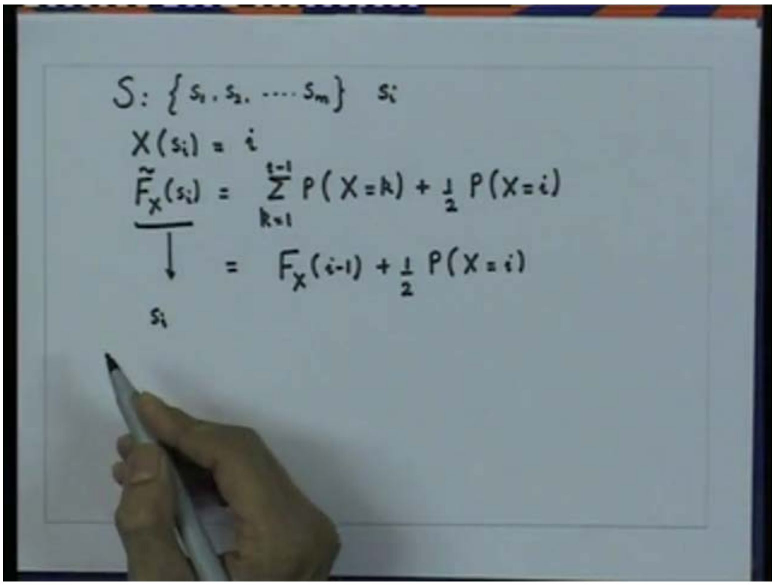
Information Theory and Coding
Prof. S. N. Merchant
Department of Electrical Engineering
Indian Institute of Technology, Bombay

Lecture - 18
Arithmetic Coding Part - I

In the last class, we initiated our study of arithmetic coding. We saw that the interval in which the tag for a particular sequence resides is disjoint from all intervals in which the tag for any other sequence resides. As such, any member of this interval can be used as a tag. One popular choice is to use the lower limit of the interval. Another possibility is to use the midpoint of this interval. For our discussion, we will use the midpoint of this interval. In order to see how the tag generation procedure works mathematically, let us start with sequences of length 1.

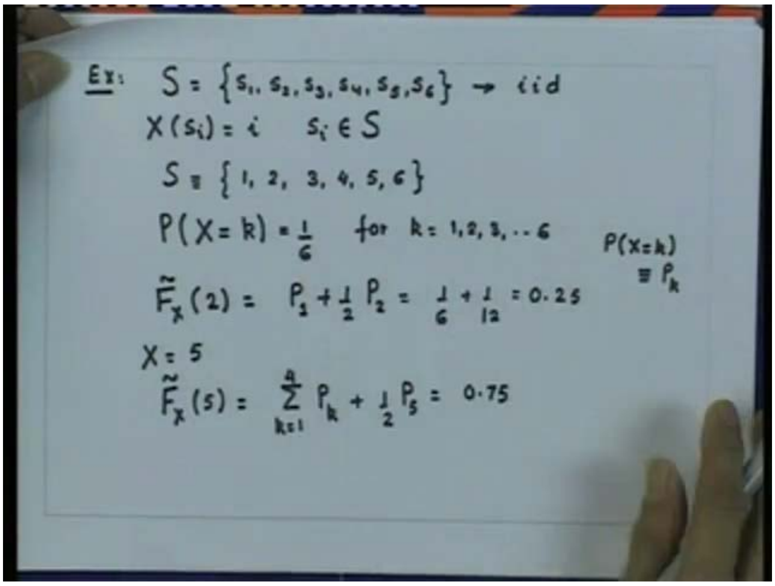
(Refer Slide Time: 01:47)



Suppose, we have a source that puts out some symbols given as s 1, s 2 up to s m. Map those symbols s i to real numbers i. Like in our study pertaining to Shannon Fano Elias coding, we will use the same mapping in this case 2 X maps. The letters of this source alphabet to integers i where s i belongs to source s. We can define a function F X tilde similar to what we did in Shannon Fano Elias coding.

This we called it as a modified p d f as plus half of probability of random variable X is equal to i. This by definition is equal to cumulative probability distribution plus half probability x is equal to i. So, for each s i, this function will have a unique value. This value can be used as a unique tag for s i. So, this value will be used as a unique tag for s i. So, let us consider an example.

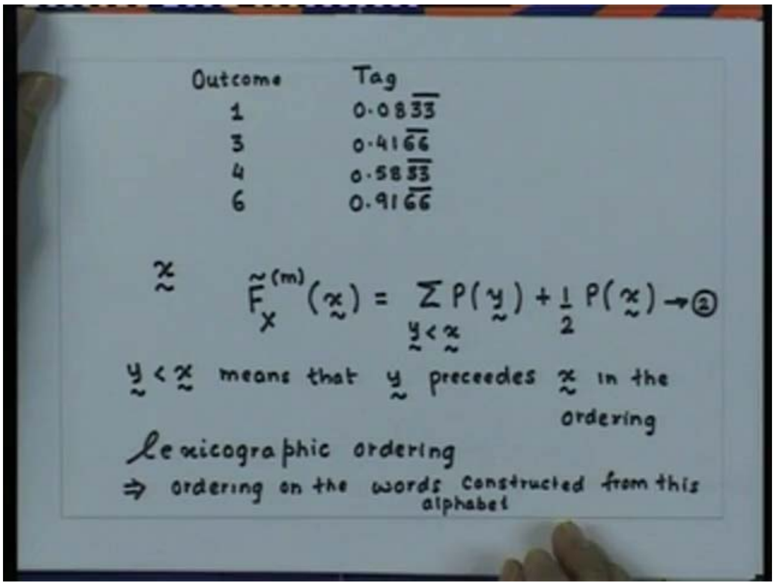
(Refer Slide Time: 04:21)



Consider source with 6 letters. Let us assume that this source is 0, memory i i d source. We use the mapping where i is integer. So, the outcomes of the source are mapped to the numbers 1, 2, 3, 4, 5 and 6. Since, we are assuming i i d probability of random variable X is equal to k where k goes from 1 to 6 is equal to 1 by 6 for k equal to 1 to 6. So, if we use this equation, let us call this equation number 1.

If we use this equation, then we can find out the value of that function for 2 i where I am using the simplified notification probability of X is equal to k is denoted by probability P suffix k. Similarly, that tag for X is equal to random variable 5 can be found out as 0.75. So, using this concept, we can find out the tags for the outcomes also 1, 3, 4, 6. They will be as follows.

(Refer Slide Time: 07:33)

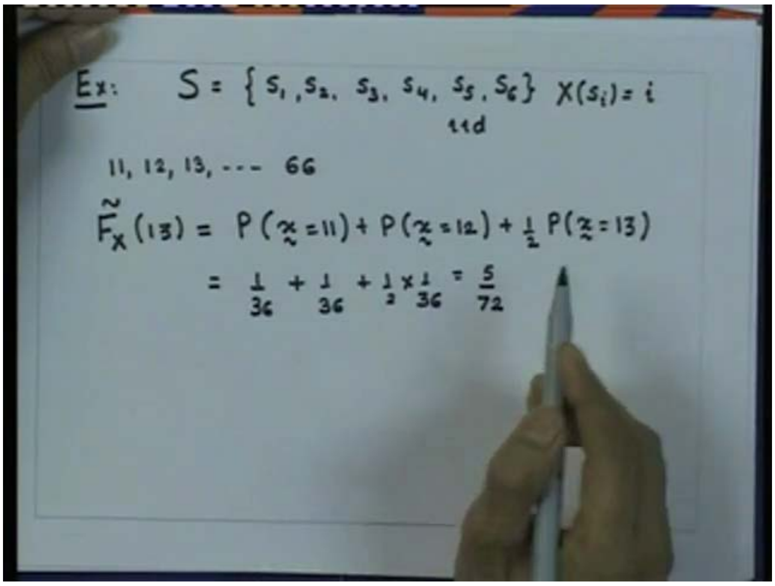


So, outcome and the tag; we have already seen the outcome for 1 and 5. So, the outcome 1, 3, 4, 6 using that equation we can find out the tag values as given here. This means that 3 3 keeps on recovering. This bar indicates 0.5833. Finally, for the outcome 6, the tag value is 0.9166. Now, this approach can be easily extended to longer sequences by imposing some kind of an order on the sequences.

We need ordering on the sequences because we really assigned a tag to a particular sequence. For example, let us denote the particular sequence as x with underscore given like this. This denotes a sequence, and then will generate a tag for this sequence by using some kind of ordering on the sequence as follows is equal to this. We will call this equation number 2.

Here, x is a sequence. m denotes the length of the sequence. This is the tag which we are generating for the sequence x. This equation is similar to what we saw in the earlier case for sequences of length 1, where I have some kind of a cumulative distribution function plus half the probability of that sequence where y is any sequence. Y sequence less than x sequence means that sequence y precedes sequence x in the ordinary. Now, in order to utilise this and easy ordering to use is what is known as lexicographic ordering. In lexicographic ordering, the ordering of letters in a alphabet induces and ordering on the words constructed from this alphabet. The ordering of words in a dictionary is a good example of lexicographic ordering. Dictionary ordering is sometime used as a synonym for lexicographic ordering. So, let us extend an earlier example to the sequences of length 2.

(Refer Slide Time: 12:15)

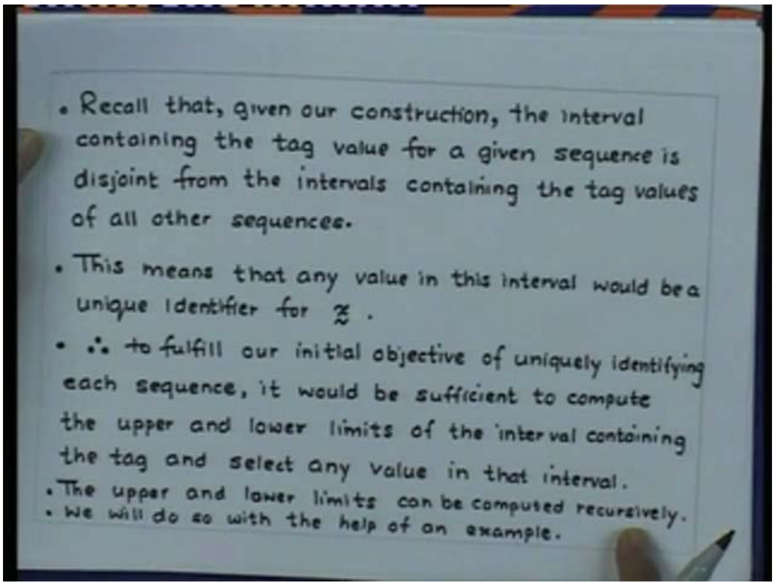


So, let me assume that I have 2 identical sources. Each source consists of 6 letters from s 1 to s 6. We use the same mapping to convert the outcome to real numbers; in this specific case, to be integers. Then I take 2 sources, identical sources. These are both i i d sources. I observed the outcome of these 2 sources as a concatenation of the outcomes of the individual sources.

So, using the ordering scheme described earlier, the outcomes in order would be 1 1, 1 2, 1 3 up to 6 6. So, if I observe both these sources together and look at the concatenation of the outcomes, I will get this is my outcome. Now, the tags for these outcomes can be generated using this equation number 2. For example, the tag for sequence 1 3 will be generated as probability of sequence equal to 1 1 plus probability of sequence 1 2 plus half the probability of the sequence equal to 1 3. So, these are the sequences 1 1 and 1 2. That is nothing but y sequences, which is less than the given sequence x 1 3. So, if I assume that these are i i d sources, then the probability for the sequence equal to 1 1 is 1 by 36.

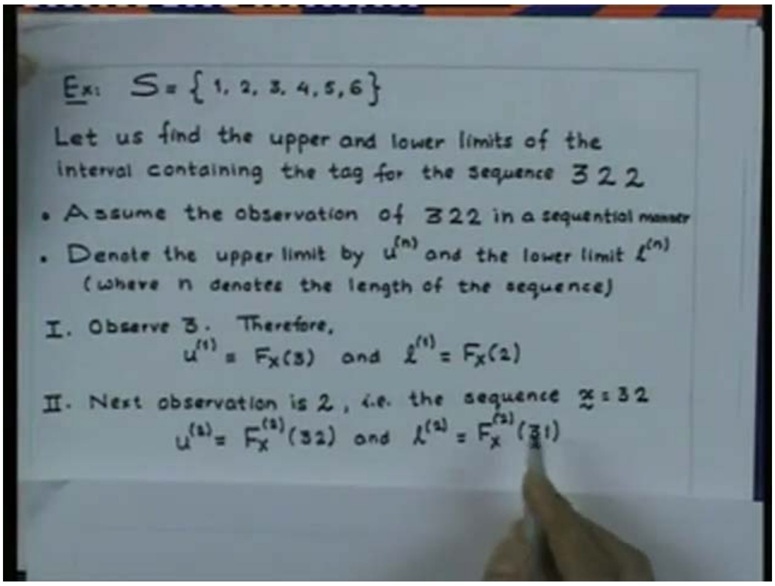
Similarly, for this is 1 by 36 and for this is half multiplied by 1 by 36. This turns out to be 5 by 72. So, notice that to generate the tag for 1 by 1 for the sequence 1 3 not have to generate a tag for every other possible message. However, we need to know the probability of every sequence that is less than the sequence for which the tag is being generated. Now, the requirement that the probability of all sequences of a given length be explicitly calculated can be as prohibitive as the requirement that we have code words for all sequences of a given length. Now, fortunately we will see that to do, so we need only the probability of individual source or the probability model for that source.

(Refer Slide Time: 16:08)



So, let us now recall that given our construction, the interval containing the tag value for a given sequence is disjoint from the intervals containing at tag values of all other sequences. This we have seen what this implies is that any value in this interval would be a unique identifier for the sequence Z. Therefore, to fulfil our initial objective of uniquely identifying each sequence, it would be sufficient to compute the upper and lower limits of the interval containing the tag and select any value in that interval. Now, we see that the upper and the lower limits can be computed recursively. We will try to understand this computation with the help of an example.

(Refer Slide Time: 17:28)

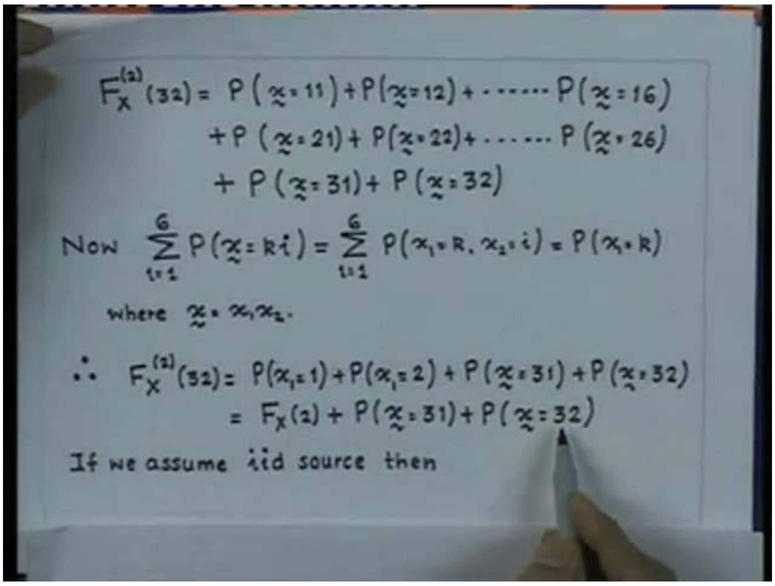


So, let me again consider the earlier source consisting of 6 letters from s 1 to s 6. I use my standard mapping X s i equal to i. So, the outcomes of this source are integers from 1 to 6. What we need is to generate the tags for the sequences of different lengths. So, let us find the upper and lower limits of the interval containing the tag for the sequence 3 2 2.

Let us first assume that we are observing this sequence 3 2 2 in a sequential manner. So, if you do that, then let us denote the upper limit of this interval as u subscribe super superscript n, and the lower limit as l superscript n where n denotes the length of the sequence. So, the first symbol or the integer which we observing the sequence is 3. Therefore, let us generate the tag at the stage. When we observe the symbol corresponding to 3, it is s 3.

So, if we follow the discussion which then u 1 is given by F X 3 and l 1 is given by F X 2. So, 1 denotes the length of the sequence at this stage. So, this is upper limit. This is lower limit after you have observed 3. The next symbol to be observed in the sequence is 2. Therefore, the sequence X becomes 3 2. Now, to generate the upper and the lower limits for the sequence by definition u 2 is F X 2 3 2 and l 2 is F X 2 3 1. Now, we need to calculate this quantity is out here. So, let us calculate these 2 quantities.

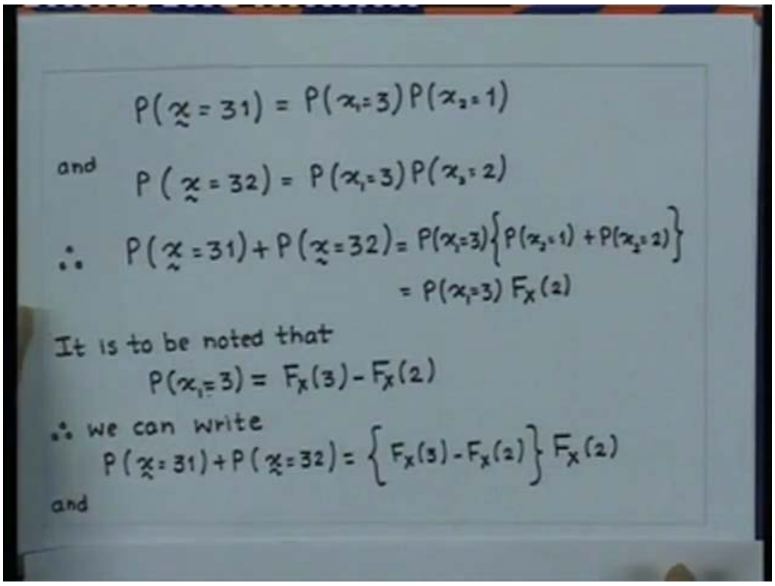
(Refer Slide Time: 20:33)



F X 2 3 2 is nothing but probability of sequences starting from 1 1. So, it is probability of sequence 1 1, 1 2 up to 1 6 plus probability of sequence 2 1 up to 2 6 plus probability of sequence 3 1 and then finally, the probability of the sequence 3 2. Now, if you look at any of this turns out here they are nothing but written as this. Look at this term out here 6 terms and 6 terms out here. This all 6 terms can be put as 1 term shown here.

So, probability of a sequence equal to k i where i goes from 1 to 6 is nothing but probability of x 1 is equal to k and x 2 is equal to i. That is equal to probability of x 1 is equal to k, where sequence x consists of 2 letters x 1 and x 2. So, using this relationship, we calculate F X 2 3 2 as probability of x 1 equal to 1. Probability of x 1 is equal to 2; probability of sequence equal to 3 1 and probability of sequence equal to 3 2. Now, this by definition is a cumulative distribution function for 2 this we have written here.

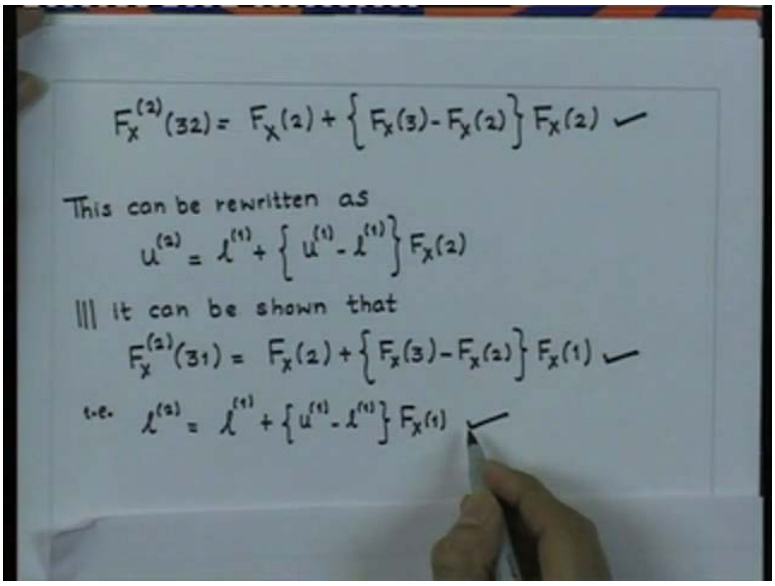
(Refer Slide Time: 22:42)



If we assume i i d sources, then we can calculate probability of the sequence 3 1 and 3 2 as follows. The probability of sequence 3 1 is given by probability of x 1 equal to 3 and probability of x 2 equal to 1 assuming these are i i d. Similarly, probability of sequence 3 2 is given by this relationship. Finally, we can write probability of the sequence 3 1 and probability of sequence 3 2. The summation of these 2 quantities is this. This on simplification turns out to be probability of x 1 equal to 3 and the cumulative distribution function for 2.

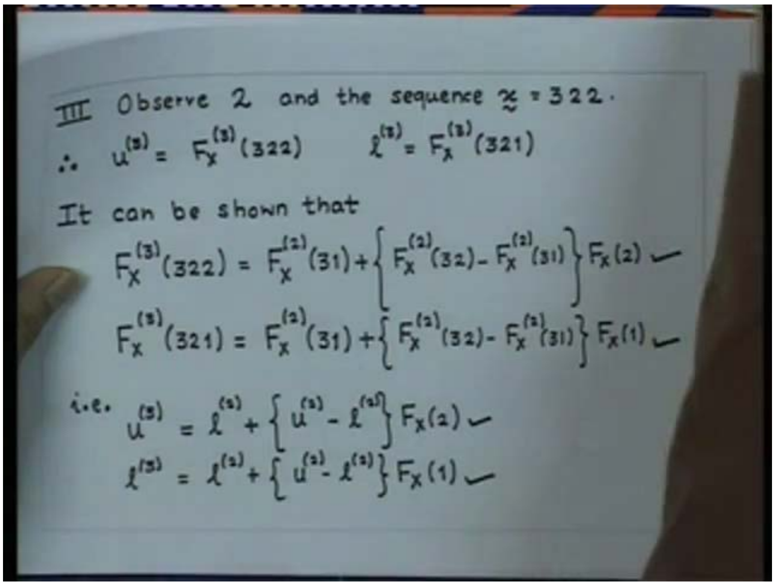
Now, it is to be noted that marginal probabilities that is P of x 1 equal to 3 is nothing but the difference of the cumulative distribution function this by definition of the cumulative distribution function. Therefore, we can write the summation of sequence probability of the sequence 3 1 and probability of sequence 3 2 as shown here. Finally, using these relationships we can write F X 2 3 2 as shown here.

(Refer Slide Time: 24:13)



So, this can be rewritten as because by definition. This is the upper limit for the interval corresponding to the sequence of length 2. So, this quantity is replaced by the u 2. F X 2 is basically the lower limit of the sequence of length 1. F X 3 is the upper limit of the sequence of length 1. Again, it is a lower limit and F X 2. Similarly, based on these arguments, we can find out F X 2 3 1. It is easy to show that this expression will be opting. Now, again this is nothing but the lower limit for the sequence of length 2. This is again lower limit of sequence 1. So, if we substitute, we get this relationship. So, at the end of observing 3 2, the upper limits for the interval where the tag for this sequence will reside are given by this and this quantity u 2 and l 2. Now, the next letter in the sequence to be observed is 2. Then the sequence becomes as 3 2 2.

(Refer Slide Time: 26:07)

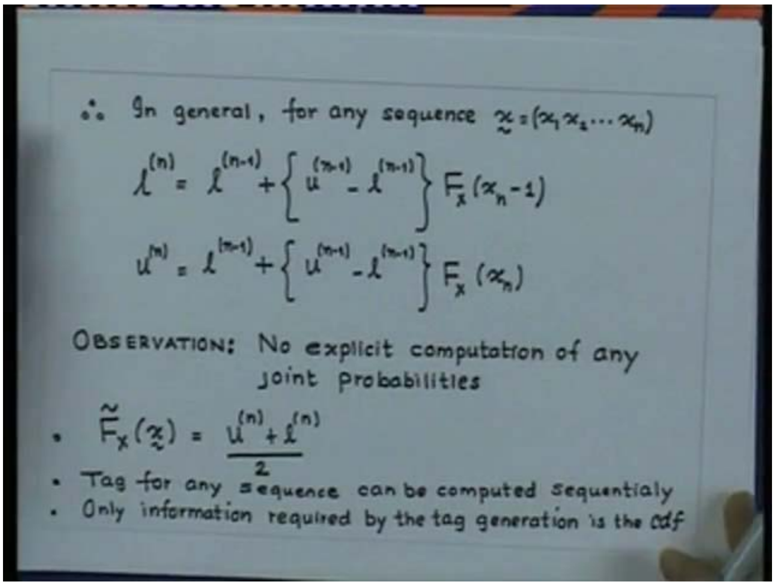


So, by definition, we are supposed to calculate the upper limit for the tag in which the tag for the sequence; the interval for the tag for the sequence. So, that will be given by as u 3. That is by definition. It is nothing but F X 3 3 2 2 and l 3 is F X 3 3 2 1. Now, it can be shown that F X 3 2 3 2 2 is nothing but given by this expression. It is not very difficult to see the relationship between this and what we saw earlier here. If you look at this relationship, if you observe this, the first term out here is the cumulative distribution function obtained for the sequence earlier to 2 that is 3 minus 1.

So, 3 minus 1 is 2. So, that is cumulative distribution function for this. This is cumulative distribution function for the sequence here. Again, the cumulative distribution functions for the sequence earlier to 2, but minus 1. So, based on this concept, we can see here that the sequence proceeding to 2 is 3 2. So, 3 2 minus 1 is 3 1. So, this is the quantity which will get and out here. You will get cumulative distribution function for the sequence which we obtain before 2. That is 3 2.

Then, again this is the same value out here. Similarly, we can obtain the lower limit for the sequence 3 2 1. That is again given by this relationship. Now, converting these values to the limits, we get these 2 relationships. Now, if you observed these relationships, we can see that in general for any sequence based on this, we can write the lower limits and upper limits as shown here.

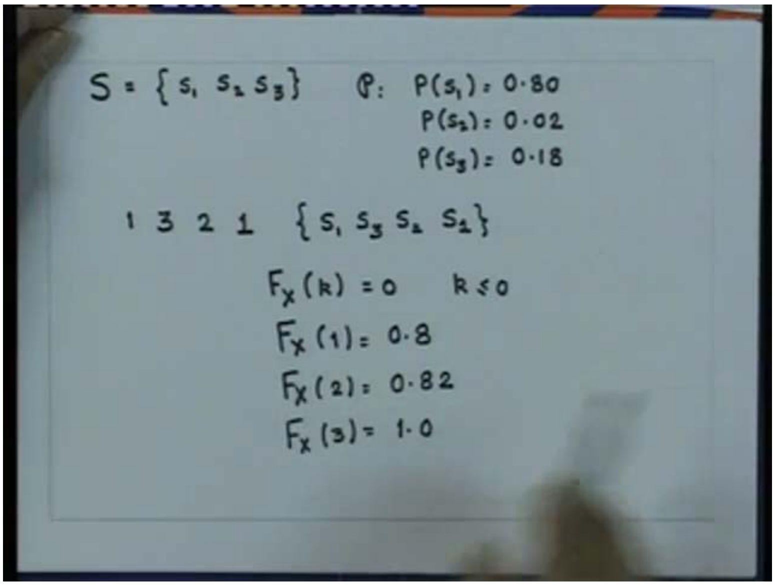
(Refer Slide Time: 28:53)



So, in general for any sequence which consists of n symbols, it will be given as l n is equal to the lower limit obtained in the sequence of length n minus 1; the difference of the upper limit and the lower limit for the sequence n minus 1 and cumulative distribution function for the current symbol minus 1. So, F X x n minus 1 and then for the upper limit, it is the cumulative distribution function for the symbol x n.

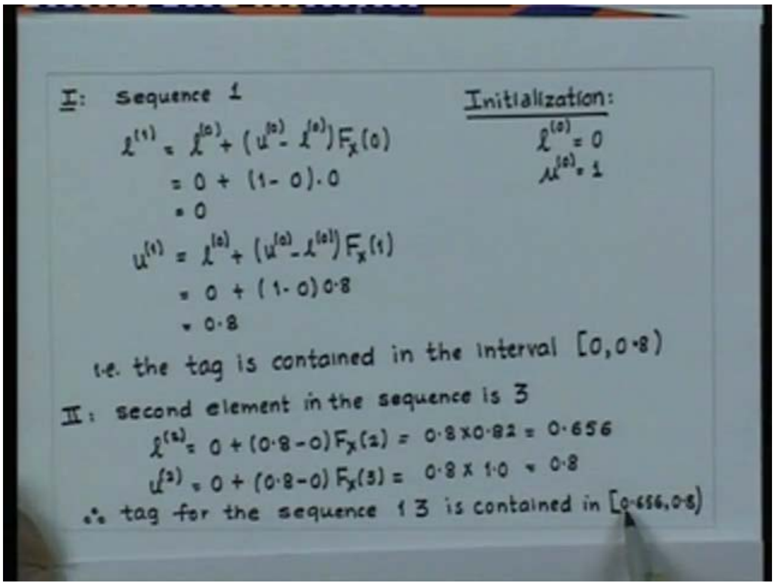
So, if you look at this procedure, we see that there is no explicit computation of any joint probabilities. Now, to generate that tag for that particular sequence as we have said earlier, we can use the midpoint of the upper limit and the lower limits. So, if we use the midpoint of the upper limit on the lower limit, get this relationship. So, this will form the tag for the sequence x. At this stage, sequence x consists of n symbols or n letters. So, tag for any sequence can be computed sequentially. Only information required by the tag generation is the cumulative distribution function. So, let us look at this tag generation procedure with a specific source.

(Refer Slide Time: 30:45)



So, let me assume that I have a source S consisting of 3 symbols with the probability model given as probability of s 1 is equal to 0.80, probability of s 2 is equal to 0.02 and finally, probability of s 3 is equal to 0.18. Now, we would like to encode the sequence 1 3 2 1. So, indirectly we are encoding the sequence s 1 s 3 s 2 s 1 because we are using a mapping x of s i is equal to i. So, from this probability model, we can calculate the cumulative distribution function as follows. This is equal to 0 for k less than equal to 0. F X 1 is equal to 0.8. F X 2 is equal to 0.82. Finally, F X 3 is equal to 1.0. So, the first symbol in the sequence is 1.

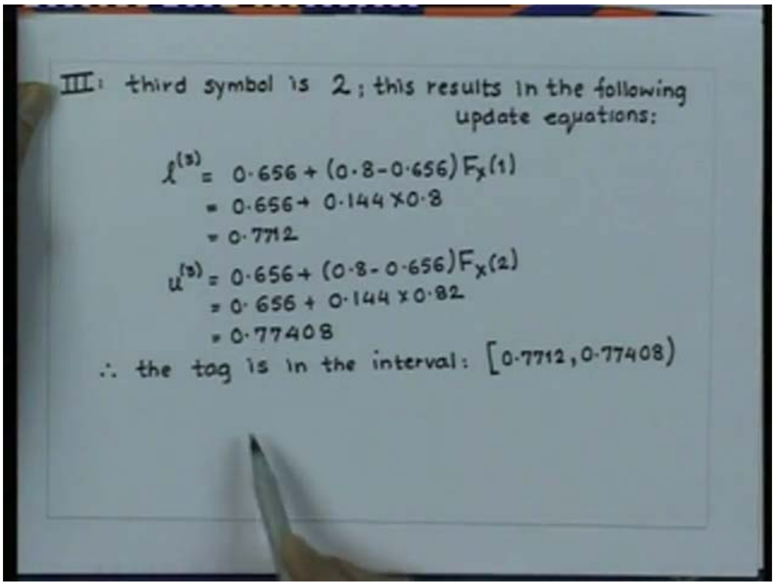
(Refer Slide Time: 32:28)



So, let us generate the upper limits and the lower limit for the sequence. At the state, we will initialise l 0 equal to 0 and u 0 equal to 1. So, l 1 will be given by this relationship, which we have derived earlier l 0 plus differences and F X of x n minus 1. In this case, x n is 1. Therefore, 1 minus 0 is 0. So, I required probability distribute cumulative distribution function for 0. So, if I substitute values out here, I get l 1 is equal to 0. Similarly, for u 1, I can substitute the values I get here. This is of F of X x n. x n in this case, x 1. That is 1. Therefore, it is F X 1. So, if I substitute values, I get u 1 equal to 0.8.

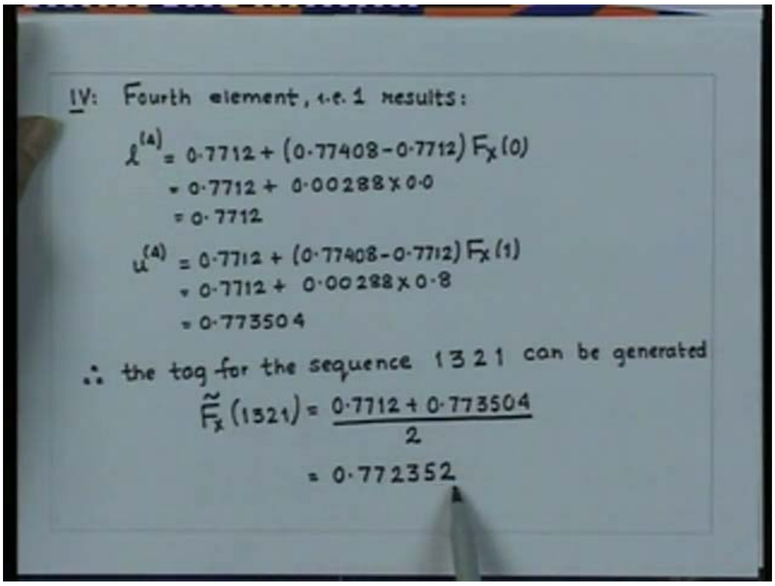
Now, the tag for sequence after observing 1 is contained in the interval between 0 and 0.8. The second element in the sequence is 3. So, u l 2 and u 2 will be given as follows. l 2 is equal to l 1 plus u 1 minus l 1 and F X of x n minus 1. x n in this case is 3. So, 3 minus 1 is 2. Therefore, we get F X 2. If you substitute the values, you get 0.656. Similarly, u 2 will be given as shown here. In this case, x n is 3. So, I write as cumulative distribution function. For 3, I get 0.8. So, at the end of observing 1 and 3, the tag for the sequence 1 3 is contained in 0.65 to 0.8. So, if we progress in this manner, the third symbol is 2.

(Refer Slide Time: 34:57)



This results in the following update equation. You get l 3. This is your l 2. This is your u 2. Similarly, u 3 is given by this. So, the tag lies in the interval given by this. Finally, we have the fourth element that is 1. If we take that into consideration finally, we get the upper limits and the lower limits as shown here.

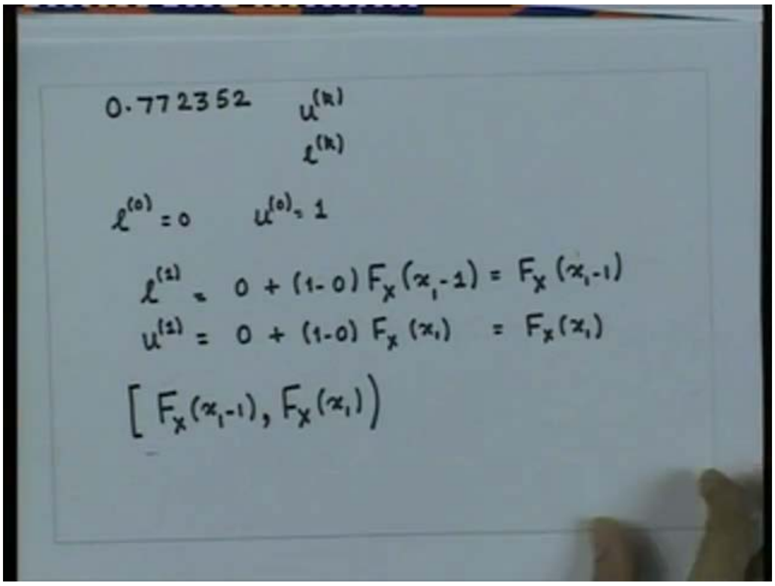
(Refer Slide Time: 35:33)



So, the lower limit is given by 0.7712. The upper limit is given by 0.773504. It is important to note again here. This becomes F X 0, because this is 1 x n is equal to 1. That is x 4 equal to 1. Therefore, you get this limit l 4 as the same as l 3. So, the tag for the final sequence 1 3 2 1 can be generated as the midpoint of l 4 u 4. That is 0.772352. Now, from this, we observed that each successive succeeding interval is contained in the preceding interval. Now, this property we will use to decipher the tag. The only undesirable consequence of this process is that the intervals get smaller and smaller and require higher precision. As a result, the sequence gets longer.

So, in order to overcome this difficulty in a practical scenario, rescaling strategy is adopted. So, we have looked at a procedure for generation of the tag. Now, let us look at the procedure to decipher the tag that is decode the tag. Now, we have seen that a sequence can be assigned a unique tag given a minimal amount of information. Now, deciphering or decoding the tag is as simple as generating it. We will study through an example. So, let us consider tag obtained in the previous example, which we use to explain the generation of the tag.

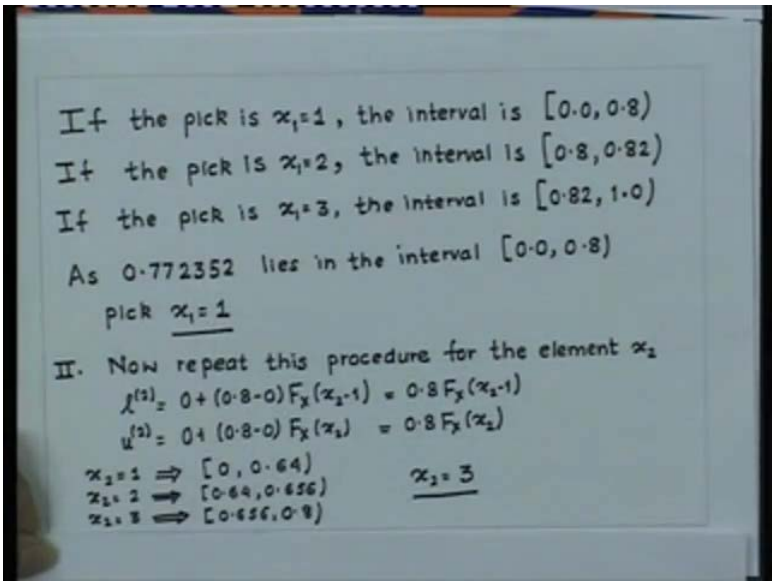
(Refer Slide Time: 37:54)



So, the tag value which we generated was given by 0.772352. The interval containing this tag value is a subset of every interval obtained in the encoding procedure. Now, our decoding strategy will be to decode elements in the sequence in such a way that the upper and the lower limits will always contain the tag value for each k. So, we will start with l 0 equal to 0 and u 0 equal to 1 after decoding the first element of the sequence x 1.

The upper and lower limits become l 1 is equal to 0 plus 1 minus 0 x 1 minus 1. That is equal to F of X 1 minus 1. The upper limit after decoding of the first element will be the F X x 1. So, in other words, the interval containing the tag is after decoding of the first element will be given by this interval. So, we need to find the value of x 1 for which 0.772352 lies in the interval.

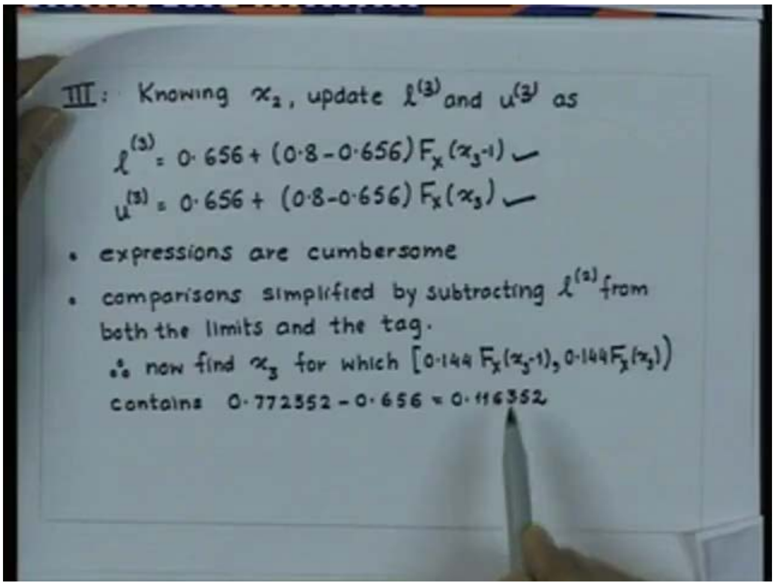
(Refer Slide Time: 40:30)



Now, if we pick, x 1 is equal to 1. The interval turns out to be 0.0 to 0.8. If you assume x 1 is equal to 2, the interval is as shown here. Finally, for x 1 equal to 3, the interval is 0.82 to 1.0. Now, we want that our tag 0.772352 should lie in the interval 1 of this interval. From here, it is very obvious that this lies in this interval. Therefore, for the first element x 1, we pick x 1 as 1. Now, we repeat this procedure for the second element in the sequence. Let us call it as x 2. Now, the upper limit and the lower limits will change as l 2 is equal to because the lower limit is 0 and upper limit is 0.8. I calculate this. I get this value. Similarly, for the upper limit, I get this value.

Now, again I have to choose my values of x 2 in such a way that might tag lies in an appropriate interval. Now, if I choose my x 2 is equal to 1, I get this interval. If I choose x 2 is equal to 2, we get this interval. If I choose my x 2 is equal to 3, I get this interval. These intervals are obtained by substitute the values next to in these equations. Now, if I choose my x 2 is equal to 3, then I get this interval. My tag lies in this interval. So, I choose my x 2 is equal to 3. So, knowing x 2, I can update my l 2 and u 2 as follows.

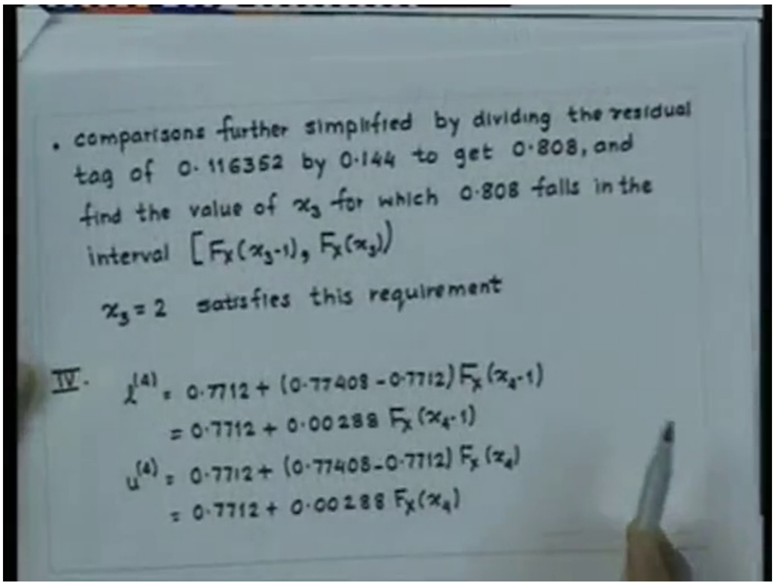
(Refer Slide Time: 42:33)



So, my l 3 and u 3 will be as 0.65. This should be l 3 and u 3 as given by this expression. Now, again I have to choose that value of x 3 for which my original tag of 0.772352 lies in this. Now, if you look at this expression, these are little bit cumbersome. So, if you want to simplify the comparison, we can subtract l 2 from both the limits. I subtract 0.65 from both the limits.

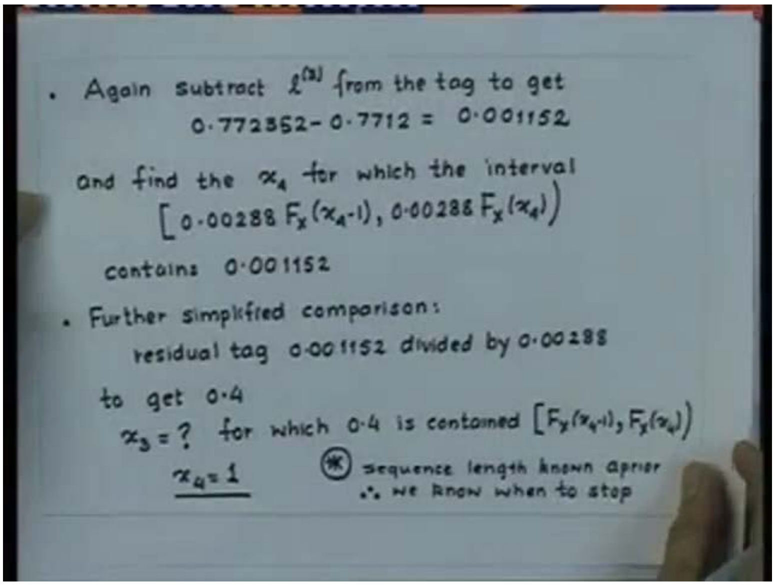
Then, you find x 3 for which it contains 0.772352 minus 0.656 because I separated from tag. Also, I have to find x 3 such that this value lies in that interval. Now, again if you look at this comparison, it is also little bit cumbersome. So, the comparisons can be further simplified by dividing the residual tag of 0.116352 by 0.144. So, if I do that, if I divide this number by 0.144, I get 0.808.

(Refer Slide Time: 44:06)



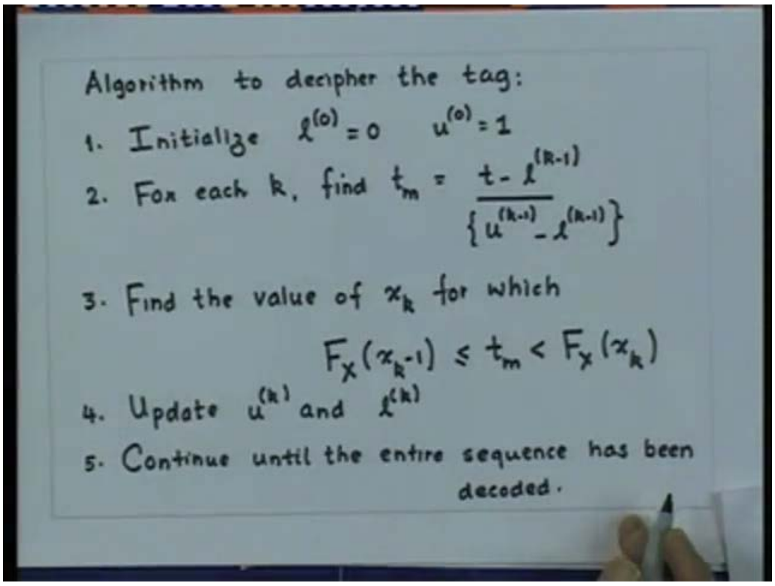
So, after comparisons for the simplified by dividing the residual tag of 0.116352 by 0.144, we get finally, 0.808. Finally, find a value of x 3 for which 0.808 falls in the interval given by this. Again, if you substitute x 3 is equal to 2, this satisfies the requirement. Finally, I get my l 4 given by this expression. Now, again I have to find out x 4 such that it lies in this interval. Again, I can subtract out this l 3 from both the thing and then divide by 0.00288. So, if I do this, what I get finally is this.

(Refer Slide Time: 45:05)



So, I subtract l 3 from the tag to get this. I also subtract it from my l 4 and u 4. Then find x 4 for which the interval is this further simplification is by dividing 0.001152 by 0.00288. So, I get 0.4. So, I have to find out x 3 for which 0.4 is contained in this. If you do that, you find x 4 is equal to 1. So, in short, we can summarise the algorithm to decipher the tag as follows.

(Refer Slide Time: 45:47)



So, the algorithm to decipher the tag is as follows. First initialise l 0 equal to 0 and u 0 equal to 1. For each k, find the modified tag value as t minus l k minus 1. Then find the value of x k for which F X x k minus 1 is less than equal to modify tag value. After having done this, update u k and l k. So, continue until the entire sequence has been decoded.

Now, we have seen the procedure for the generation of the tag and also how to decipher the tag. The next logical step is basically to represent this tag by binary representation if you are using the binary alphabet. Now, in the next class, we will look at the binary representation of the tag that forms a unique binary code for the sequence. We will also look at the uniqueness and efficiency of the arithmetic code.