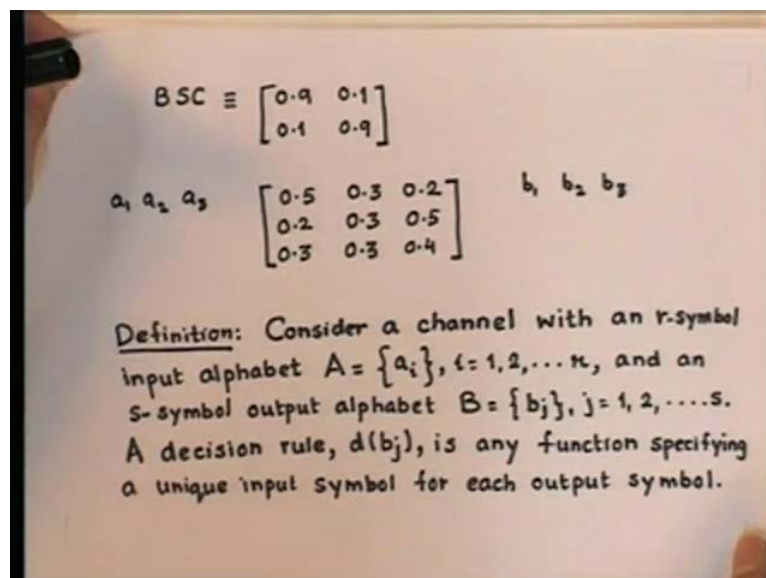


**Information Theory and Coding**  
**Prof. S. N. Merchant**  
**Electrical Engineering**  
**Indian Institute of Technology, Bombay**

**Lecture - 26**  
**Shannon's Second Theorem**

Shannon's second theorem deals with the amount of error free information. We can get through a channel. In order to appreciate more fully the significance of this theorem, let us look into the question of the error probability of a channel. For some of the simple channels, which we have studied such as binary symmetric channel and r symmetric channel, it is intuitively clear what it is reasonable to call the error probability of a channel. Nevertheless, in the simple cases, also the error probability of a channel is dependent upon a factor not yet discussed in the study of information channel. Let us consider a simple binary symmetric channel with the channel matrix as follows.

(Refer Slide Time: 01:53)

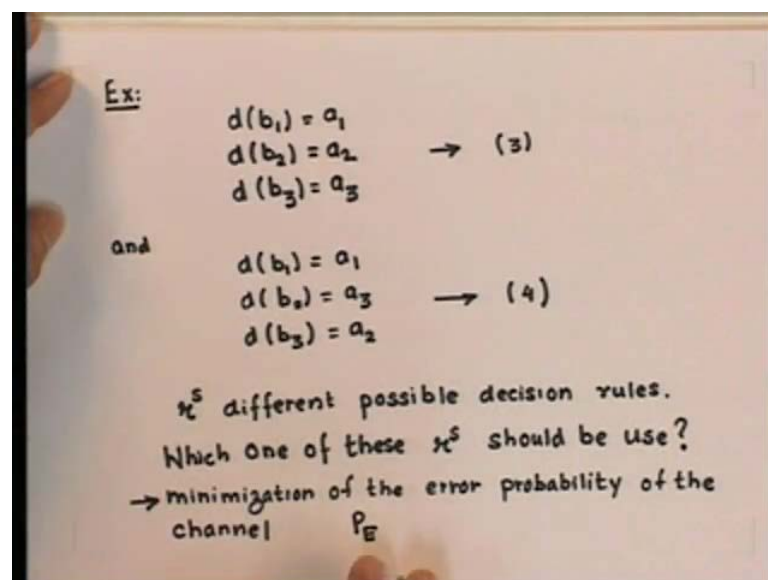


Ordinary, we would say that the probability of error of this channel is 0.1, note however that this statement depends upon the assumption that the channel is being used in a reasonable manner. If the receiver were to examine the channel output and decide that a one was sent when a 0 is received and vice versa the probability of error of this channel would be 0.9. Of course, this is not a reasonable use of the channel, but it is a possibility which we should also consider, the error probability of the channel depends upon how

the receiver interprets the output symbol of the channel. To bring out this point in a more meaningful case take the channel with a channel matrix given as follows. This channel has three inputs  $a_1$ ,  $a_2$  and  $a_3$  and three outputs  $b_1$ ,  $b_2$  and  $b_3$ .

When a particular channel output is received which input should we say was sent, this question prompts us to the following definition. Consider a channel with an  $r$  symbol input alphabet identified as capital A is equal to the letters of the alphabet  $a_i$ , for  $i$  equal to 1, 2 up to  $r$  and an  $s$  symbol output alphabet identified as capital B with the letters of that alphabet  $b_j$  where  $j$  is equal to 1, 2 up to  $s$ . Now, a decision rule define as  $d$  of  $b_j$  is any function specifying a unique inputs symbol for each output symbol, in order to understand this definition in a more meaningful manner. Let us take a simple example.

(Refer Slide Time: 07:05)



Two possible decision rules for the channel which is given by this channel matrix, let us indicate it by as equation two and this channel matrix as equation one. So, two possible decision rules for the channels of two possible decision rules for the channel given by this channel matrix are  $d(b_1)$  is equal to  $a_1$ ,  $d(b_2)$  is equal to  $a_2$  and  $d(b_3)$  is equal to  $a_3$ . Another decision rule for the same channel could be as follows  $d(b_1)$  is equal to  $a_1$ ,  $d(b_2)$  is equal to  $a_3$  and  $d(b_3)$  is equal to  $a_2$ . So, for a channel with  $r$  inputs and  $s$  outputs there are  $r^s$  different possible decision rules.

Therefore, the question which prompted our definition of this, therefore the question which prompted our definition of a decision rule may therefore be rephrased as, which

one of these decision rules should we use. Now, the answer to this question will depend upon what we are trying to accomplish, but a reasonable goal for our purpose is the minimization of the error probability of the channel. Hence, we seek a decision rule which minimizes the error probability of the channel. To find such a rule we calculate the probability of error denoted by  $P_E$ .

(Refer Slide Time: 10:53)

average of  $P(E/b_j)$

$$P_E = \sum_B P(E/b)P(b) \rightarrow (5)$$

$d(b_j)$      $P(b_j)$

For a fixed decision rule,  $d(b_j) = a_i$

$$P(E/b_j) = 1 - P[d(b_j)/b_j] \rightarrow (6)$$

$P(a_i/b_j)$      $d(b_j) = a^*$      $\rightarrow 7(a)$

$$P(a^*/b_j) \geq P(a_i/b_j) \text{ for all } i \rightarrow 7(b)$$

Now, this probability of error may be written as the average of probability of  $E$  given  $b_j$ . This is the conditional probability of error given that the output of the channel is  $b_j$ . So, we can write the probability of error of the channel is  $P_E$  equal to summation of  $P$  of  $E$  given  $b_j$  multiplied by probability of  $b$  multiply  $P_E$  is equal to summation of probability of  $E$  given  $b$  multiplied by probability of  $b$  summed over the alphabet  $B$ . As discussed earlier and without any loss of generality ((Refer time 11:53)) we will drop the subscript  $j$  with  $b$ .

Now, this equation expresses the error probability as a sum of non negative terms, therefore in order to minimize the probability of error by choice of a decision rule  $d(b_j)$ , we may select  $d(b_j)$  to minimize each term in the sum separately. Now, probability of  $b_j$  does not depend upon the decision rule we use, so it is equivalent to choose  $d(b_j)$  to minimize the conditional probability of error that is  $P$  of  $E$  given  $b_j$ .

So, for a fixed decision rule that is  $d(b_j)$  equal to some  $a_i$  probability of error given  $b_j$  is equal to 1 minus probability of  $d(b_j)$  given  $b_j$ . Now, since our decision rule is fixed,

probability of  $d(b_j)$  given  $b_j$  is the backward probability; that is probability of  $a_i$  given  $b_j$ . So, in order to minimize this equation for each  $b_j$ , we choose  $d(b_j)$  equal to  $a^*$ , where  $a^*$  is defined by probability of  $a^*$  given  $b_j$  is greater than equal to probability of  $a_i$  given  $b_j$  for all  $i$ . Let us denote this equation by 7 a and this by 7 b, so in other words the channel probability is minimize. If we use that decision rule, we chooses for each output symbol the input symbol with the highest probability.

(Refer Slide Time: 15:25)

Conditional maximum-likelihood decision rule.

$$\frac{P(b_j/a^*)P(a^*)}{P(b_j)} \geq \frac{P(b_j/a_i)P(a_i)}{P(b_j)} \text{ for all } i \quad (9)$$

$$d(b_j) = a^* \rightarrow q(a)$$

where

$$P(b_j/a^*) \geq P(b_j/a_i) \text{ for all } i \rightarrow q(b)$$

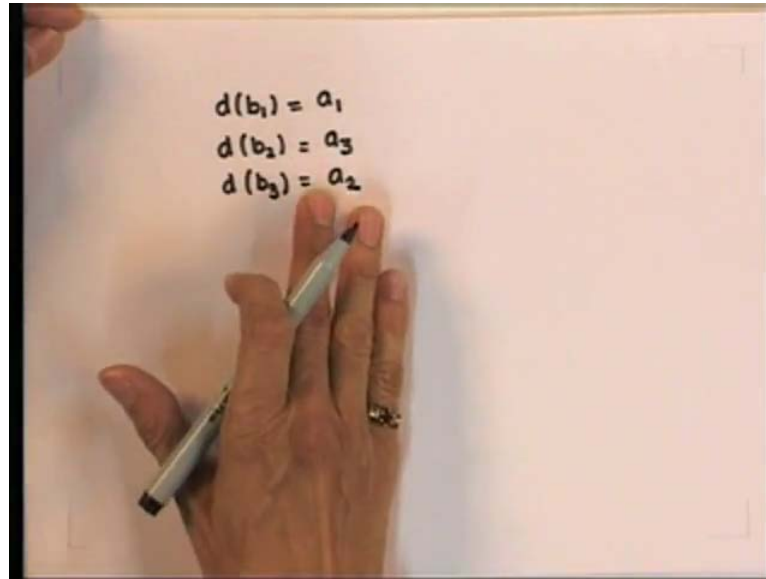
"maximum-likelihood decision rule"

Now, this decision rule is sometime called conditional maximum likelihood decision rule. The conditional maximum likelihood decision rule depends upon the a priority probabilities  $p$  of  $a_i$ . So, we may use Base law to write this equation 7 b as probability of  $b_j$  given  $a^*$  multiplied by probability of  $a^*$  divided by probability of  $b_j$ . This should be greater than equal to probability of  $b_j$  given  $a_i$  multiplied by probability of  $a_i$  divided by probability of  $b_j$  for all  $i$ .

Hence, when all the a priority probabilities are all equal, the conditional maximum likelihood decision rule may be written as  $d(b_j)$  is equal to  $a^*$ , where probability of  $b_j$  given  $a^*$  is greater than equal to probability of  $b_j$  given  $a_i$  for all  $i$ . The decision rule defined by this equation 9 b is known as the maximum likelihood decision rule. The maximum likelihood decision rule does not depend upon the priority probabilities. So, when the priority probabilities are all equal the maximum likelihood decision rule results in a minimum value for the probability of error.

Even if the priority probabilities are not equal or in those cases where these probabilities are unknown, we could still employ the maximum likelihood decision procedure. But in such a case we may not be assured of obtaining a minimum value of the channel error probability. Let us take the earlier example, which we had discussed for this channel matrix.

(Refer Slide Time: 19:10)



Now, we can write a maximum likelihood decision rule as follows  $d(b_1) = a_1$ ,  $d(b_2) = a_3$  and  $d(b_3) = a_2$ . Now, it is important to note that this rule is not unique, there are in fact three maximum likelihood decision rules for this channel. Now, the error probability using any given decision rule is easily obtained with the help of equation 5 and equation 6 as follows.

(Refer Slide Time: 20:05)

$$\begin{aligned} P_E &= \sum_B P(E/b)P(b) \\ &= \sum_B P(b) - \sum_B P[d(b)/P(b)] \\ &= 1 - \sum_B P[d(b), b] \rightarrow (10) \end{aligned}$$

$d(b_j) = a^*$  and  $b_j$  for each  $j$

$$\bar{P}_E = 1 - P_E$$
$$\bar{P}_E = \sum_B P(a^*, b) \rightarrow (11)$$
$$\sum_{A, B} P(a, b) = 1 \rightarrow (12)$$

Probability of error is equal to summation of probability  $E$  given  $b$ , probability of  $b$  summed over alphabet  $b$ . This is equal to probability of  $b$  summed over alphabet  $b$  capital  $B$  that is minus probability of  $d$   $b$  given probability  $b$ , summed over alphabet  $b$ , which can be simplified as 1 minus summation over capital  $B$  probability of  $d$   $b$  comma  $b$ . The terms in the summation of this equation number ten are just the joint probability that  $d$   $b$   $j$  is equal to  $a$  star is transmitted and  $b$   $j$  is received for each  $j$ . Hence, defining probability  $E$  compliments as 1 minus probability  $P_E$ , we can write  $\bar{P}_E$  is equal to summation of probability  $a$  star  $b$  summed over  $b$ . Now, we know that summation of probabilities that is joint probability  $a$   $b$  over alphabet  $a$   $b$  is equal to 1. So, using this relationship we can rewrite the equation number 10 as follows.

(Refer Slide Time: 22:49)

$$P_E = \sum_{B, A-a^*} P(a, b) \rightarrow (13)$$

$$\sum_{A-a^*} d(b_j) = a^*$$

$$P_E = \sum_{B, A-a^*} P(b/a) P(a) \rightarrow (14)$$

$$P_E = \frac{1}{r} \sum_{B, A-a^*} P(b/a) \rightarrow (15)$$

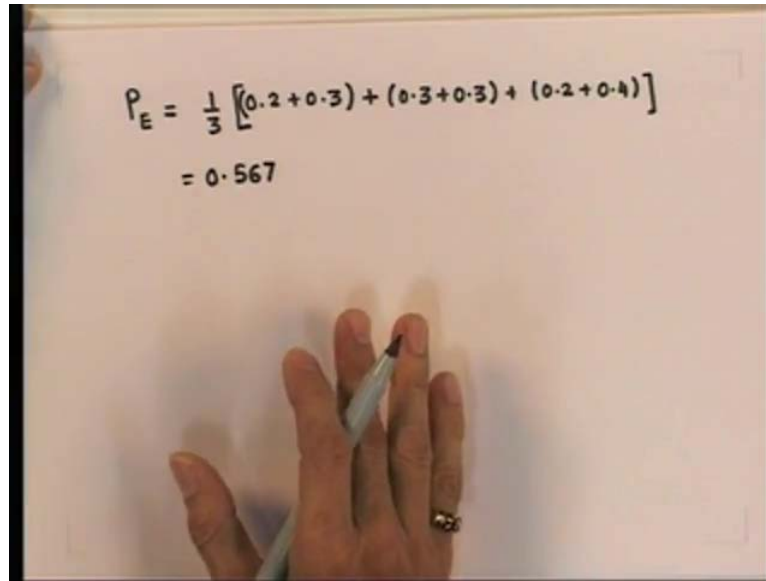
$$d(b_j)$$

Probability E is equal to summation of joint probability of a, b summed over the alphabet B and A minus a star. The notation summation A minus a star is meant to indicate that this sum is over all members of the an alphabet except d b j is equal to a star. So, an alternative way of writing this same equation number thirteen is probability of E is equal to probability of b given a multiplied by probability of a. Now, if you assume that the a priority probability that is P a are all equal, then this equation can be written as P E is equal to 1 by r summation over capital B capital A minus a star probability of b given a.

Now, this equation is of some interest since for the special case of equal priority probabilities, it provides an expression for the channel error probability in terms of sum over the elements of the channel matrix P b given a. Now, this sum is over all elements of the channel matrix except that one term corresponding to d of b j, which is omitted from each column.

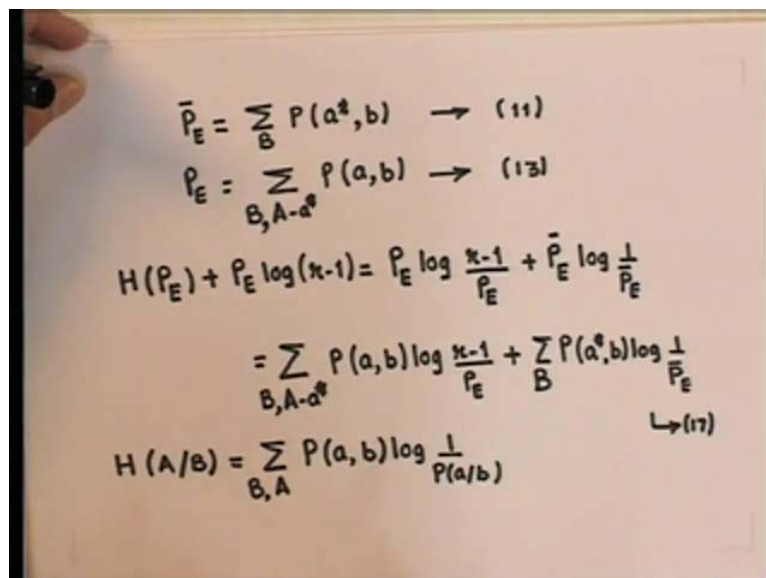
Now, for the same channel matrix which we considered earlier, that is this channel matrix, if we assume that all the three input symbols that is a 1 a, 2 a, 3 are chosen with equal probabilities, then we can apply a maximum likelihood decision rule. If we use that maximum likelihood decision rule, then the minimum probability of error, which will get for such a channel would be given by using this equation as follows.

(Refer Slide Time: 26:15)


$$P_E = \frac{1}{3} [(0.2+0.3) + (0.3+0.3) + (0.2+0.4)]$$
$$= 0.567$$

Probability of E is equal to 1/3 summation over. So, now we have studied how to find the error probability for a channel, but this we have done without reference to entropy or equivocation or mutual information. Now, we will see the connection between these two separate sets of ideas, as a first step in this direction we will derive upper and lower bounds on the equivocation in terms of the error probability of a channel. So, in our derivation to follow we shall make repeated use of these two equations which we derived earlier.

(Refer Slide Time: 27:34)


$$\bar{P}_E = \sum_B P(a^*, b) \rightarrow (11)$$
$$P_E = \sum_{B, A-a^*} P(a, b) \rightarrow (13)$$
$$H(P_E) + P_E \log(\kappa-1) = P_E \log \frac{\kappa-1}{P_E} + \bar{P}_E \log \frac{1}{\bar{P}_E}$$
$$= \sum_{B, A-a^*} P(a, b) \log \frac{\kappa-1}{P_E} + \sum_B P(a^*, b) \log \frac{1}{\bar{P}_E}$$
$$H(A/B) = \sum_{B, A} P(a, b) \log \frac{1}{P(a/b)} \quad \rightarrow (17)$$



That is  $P$  complement  $E$  is equal to summation of joint probabilities  $a$  star  $b$ , this was equation number eleven and probability of  $E$  equals summation of joint probabilities  $a$ ,  $b$  summed over alphabet  $B$  and  $A$  minus  $a$  star. This was equation number thirteen. Using these two relationships we construct the identity  $H$  of  $P$   $E$  plus  $P$   $E$  log to the base 2 of  $r$  minus 1 is equal to probability of error log of  $r$  minus 1 upon  $P$   $E$  plus  $P$  complement log of 1 by  $P$  complement  $E$ .

This can be rewritten using this equations as follows, probability  $a$ ,  $b$  log of  $r$  minus 1 upon probability of 1 plus this term can be rewritten from this expression as probability  $a$  star  $b$  log of 1 by probability, complement probability error summed over alphabet  $B$ . Now, the equivocation that is  $H$  of  $A$  given  $B$  is by definition, we have seen equal to summation of probability  $a$ ,  $b$  log of 1 by probability  $a$  given  $b$  summed over  $B$  and  $A$  alphabet. Now, this equivocation that is  $H$  of  $A$  given  $B$  can be also written in terms of the same sort of summations which we have written in this equation number 17.

(Refer Slide Time: 31:22)

$$H(A/B) = \sum_{B, A-a^*} P(a,b) \log \frac{1}{P(a,b)} + \sum_B P(a^*,b) \log \frac{1}{P(a^*/b)} \quad \rightarrow (18)$$

$$H(A/B) - H(P_E) - P_E \log(r-1) = \sum_{B, A-a^*} P(a,b) \log \frac{P_E}{(r-1)P(a,b)} + \sum_B P(a^*,b) \log \frac{\bar{P}_E}{P(a^*/b)} \quad \rightarrow (19)$$

$$\log_a x = \frac{\log_b x}{\log_b a}$$

So, if we follow the same procedure we can write  $H$  of  $A$  given  $B$  equal to summation of joint probabilities  $P$   $a$ ,  $b$  log of 1  $P$   $a$  given  $b$  summed over  $B$ ,  $A$  minus  $a$  star plus summation of  $P$   $a$  star  $b$ , log of 1 by  $P$   $a$  star given  $b$  over alphabet  $B$ . Now, if we subtract this equation, that is equation number seventeen from equation number eighteen, then we get the relationship as follows.  $H$  of  $A$  given  $B$  minus  $H$  of probability of error minus probability of error, log to the base 2  $r$  minus 1 is equal to summation  $B$ ,  $A$  minus  $a$  star,

probability of a, b log of probability of error divided by r minus 1 into probability of a given b plus summation over alphabet B of probability of a star b log of P compliment divided by probability of a star given b. Now, this logarithms are to the base 2, so if we change the base of the logarithm by using the following relationship log of a x is equal to log b x upon log of a to the base b. We can rewrite equation number nineteen in terms of natural logarithms as follows.

(Refer Slide Time: 35:05)

The image shows a handwritten derivation on a whiteboard. The first line is:

$$(\log e)^{-1} [H(A/B) - H(P_E) - P_E \log(r-1)]$$

The second line is:

$$= \sum_{B, A-a^*} P(a,b) \ln \frac{P_E}{(r-1)P(a/b)} + \sum_B P(a^*,b) \ln \frac{\bar{P}_E}{P(a^*/b)}$$

Below this is the label  $\hookrightarrow (20)$ . The third line is the inequality:

$$\ln x \leq x-1 \quad \rightarrow (21)$$

The fourth line is:

$$\sum_{B, A-a^*} P(a,b) \left[ \frac{P_E}{(r-1)P(a/b)} - 1 \right] + \sum_B P(a^*,b) \left[ \frac{\bar{P}_E}{P(a^*/b)} - 1 \right]$$

The fifth line is:

$$= \left[ \frac{P_E}{(r-1)} \sum_{B, A-a^*} P(b) \right] - \underbrace{\sum_{B, A-a^*} P(a,b)}_{P_E} + \left[ \bar{P}_E \sum_B P(b) \right] - \underbrace{\sum_B P(a^*,b)}_{\bar{P}_E}$$

The sixth line is:

$$= 0$$

Below this is the label  $\hookrightarrow (22)$ .

Log of e inverse H of A given B minus H of P E minus P E log r minus 1 equals probability of a, b natural logarithm of P E divided by r minus 1 probability of a given b summed over alphabet B minus a star plus summations of probability a star b log of P compliment P a star b over the alphabet B. Now, the advantage of writing in the natural logarithm is we can use the inequality, which we had seen in our earlier lectures that is log of x is always less than equal to x minus 1. So, if we use this inequality, then we can write each term in this summation on the right hand side of equation number 20 as follows.

Summation of probability a, b summed over B minus a star, we apply this relationship to the quantity here, as shown here, will give us P E r minus 1 probability of a given b minus 1 plus summation probability a star b compliment P E upon probability a star b minus 1. Now, this can be so the right hand side of the expression of the equation number 20 can be shown to be less than equal to this summation. This can be further

simplified as  $P_E$  over  $r$  minus 1 summation of probability  $b$  minus probability of  $a, b$  B A minus  $a$  star plus  $P$  compliment summation of  $P b$  B minus  $a$  star  $b$  summed over B. Now, this quantity is equal to probability of E and this quantity is equal to  $P$  compliment E.

So, if we substitute  $P_E$  for this quantity and  $P$  compliment for this quantity and then further simplify this, then this equation number 22 will reduce to 0. So, what we have shown that this quantity on the left hand side is less than equal to 0, so what it implies is that  $h$  of  $a$  given  $b$  is less than equal to  $h$  of  $p_e$  plus  $p_e \log$  of  $r$  minus 1. So, this is the inequality we wanted to derive and this important inequality which relates the probability of a channel probability.

(Refer Slide Time: 41:40)

$H(A/B) \leq H(P_E) + P_E \log(r-1)$

FANO'S INEQUALITY

uncertainty of whether an error has been made or not  
 $\hookrightarrow H(P_E)$

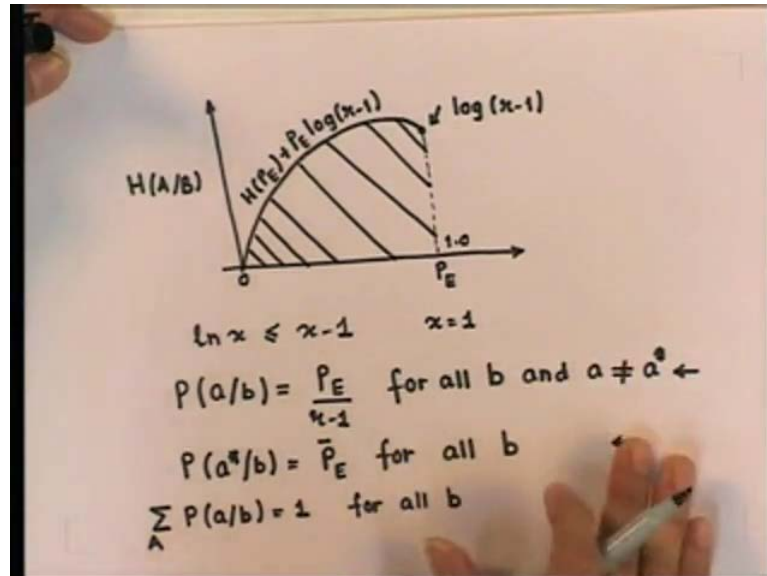
$P_E$   
 uncertainty of which of the  $(r-1)$  remaining symbols has been transmitted  
 $\hookrightarrow P_E \log(r-1)$

This inequality which relates equivocation to the error probability of a channel was first derived by Fano and it is known as Fano's inequality. It is useful for proving the Shannon's second theorem in a very general case. Now, this inequality is valid no matter what decision rule we use, although the error probability can depend drastically upon the decision rule. Now, this inequality has a very interesting interpretation, the average uncertainty about A if B is known, that is this quantity can be regarded as the uncertainty of whether an error has been made or not?

If an error has been made with probability  $P_E$ , then the uncertainty of which of the  $r$  minus 1 remaining symbols has been transmitted. Now, the first amount of uncertainty

that is this quantity is given by  $H$  of  $P E$ , while the second is at most equal to  $\log$  of  $r$  minus 1 waited with the probability of error that is  $P E$ . So, from this inequality it follows that a small average error probability means that equivocation is also small.

(Refer Slide Time: 44:45)



Now, the inequality is depicted graphically as shown here. So  $H$  of  $A$  given  $B$  will be always lying below this curve, this curve is given by  $H$  of  $P E$  plus  $P E \log r$  minus 1 for  $P E$  equal to 1 the value would be  $\log r$  minus 1 and for  $P$  equal to 0  $H$  of  $A$  given  $B$  would be obviously equal to 0. So, this is the relationship we have between error probability and equivocation. Now, let us examine the condition for equality in the Fano bound, which is given by this equation, now the inequality  $\log x$  less than equal to  $x$  minus 1 becomes an equality if and only if  $x$  is equal to 1.

So, if we use this condition in this equation we find that the Fano bound becomes an equality if and only if probability of  $P a$  given  $b$  is equal to  $P E$  divided by  $r$  minus 1 for all  $b$  and  $a$  naught equal to  $a^*$ . Probability of  $a^*$  given  $b$  is equal to  $P$  compliment  $E$  for all  $b$ . So, if this two conditions are satisfied then we would get the equality in the Fano bound. Now, since probability of  $P a$  given  $b$  summed over  $a$  is equal to 1 for all  $b$ , it implies that the second condition, that is this condition follows from the first condition.

So, this equation probability of  $a$  given  $b$  is equal to probability of  $E r$  minus 1 implies that for each  $b$  all input symbols except the only, it implies that for all input symbols except the one selected by a decision rule are equip probable. Now, this condition thus

serves to reinforce our interpretation of the Fano bound. The objective of Shannon's second theorem is to describe the fundamental limitations on the transmission of reliable error free messages through an unreliable channel. In the next class, we will initiate our discussion on the procedure to obtain error free communication over noisy channel.