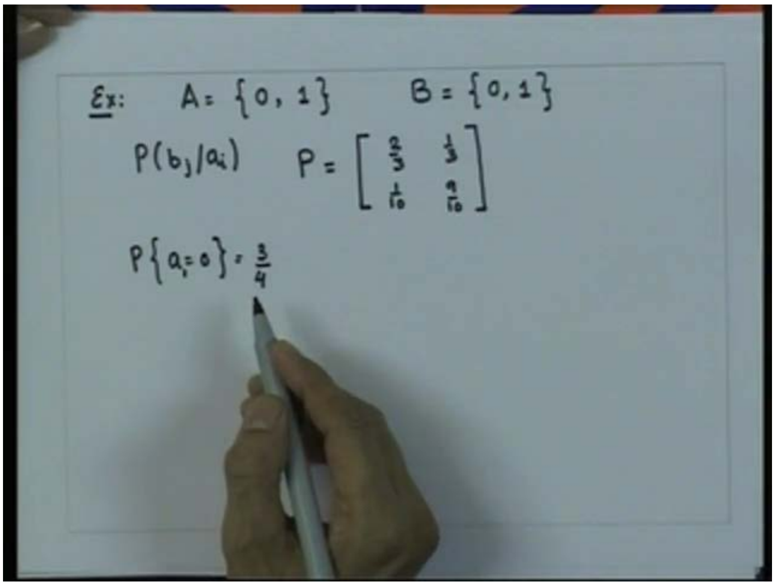


For example, if you take a binary symmetric channel with P equal to half, then any set of input probabilities will lead to output symbols which are equi-probable. So, in general, there may be many input distribution which lead to the same output distribution. If you are given the input distribution on the other hand, we may always calculate a unique output distribution with the aid of these equations.

In addition to the marginal probabilities of the output symbol, there are 2more sets of probabilities associated with an information channel which we may calculate form the input marginal probabilities and the channel matrix. So, let us look at those probabilities. So, we have been given probability of P ai and we have the channel matrix. So, we have been given conditional probabilities b j given a i. Now, according to Baye'slaw, the conditional probability of an input a i given that an output b j has been received is probability a i given b j is equal to probability of b j given a i multiplied by probability a i divided by probability b j and this can be written as this expression has been substituted by this based on our earlier observation given by the sets of equation.

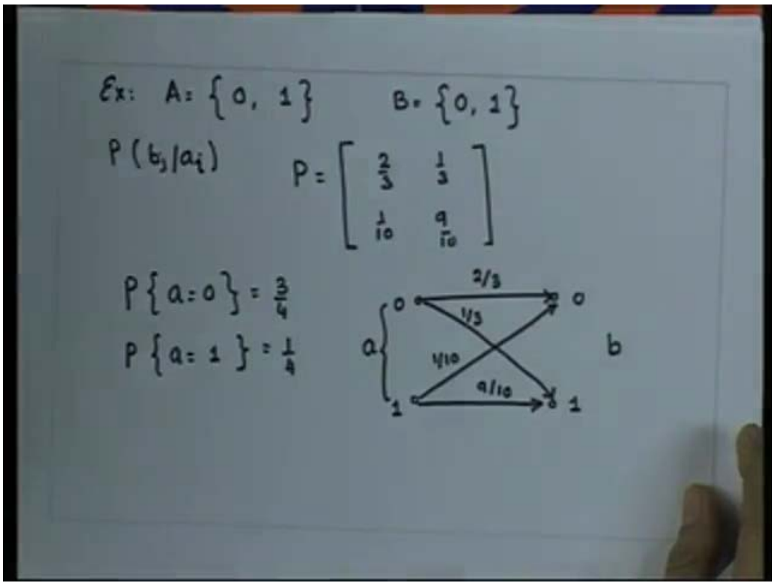
So, the probabilities a i given b j are sometime referred as backward probabilities in order to distinguish them from the forward probabilities which are indicated by probability b j given a i. Now, the numerator of the right side of this equation is the probability of the joint event a i b j. So, this is equal to probability b j given a i, probability a i. And this quantity is also equal to probability of a i given b j multiplied by probability b j. Now, let us illustrate the calculation of the various probability associated information channel.

(Refer Slide Time: 36:53)



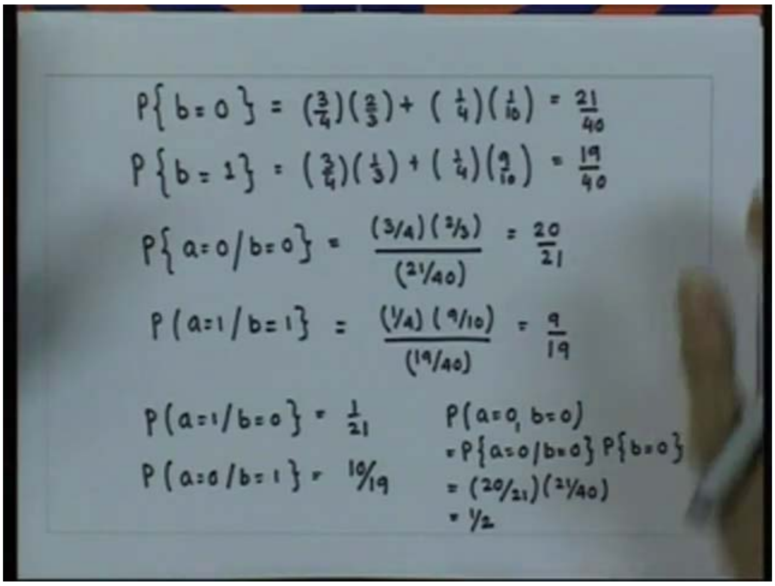
We again take an example of a binary channel, that is we have input alphabet a consisting of 2 letters 0 1 and output alphabet consisting of 2 symbols again 0 and 1 and we assume that the channel matrix, that is the condition set of probabilities b j given a I, has been given to us as follows. So, some of this row is 1. Again some of this row is equal to 1 as discussed earlier. Now, we associate the rows and columns of this matrix with the input and output symbols in the natural order. Therefore, probability of b equal to 0given a to 0 is two-third probability of b equal to 1 given a is equal to 0 is one-third and similarly, for this entries.

(Refer Slide Time: 38:53)



So, we also assume that probability of a equal to0, a1 equal to 0 is equal to three-fourth and probability a equal to 1 is equal to one-fourth. Now, this information can be neatly summarisedas shown in the figure here.

(Refer Slide Time: 41:00)



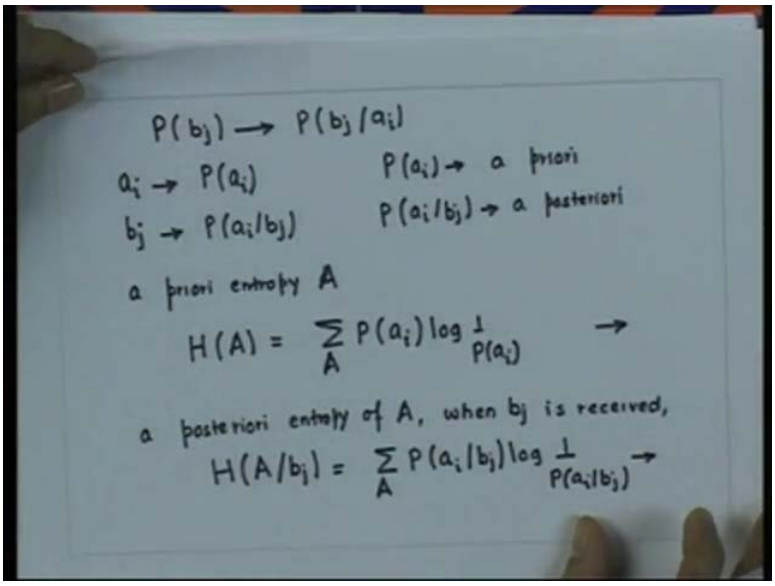
So, the probability of the output symbols 0, and 1 can be obtained with the help of these equations, and probability of b equal to 0 is equal to probability of 0 given 0 was transmitted multiplied by probability of 0 transmission or probability of receiving 0 given 1 was transmitted multiplied by probability of 1 being transmitted. So, using this you can write this as three- fourth multiplied by two-third plus one-fourth multiplied by one-tenth. This comes out to be 21 by 40.

So similarly, probability b equal to 1is equal to three-fourth by one-third plus one-fourth multiplied by9by 10 and this is equal to 19 by40. So, as I check probability of b equal to 0plus probability of b equal to 1. This should sum up equal to 1. Now, the conditional input probabilities can be obtained as probability of a equal to 0 given b equal to0 is equal to three-fourth multiplied by two-third divided by 21 by40is equal to 20 by 21.One-fourth multiplied by 9 by 10 whole over19 by 40 is equal to 9 by 19.Theother two backward probabilities maybe similarly, obtained.

So, probability of a equal to 1 given b equal to 0 is equal to 1by 21 and probability of a equal to 0 given b equal to 1 is equal to 10 by 19. It is to be noted that this plus this equals to 1 again this plus this is equal to 1 and if you are interested in calculating the probability of area of joint events then we can calculate the probability of a equal to 0, b equal to 0 as probability of a equal to 0 given b equal to 0 multiplied by probability of b equal to 0 and this is equal to 20 by 21 multiplied by 21 by 40. So, this is equal to half.

Now, the various outputs symbols of a channel occur according to the set of probabilities which is given by probability b j. Note that a probability of a given output symbol, bj, is P b j, if you do not know which input symbol is sent. On the other hand, if we do know that input symbol is a i then the probability that a corresponding output will be b j changes from probability b j to a probability b j given ai.

(Refer Slide Time: 45:48)

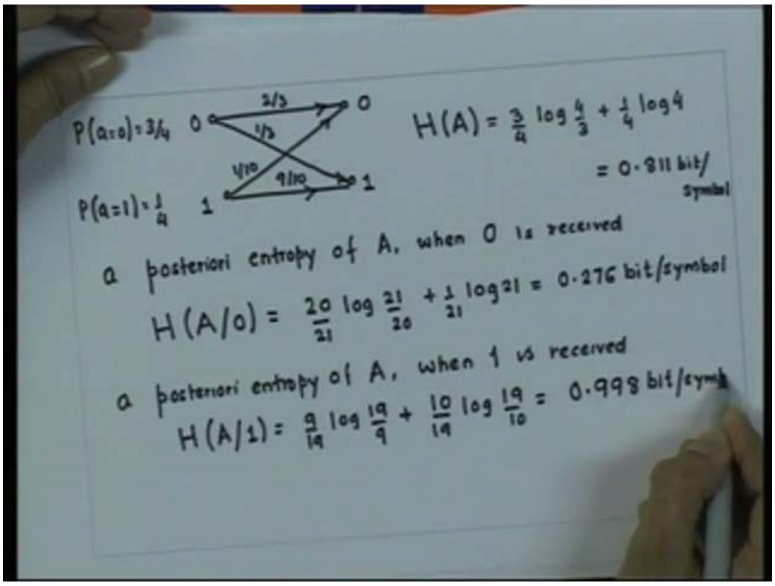


So, once the transmission of symbol a i is known, the output symbol probability P b j changes to probability of b j given a i. Likewise, we recall that the input symbol, a i, is chosen with probability P a i. If you observe the output symbol b j, however, we know the probability that a i is the corresponding input symbol is a i given bj.

Now, let us focus our attention on this change in news on the probabilities of the various input symbols by the reception of a given output symbol bj. We shall refer to probability P a i as the a priory probabilities of the input symbols that is the probabilities of the a i before the reception of an output symbol. And probability a i given b j will be denoted as a posteriori probabilities of the input symbol. That is the probability after the reception of b j. We know that we can calculate the entropy of the set of input symbols with respect to both the sets of probabilities.

So, let us calculate the a priory entropy of a. So, a priory entropy of the source a will be given as H A equal to probability P ai log 1 P a I summation over all a i and that will denoted by A and a posteriori entropy of a when b j is received is given as a i b j log of overall a i. Now, the interpretation of these two quantities follows directly from Shannon's first theorem. H A is the average number of bin, it is needed to represent a symbol from a source with the a priori probabilities given by P ai. i is equal to 1to r. This is the average number of bin its needed to represent a symbol from a source with the a posteriori probabilities P a i given b j for all i is equal to 1 to r. So, let us calculate these two quantities for a binary symmetric channel which we considered earlier in the class.

(Refer Slide Time: 50:35)



Today, so we will take that example again. So, binary symmetric channel is given by this channel diagram probabilities of 0 transmission. Let us assume P a equal to 0 equal to three-fourth. Probability of transmitting 1 is equal to one-forth. Now, for this we can calculate entropy of the source that is equal to three-forth log4 by 3 plus 14 log4. This is equal to 0.811 bits per symbol. If you receive the symbol 0 at the output of the channel then our a posteriori probabilities which we have calculate earlier and using those a posterior probabilities, we can calculate it a posterior entropy as follows. So, a posteriori entropy of a when 0 is received is given by this expression.

This is equal to0.276 bit per symbol. Now, if we receive the symbol 1 on the other hand, a posterior entropy of a will be given by is equal to 0.998 bit per symbol. Hence, if a 0 is received the entropy that is uncertainty about which input was sent, decreases whereas, if a1 is received uncertainty increases. In today's class, we have looked at a definition of information channel and seen how to represent an information channel and also looked into the calculation of the various probabilities associated with an information channel. In the next class, we will revisit Shannon's first theorem and generalise this Shannon’s first theorem from the viewpoint of an information channel.