Stochastic Hydrology

Prof. P.P. Mujumdar

Department of Civil Engineering

Indian Institute of Science, Bangalore

Module No. # 06

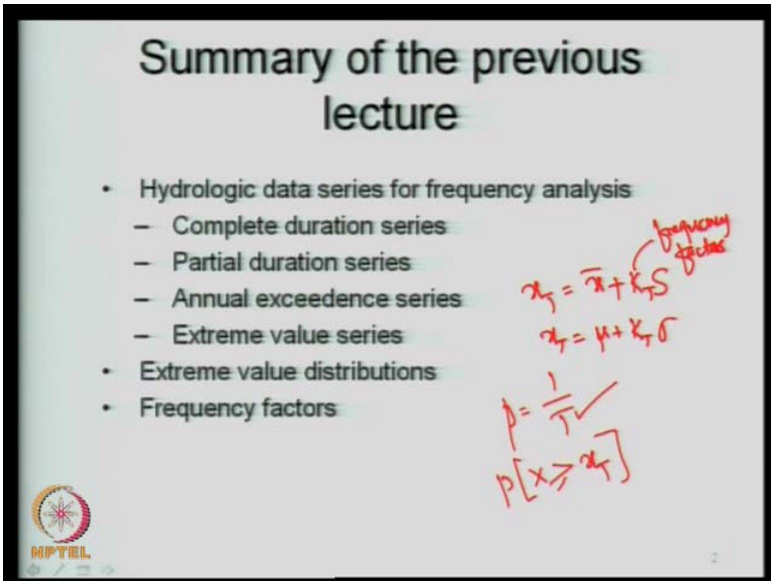
Lecture No. # 26

Frequency Analysis – III and Probability Plotting – I

Good morning and welcome to this the lecture number 26 of the course Stochastic Hydrology. We have been discussing now, the Frequency Analysis especially with related with as related to flood frequencies. Although, the methods that I am discussing in these classes can as well be used for drought frequency analysis but, the examples that we are dealing with in these lectures, are mostly related to the flood frequency.

So, if you recall in the last lecture, I first introduced the type of data, that we use for the hydrologic frequency analysis.

(Refer Slide Time: 00:57)



We may use the complete duration series, where the entire available data will be used for the frequency analysis. Let us say you have monthly data for the last 40 years, all the data will be used for the frequency analysis, in such a case the series, that is used is called as a complete duration series.

Then, we have the concept of the partial duration series, where you put a threshold value and then pick up only those values, which cross the threshold. Let say you put a threshold of 2000 cubic meters per second. So, whenever the discharge crosses 2000 cubic meters per second, we collect those values and then constitute a series and use only such series for frequency analysis, that constitutes a partial duration series.

In the annual exceedence series, we put this threshold such that, the number of values that you get will be equal to the number of years that you have, the number of years for which you have the data. And in the extreme value series, we just pick up the extreme values, for example, annual maximum values we pick up for every year, if you have 40 years of data every year you pick up the maximum value of the discharge, if you are doing it for discharges stream flow discharges; and then constitute the series of maximum values.

Most commonly, we use the extreme value series and in some cases we also use annual exceedence series. Although other methods are also not very uncommon. Then, we introduced extreme value distributions, which I had discussed earlier in some lectures but, we just recapitulated the extreme value type one, type two and type three distributions. If you recall extreme value, type one is also called as a gumbel distribution and then extreme value type three is also called as the wible distribution. The gumbel distribution can be used both for maximum values as well as the minimum values whereas, the wible distribution is typically used mostly for the minimum values.

Gumbel distribution is typically used for maximum values, although it can also be used for minimum values. Then, we introduced the concept of the frequency factors, remember in frequency analysis, what is it that we are interested in; we are talking about the magnitude of the particular hydrologic variable, which equals or exceeds a given magnitude. Let us say we are talking about the probability of x being greater than or equal to x t, where t is the return period.

So, typically we are interested in that particular magnitude, flood magnitude typically, which will be exceeded once in about 20 years, once in 50 years, once in 100 years etcetera, so this is the concept of the return period. And we recall that, the return period determines the probability, so this we have shown earlier probability is equal to 1 by t, where probabilities probability of x being greater than or equal to x T and x T is that particular magnitude.

So, typically what we do is, we fix this return period T let say 20 years, 50 years, 100 years, 1000 years, etcetera. And then ask the question, what is that value of x T for corresponding to this particular T. Now these are important for our design recedence, that is what is the magnitude of flood that you need to design for, if you are planning your design for 100 year flood, 100 year flood 1000 year flood and so on.

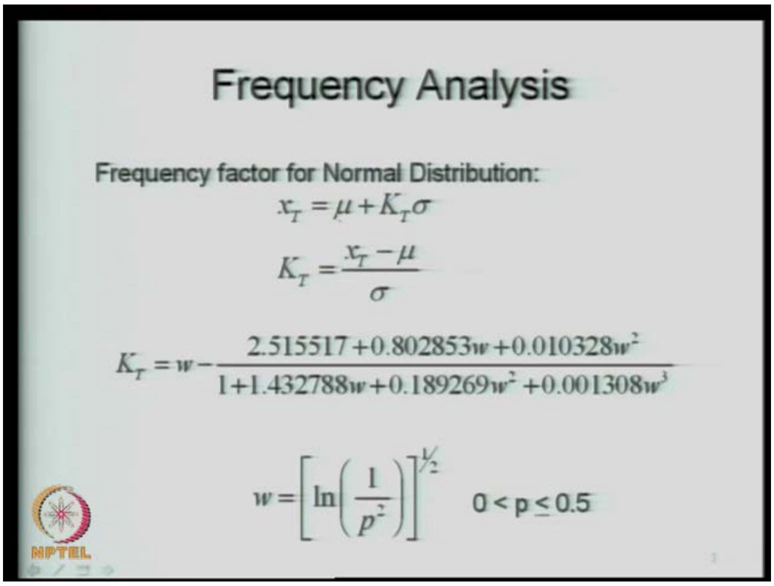
Now, this can be obtained for some distributions like exponential distribution, etcetera by analytical methods, directly you fix your T and therefore, you are fixing probability of x being greater than or equal to x T. And then, analytically you can solve for using the expressions for the cumulative distribution function for that particular distribution; and then solve for x T. However, there are many distributions, which are not invertible in the sense that, you cannot get x T directly from your p d f.

In such situations, we use the concept of frequency factors. So, in the frequency factors you recall that the basic idea is that you are writing x T as x bar plus k T times sigma or S if you are writing for sample; or more generally you write it as for the population mu plus k T times sigma. This is S here, so if you are writing for the sample, you write k T times S and these are called as the frequency factors.

Now, frequency factors for most distributions, that we use in hydrology have been specified by earlier studies and these are available in some empirical relationship form for different distribution. For example, normal distribution, log normal distribution, log pearson type three distribution, extreme value type one distribution, etcetera. We have the k T factors, which are the frequency factors and then we use that to obtain your x T.

Finally, our aim is to obtain x T for a given return period T, given a particular distribution. Now, we will continue the discussion last towards the end of the last lecture, I introduced the frequency factor for normal distribution; initially we also discuss for discussed the analytical method for the exponential distribution but, we now move on to the frequency factors, starting with the normal distribution.

(Refer Slide Time: 07:22)

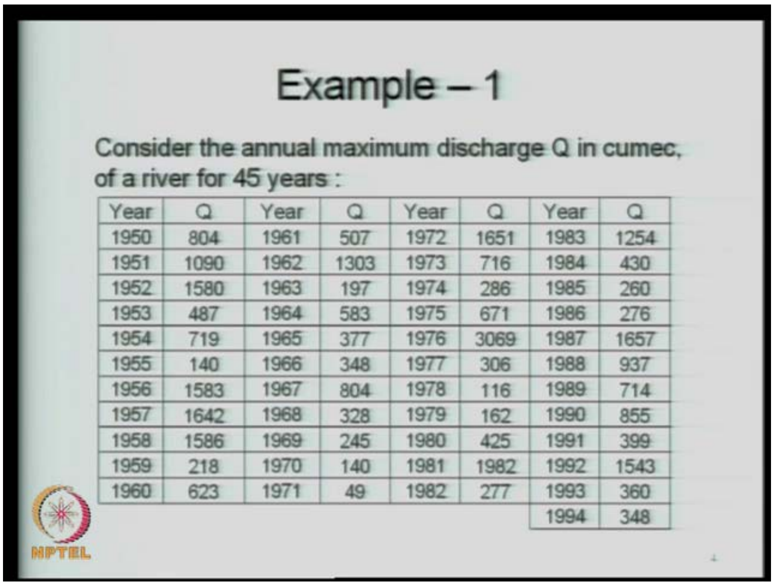


So, we use the expression x T is equal to mu plus K T into sigma, this is our fundamental equation for frequency analysis. So, K T is in general x T minus mu over sigma from this. You can identify that this is in fact, the standard normal deviate z is equal to x T minus mu over sigma, so this expression for normal distribution turns out to be simply the standard normal deviate z. And therefore, you can go to the tables and then obtain the associated value associated value of K T. That is, the tables of standard normal deviate the z tables; corresponding to a particular T, because you know p is equal to 1 by T which is probability of x being greater than or equal to x T is 1 by T; from that, you can convert that into z, go to the z table obtain the associated z value that itself becomes k T.

However, there is a expression available for the z value itself for the normal distribution, that expression we introduced towards the end of the last lecture that is z, which in the in this case also equal to k t will be a function of w here, like this, where w is log 1 over p square, the whole to the power half. And this is valid for p less than or equal to 0.5 and p is our probability of x being greater than or equal to x T, which is given by 1 by T, so this is 1 by T.

Now, if p is greater than 0.5, we use 1 minus p here in this case and obtain k T using 1 minus p here and that k T we use it to the negative sign here. So, if p is greater than 0.5 use 1 minus p here, get the k T corresponding to that and use mu minus k T into sigma with k T as obtained here. However, the cases of p being greater than 0.5 are very rare in most hydrologic applications, so most of the cases this expression works.

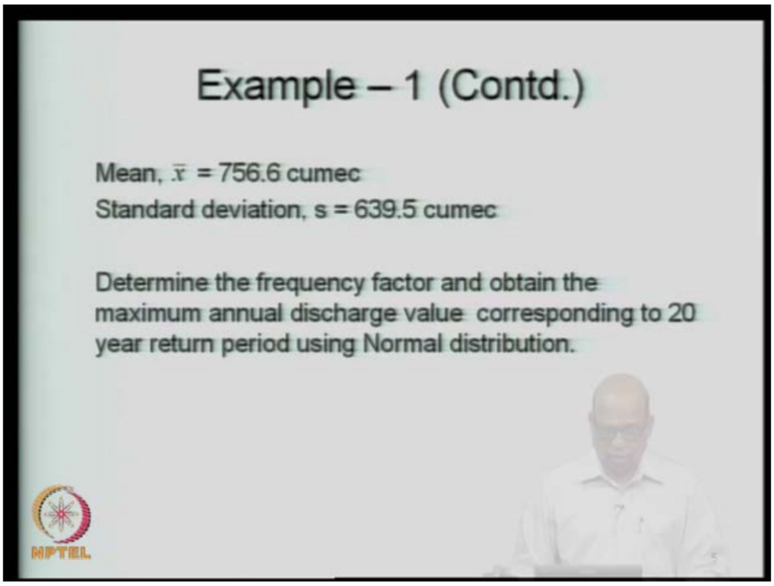
(Refer Slide Time: 09:52)



Year	Q	Year	Q	Year	Q	Year	Q
1950	804	1961	507	1972	1651	1983	1254
1951	1090	1962	1303	1973	716	1984	430
1952	1580	1963	197	1974	286	1985	260
1953	487	1964	583	1975	671	1986	276
1954	719	1965	377	1976	3069	1987	1657
1955	140	1966	348	1977	306	1988	937
1956	1583	1967	804	1978	116	1989	714
1957	1642	1968	328	1979	162	1990	855
1958	1586	1969	245	1980	425	1991	399
1959	218	1970	140	1981	1982	1992	1543
1960	623	1971	49	1982	277	1993	360
						1994	348

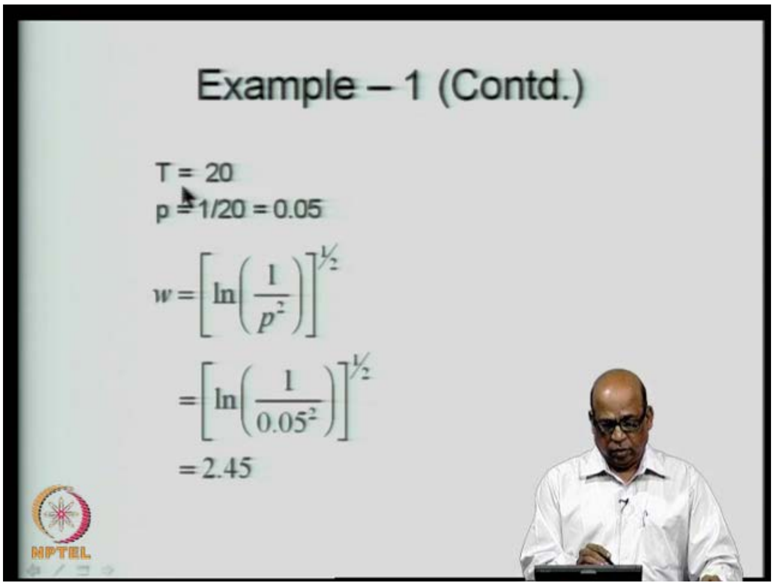
So, let us look at a simple example, we will take the same data as we considered in the last lecture, this is the data we are talking about. Now this is the maximum discharge which means, what every year among all the values that you have observed, you pick up one value which is a maximum among all those values. And therefore, every year you have exactly one value, so this is form 1950 to 1994, which is a data for 45 years.

(Refer Slide Time: 10:30)



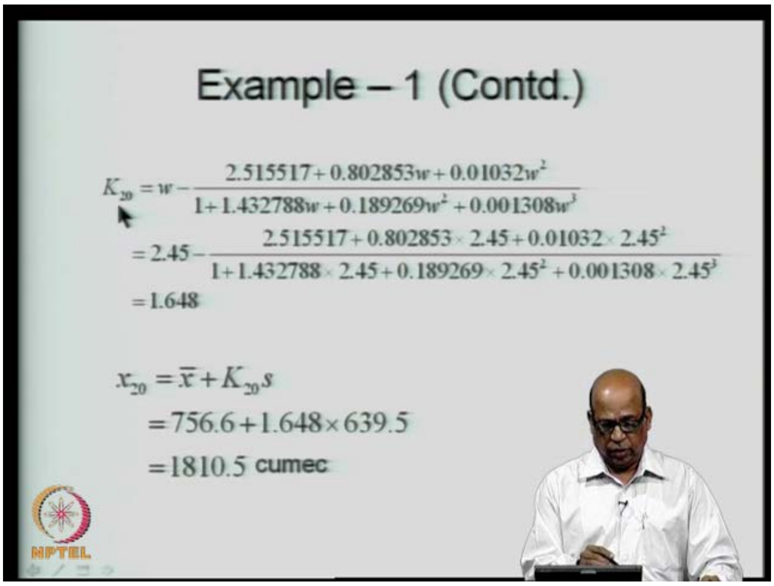
Now, assuming that this follows normal distribution, let us try to get the maximum annual discharge value corresponding to 20 year return period using normal distribution. Now, the mean of this values is 756.6 cumecs, standard deviation is 639.5 cumecs.

(Refer Slide Time: 10:55)



Because, T is given to be 20 years what do we do first calculate your p, so T is 20 therefore, P will be 1 by T which is equal to 0.05. And then calculate w which is log 1 over p square, the whole to the power half and that comes to 2.45.

(Refer Slide Time: 11:15)

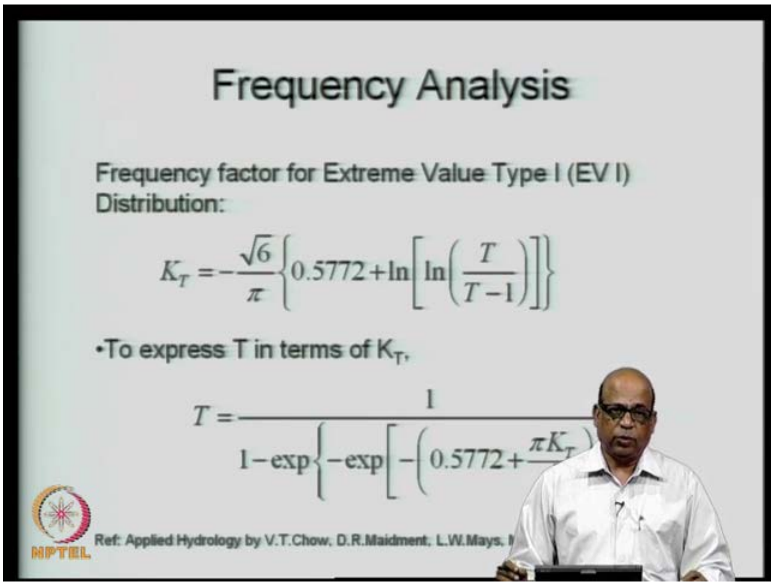


Once you calculate w put it in the expression for K T. T being 20 I write it as k 20 and then you put w value of 2.45 here, and then you get k 20 as 1.648; this is a frequency factor you can just verify also using (Refer Slide Time: 11:34) this particular probability, which is in fact, probability of x being greater than or equal to x T and convert that into a c d f form of z, namely a z being less than or equal to z, probability of z being less than or equal to z go to the z tables get that particular z value, you will still get around 1.65.

Once you get the K T you can get X T in this case x 20 is equal to x bar plus k 20 into s, which comes out to be 1810.5 cubic meters per second. This means that, do not lose the physical picture here, what we are looking at is given the historical data. We are now trying to get the particular magnitude of the flow or the peak discharge, which may be expected to occur once in about 20 years.

So, once in about 20 years, we may expect a discharge of 1,810 cubic meters approximately to be equaled or exceeded that is the idea, so this concept you should not forget, that is about the normal distribution.

(Refer Slide Time: 12:53)

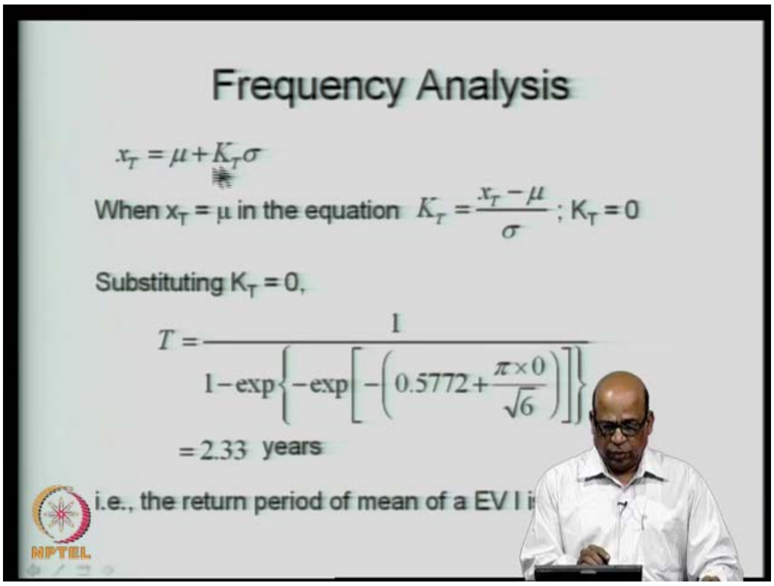


Let us look at the extreme value type one distribution, as I said the frequency factors have been worked out earlier in research papers and most of them are due to V T Chow, Ven Te Chow all students of hydrology are familiar with this particular name, Ven Te Chow and then I am referring to this particular text book by Ven Te Chow maidment and mays.

For the extreme value type one distribution the K T factor or the frequency factor is given by this expression, pi is constant, T is the return period. So, you can get for a given return period you can get K T it will sometimes advantageous also to express T in terms of K T you will presently see what is advantage. So, t is equal to 1 divided by this expression, so if you know K T you can get T.

So, fixing the T let us say you are talking about 20 year return period, 50 year return period, etcetera once you fix T you will be able to get K T and that K T we will use in the expression for obtaining your X T.

(Refer Slide Time: 14:11)



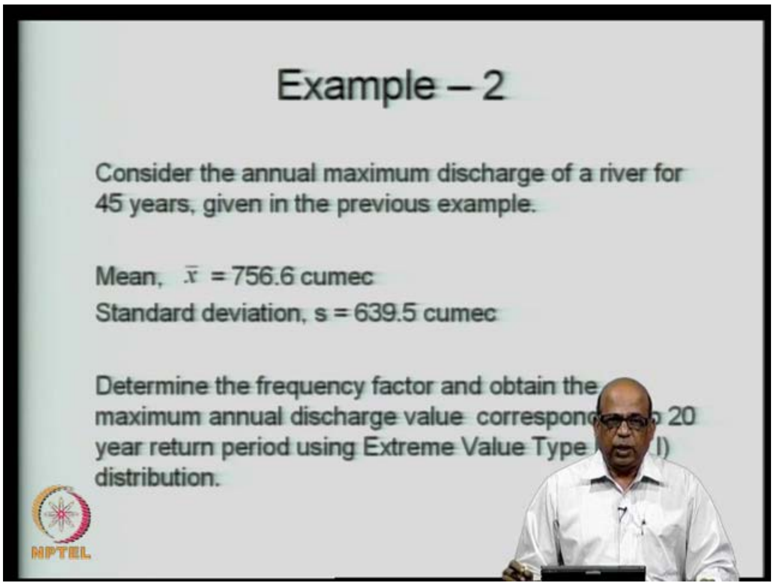
So, look at this expression X T is equal to mu plus K T into sigma this is a general expression for any frequency analysis, not necessarily only for normal distribution remember this.

So, when X T is equal to mu in this expression, you can get K T is equal to X T minus mu or sigma from the here and therefore, when you put X T is equal to mu you get a K T is equal to 0. So, if you are dealing with extreme value type one distribution and then your X T is equal to mu, what does it mean you are saying that the average value or the mean value of the particular sample, that you have or the population that you have, what is the return period corresponding to the mean value is the question that you are trying to answer. So, X T will be equal to mu because that is a magnitude that is equaled or exceeded.

So, we are seeking the answer to the we are seeking the answer to the question of the return period associated with the mean value of an extreme value type one distribution. So, once K T is equal to 0, we put this in this expression put K T is equal to 0 here and then calculate T. So, I put K T is equal to 0 and then I get T is equal to 2.33 years, this is an interesting result in the sense that the return period of a mean value of the mean value of a extreme value type one distribution is about 2.33 years.

Which means what, you have the sample data and then you are reasonably sure that it fits the extreme value type one distribution then straight away you know, that the mean value of that is likely to occur once in about 2.33 years.

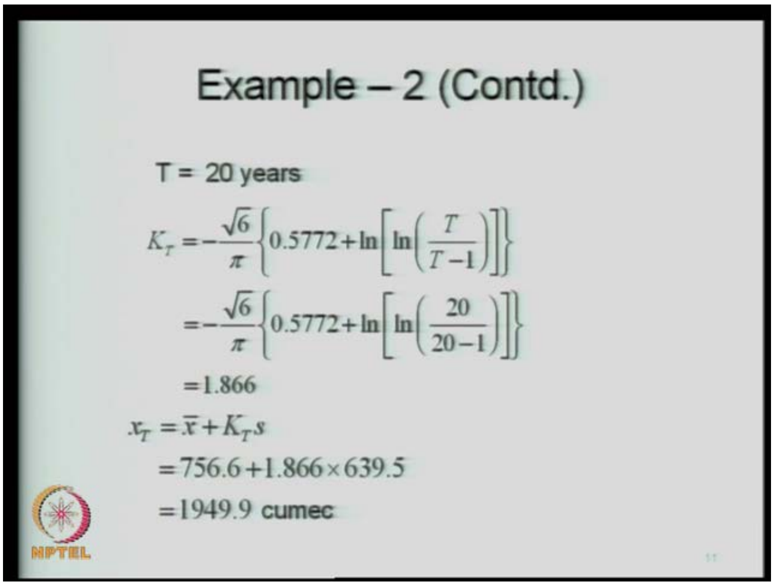
(Refer Slide Time: 15:59)



Then we will try to do the same example, we will do the same example as we did for the normal distribution use the same data and then use extreme value type one distribution and then see what is the magnitude that we get.

So, the 45 years data that we considered in the previous example, the mean of that is 756.6 cubic meters per second, standard deviation is 639.5 cubic meters per second. We will now use the extreme value type one distribution, still determine the 20 year return period value associated with the extreme value type one distribution.

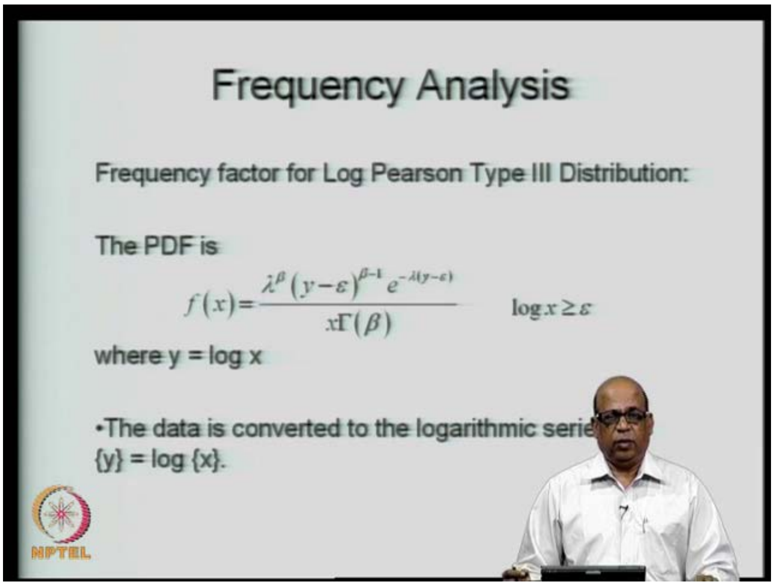
(Refer Slide Time: 16:37)



So, T is equal to 20 years this is straight forward substitution of the values but, let us go through the exercise because we want to compare what different distribution yield. So, K T I simply substitute T is equal to 20 years here and then I obtain K T is equal to 1.866 which is higher than what I got for the normal distribution 1.648 or something.

And therefore, the X T or the 20 year return period annual discharge, maximum annual discharge will correspond to will be equal to 1949.9 cubic meters per second or around 1950 cubic meters per second for the data, that we considered this is higher than the normal distribution (Refer Slide Time: 17:35). Where we got if you look back for the normal distribution, you got somewhere around 1800 cubic meters per second.

(Refer Slide Time: 17:43)

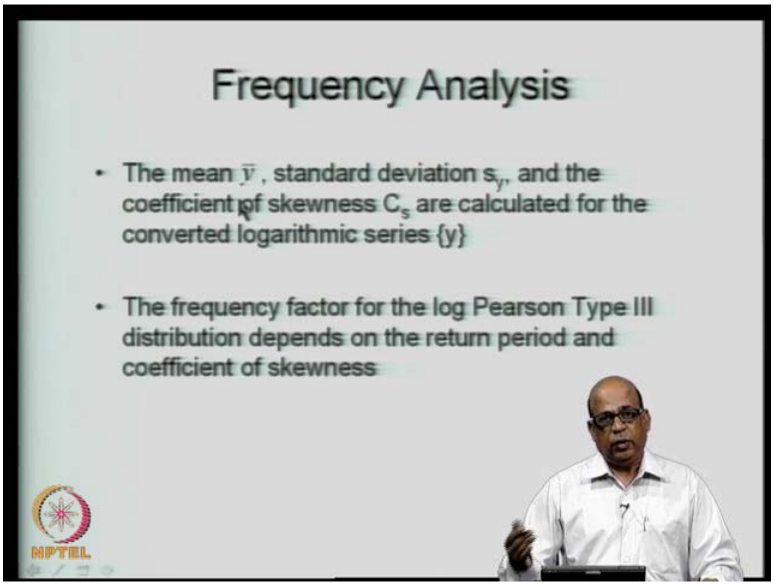


Then there is another important distribution, that we normally consider in the frequency analysis, this is log Pearson type III distribution, we have a Pearson type III distribution and this is the log Pearson type III distribution, when we are considering the logarithm of the values. Because I have not discussed this distribution earlier we will just go through the PDF the PDF of the log Pearson type III distribution is given by lambda to the power beta y minus epsilon beta minus to the power beta minus 1 e to the power minus lambda y minus epsilon divided by x gamma function of beta. And the condition for this to be valid as a p d to be a valid PDF is log of x must be greater than or equal to x epsilon.

So, this has three parameters epsilon, beta and lambda, so it has three parameters where y is equal to log x, x is your original data let say your talking about annual maximum discharges, then the series x consist of the annual maximum discharges. Convert x into y by taking logarithm of x and then use the y in this particular expression to get the PDF.

However, we are not interested in getting the PDF, we are interested in getting the flows values corresponding to a particular return period. So, we convert x into y by taking the logarithm and then construct the series y is equal to log of x.

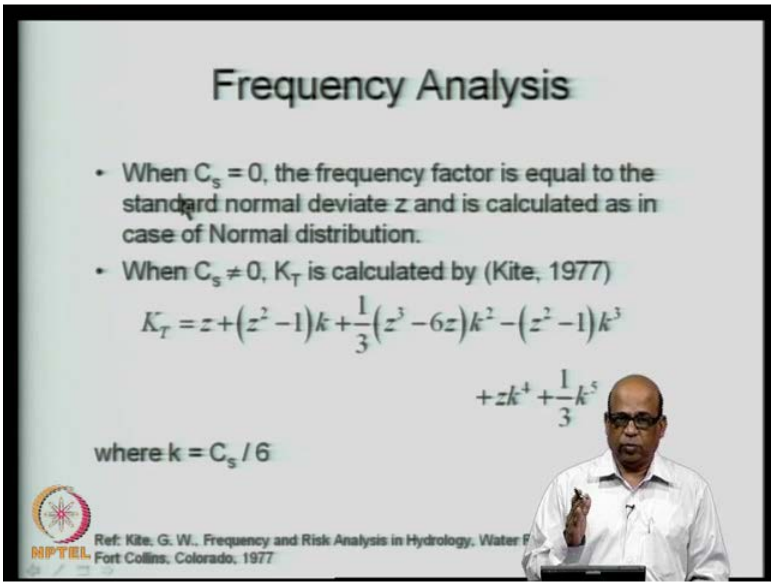
(Refer Slide Time: 19:27)



Then we obtain the mean y bar, this I am leading to the frequency factors as I said frequency factors, we have expressions arrived at different research papers earlier, which are all referred in the text book by Ven Te Chow. And those frequency factors is what we will be using for which, we obtain mean y bar standard deviation S y and the coefficient of skewness C s are calculated for the logarithmic series y.

For the log Pearson type III distribution the frequency factor depends on the return period and the coefficient of skewness, so coefficient of skewness is important for getting the frequency factor for log Pearson type III distribution.

(Refer Slide Time: 20:23)



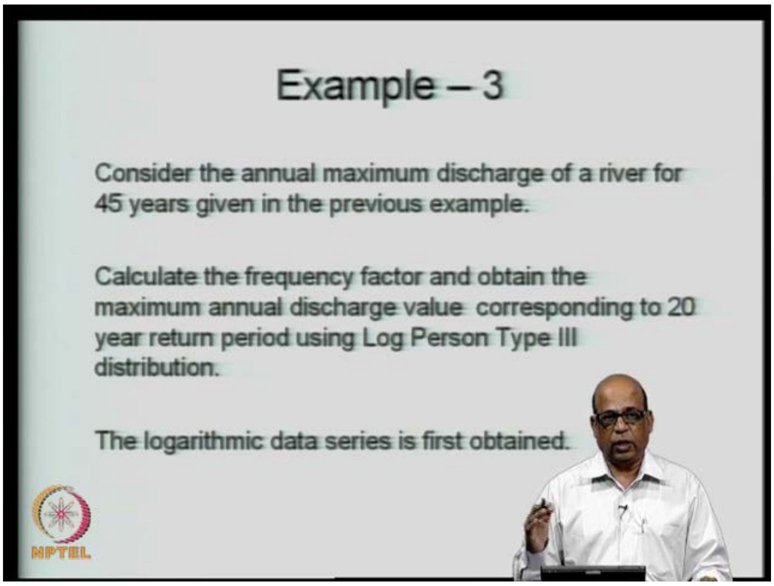
The frequency factor when C s is equal to 0; that means, the coefficient of skewness is 0 what does it mean recall that it is a symmetric distribution. So, when you have symmetric distribution for the log Pearson type III distribution, you still use the method of normal distribution; that means, use the same frequency factors as a normal distribution and then proceed as long as your distribution is symmetric C s is equal to 0.

However, when C s is not equal to 0, which is typically the case, we calculate the K T by this expression, where K is C s by 6 and this z is the z that is given by the normal distribution. So, you can obtain z, because you know T the return period T and therefore, probability of x being greater than or equal to x T is known and from that you can obtain z. And then you know C s which is a coefficient of skewness which you have estimated from the sample data and therefore, you can get K.

Once k and z are known you can obtain your frequency factor K T corresponding to that particular return period therefore, the K T depends for the log Pearson type III distribution on the return period T. And on the coefficient of skewness if the coefficient of skewness is 0, indicating that it is a symmetric distribution just go ahead with the normal distribution, which means use a z value straight away as the K T values for that particular return period.

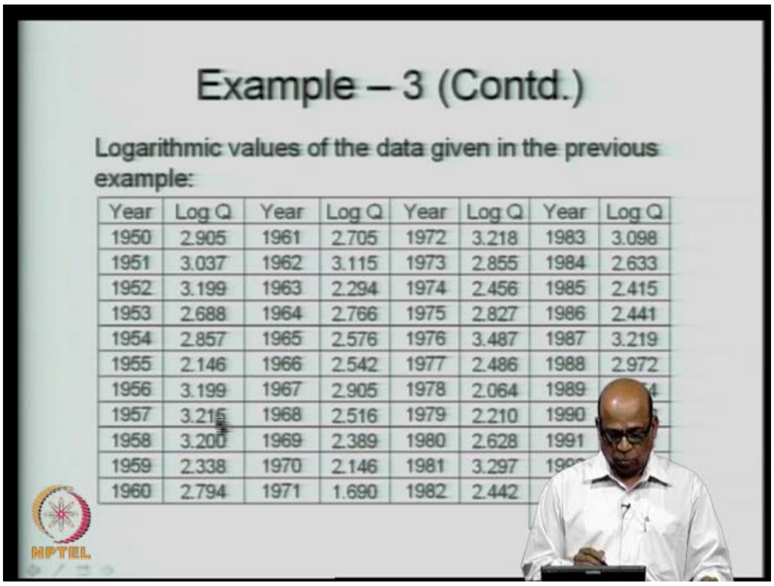
If C s is not equal to 0, which indicates a skewed distribution, then use this expression with the same z value that you get for the normal distribution and then get k small k here by c C s by 6 and then obtain your K T value.

(Refer Slide Time: 22:29)



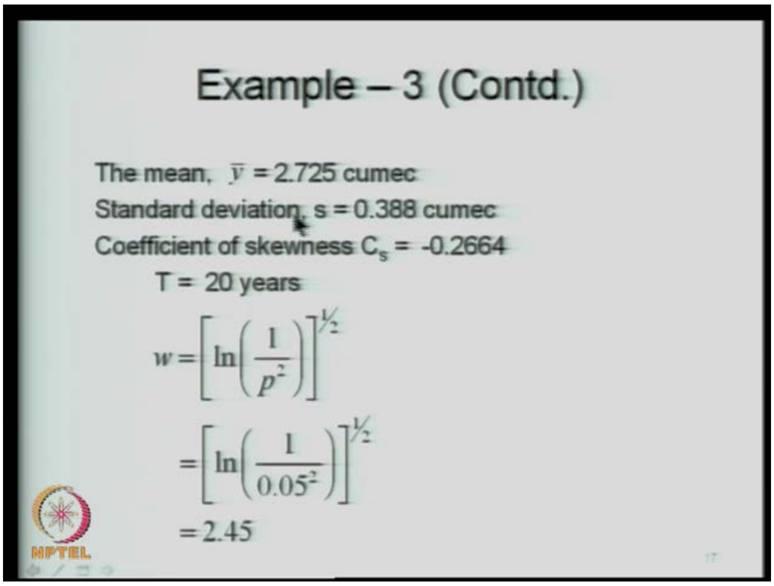
So, let us look at the same example again and then use the log Pearson type III distribution. So, we consider the same data and use the log Pearson type III distribution to get the x t values, so first we construct the logarithmic series, so for the given 45 years data.

(Refer Slide Time: 22:48)



We take the log q and then construct the series of the logarithm values. So, 1950 to 1994 you get the log values and then you obtain the mean standard deviation and the coefficient of skewness with respect to these values now.

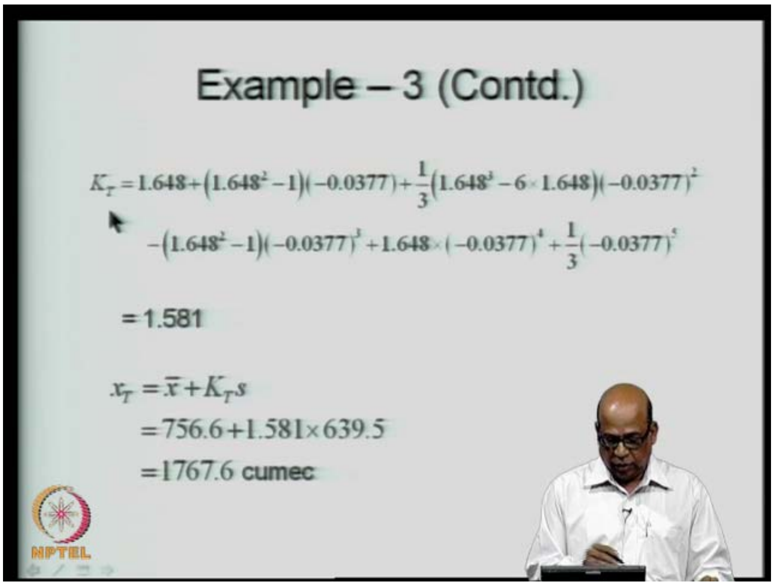
(Refer Slide Time: 23:06)



So, the mean of this series, log series comes out to be 2.725 cubic meters per second and the standard deviation is 0.388 cubic meters per second coefficient of skewness is minus 0.2664 it is a negatively skewed distribution and we have T is equal to 20 years; that means, return period is 20 years.

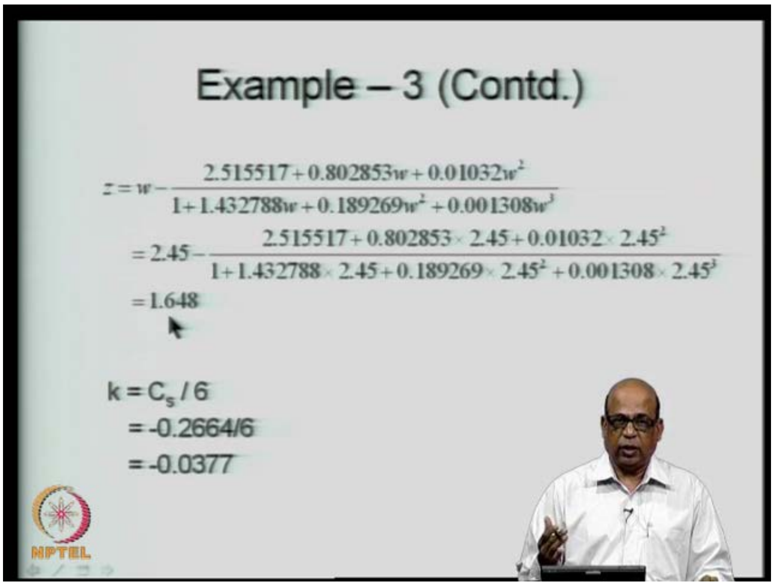
So, we get for this example, w as 2.45 and z as 1.648, what I am doing first I am calculating the z instead of going to the tables. I am using the same expression that, I gave for the normal distribution and I use that with P is equal to 0.05 which corresponds to 20 years and therefore, I get a w of 2.45.

(Refer Slide Time: 24:09)



And I get K T here where K is C s by 6, am sorry I repeat this (Refer Slide Time: 24:21) I get w from here and use this w to get the z value and z turns out to be 1.648 as in the previous case.

(Refer Slide Time: 24:33)



And k as in the previous case by that I mean as in the case of the normal distribution, because we were using the same return period of 20 years then and k is C s by 6 C s is 0.2664. So, the k comes out to be minus 0.0377, this we use to obtain K T. So, K T is given by that expression recall that K T, I expressed in terms of z as well as the small k.

(Refer Slide Time: 25:05) So, we have determined the z we have determined k therefore, we will determine now K T, which is a frequency factor. So, that frequency factor we substitute the values and get 1.581 (Refer Slide Time: 25:16).

Once, you get the frequency factor for which is a function of the return period as well as the distribution always. So, K T factor for return period of 20 years for normal distribution is different from K T factor for a return period of 20 years, for extreme value type one distribution, different from log Pearson type III distribution and so on.

So, once you fix the distribution and the return period you must know you must have expressions to determine K T, which I have just provided for three distributions, normal and extreme value type one distribution and the log Pearson type III distribution. Once you know the K T factors, the expression still remains the same X T is equal to x bar plus K T into s, so we obtain x t as 1767.6 cubic meters per second.

So, for the normal distribution you got a X T value of about 1800 cubic meters per second and for extreme value type one distribution you got somewhere around 1900 cubic meters per second and for the log Pearson type III distribution you got somewhere around 1767 or 1768 cubic meters per second.

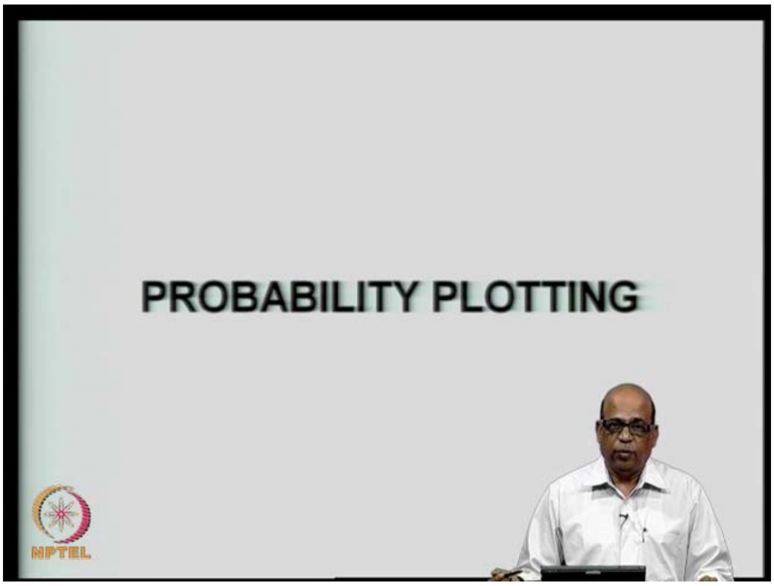
Your hydrologic designs may be quite sensitive to these differences for example, if you use thousand 700 or 1767 against the 1900 that extreme value type one distribution is that is giving, your designs may be quite conservative. In the sense of passing the discharges or some such designs if you are talking about the spillway capacity and so, on. So, the spillway capacities may not be adequate, adequately represented to actual scenarios if, in fact, the data follows extreme value type one distribution.

And therefore, in the frequency analysis it is important to use the correct distribution also the correct data series, whether you use the complete data series or the partial duration series or the annual exceedence series or the extreme value series. And which of the distribution that you use which among several of these extreme value distribution that we talked about which ones you use.

So, there there is also a judgment involve and also analytically or otherwise you must be able to judge, which distribution best fits the particular data that you have and always remember the data is just a small sample of a big population typically in hydrology you get about 30 to 40 years of data and in India it is even smaller in many cases.

So, the sample size is always small and therefore, you must use the right distributions and the right type of return period to get the K T values and then associated X T values for your hydrologic designs.

(Refer Slide Time: 28:23)



Now, that brings us to the question of how do we decide, which particular distribution do we use for the sample data that we have; that means, let say you have last 30 years of observed the stream flow data, can we say that we can approximate this data by normal distribution by log normal distribution exponential distribution and so on.

Now, there are ways available to examine, how best a particular distribution fits the data set that we have and that is what we will start looking at now. The simplest way of doing that is through what is called as a probability plotting, that is let say that you have 50 years of data and then you simply calculate the frequencies you distribute that into number of classes and then calculate the frequencies. And these frequencies you plot relative frequency, let say relative frequency you plot you get a rough idea of whether this relative frequency diagram, may be a histogram that you prepare out of the relative frequencies can be approximated as a normal distribution or can be approximated as a skewed distribution.

So, it may give you some clue as to what kind of distribution that you may use, you may want to use for the type of data that you have; however, more formally what we do is that we plot the CDF the cumulative distribution function of the particular distribution; and then prepare a paper graph paper called as the probability paper. So, for different distributions you may prepare different probability papers on which if the data plots as a straight line, then that data follows that particular probability distribution.

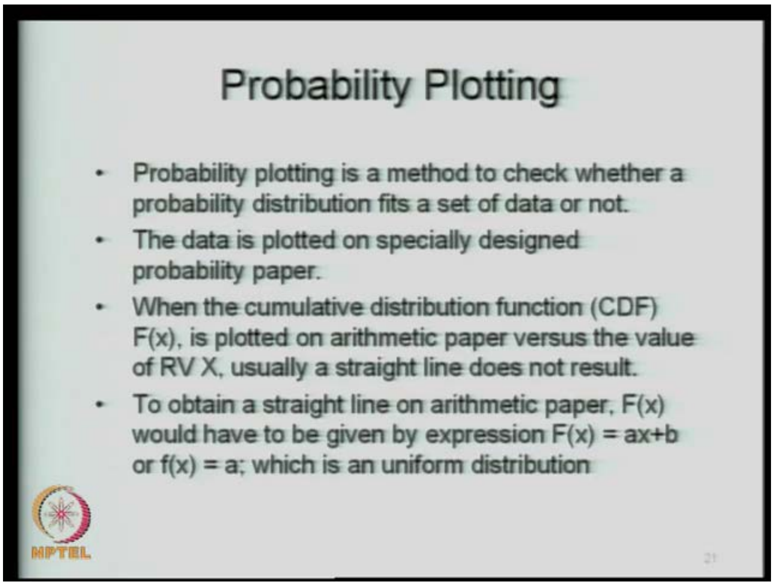
So, you prepare probability papers for let us say normal distribution exponential distributions log normal distribution gumbel distribution and so on, such that if the data that you have actually follows that particular distribution normal log, normal exponential gumbel, etcetera. Then the data will plot as a straight line construction of the probability papers is a very simple exercise, all we are doing is the arithmetic scale that you have in normal graph paper we convert that into a probability scale. We transform not convert really, we transform the arithmetic scale into a probability scale and then prepare a fresh scale for the y axis or x axis depending on how your paper.

Both analytical methods as well as graphical methods are possible for preparing the probability paper, the analytical methods are possible for some of the distributions, exponential distribution and so on. But, when the C D F's are not easily convertible or invertible as I said in the last lecture, we adopt the non analytical method and in this particular case, we use the graphical method.

So, the purpose of probability plotting is to make sure or convince yourself, that the data that you have, the sample data that you have can be approximated to follow a particular distribution; either the exponential distribution or the extreme value type one distribution or the normal distribution log normal distribution and so on.

So, that is topic that we will cover now, in the frequency analysis identification of the distribution is important and probability plotting is one of the ways of identifying the distribution for some of the distribution straightaway. May be gamma distribution you can construct a probability paper for gamma distribution, which defer from one parameter set to another parameter set, because recall that gamma gamma distribution is in fact, a family of distribution. We will start with some analytical techniques with exponential distribution which is extremely amenable to analytical treatment.

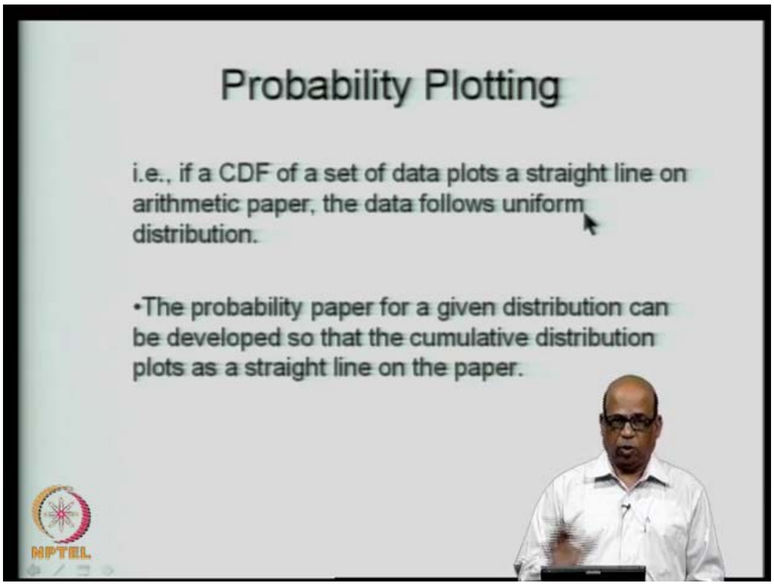
(Refer Slide Time: 32:57)



So, probability plotting is a method to check, whether a particular probability distribution fits set of data or not and the sample data that we have we plot on specially designed probability papers. So, when the cumulative distribution function C D F is plotted on arithmetic paper versus the value of R V, let say that I plot f of x on the y axis and then x the random variable x on the x axis, usually it is not a straight line.

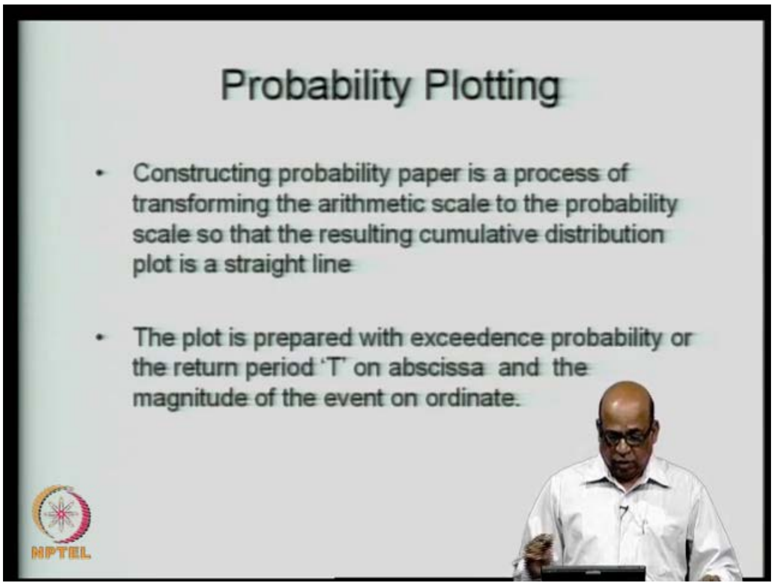
Our aim is to obtain a straight line of F of x that is capital F of x. So, if F of x has to plot as a straight line on a arithmetic paper, what does it mean F of x has to be equal to a x plus b. So, if your F of x actually plots as a straight line then F of x has to be a x plus b or f of x is equal to a which is a uniform distribution. So, if you have a uniform distribution, on the arithmetic paper the uniform distribution will straightaway plot as a straight line, because F of x will be equal to a x plus b.

(Refer Slide Time: 34:26)



So, if a C D F of a set of data plots as a straight line on arithmetic paper; that means, we have not done any transformation of the scales simply we have used the arithmetic graph paper and then, we see that the C D F plots as a straight line. In that case it is straightaway a uniform distribution; however, for most distributions for most other distributions, it would not plot as a straight line. And therefore, we construct a graph paper, which is a probability paper with the arithmetic scale transform to probability scales. The idea here is, that if you plot the f of x it should plot as a straight line, on the probability paper.

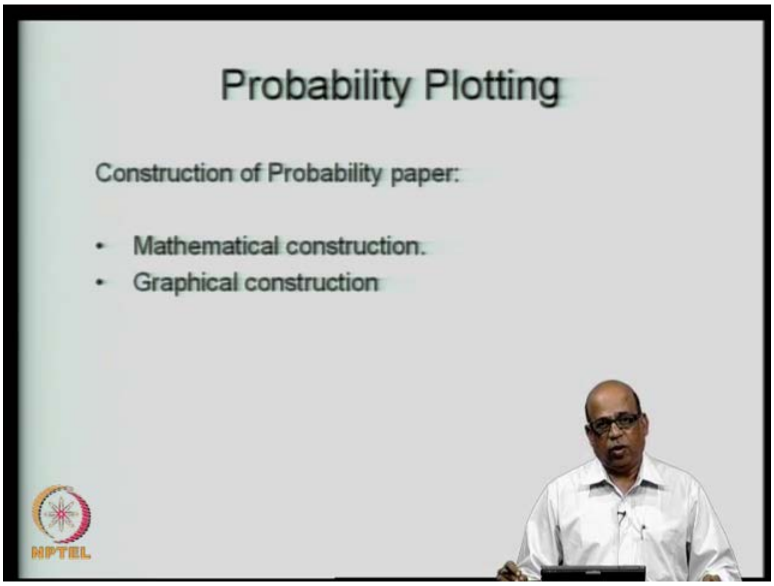
(Refer Slide Time: 35:15)



So, how do we transform this, so this is a process of transforming the arithmetic scale to the probability scale, so that the resulting C D F plots as a straight line on this particular paper. And typically when we are using the probability paper as I will demonstrate presently, we may use the return period T, you know because, you are constructing the probability paper for a particular application.

You can prepare it in a way that is most convenient to you, you may use t return period itself or you may use the probability of exceedence or probability of x being less than or equal to x and so on. So, depending on the application that you want to use you transform the axis in a convenient fashion.

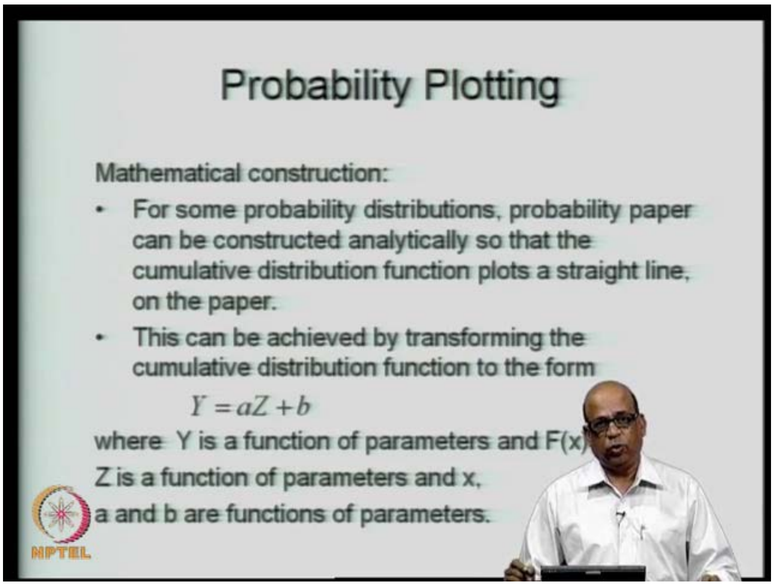
(Refer Slide Time: 36:08)



So, as I mentioned you can have analytical or mathematical construction or you may use graphical construction. The mathematical construction is simple for distribution such as exponential distribution, where you can express x in terms of f of x, that is they are invertible. Whereas, for distribution such as normal distribution or may be log normal distribution, log Pearson type III distribution, etcetera.

Analytical construction will be difficult and you may want to go with graphical construction both of them are very simple procedures, we will just go through both these procedures, for I will demonstrate for exponential distribution as well as for the normal distribution. The same procedure can be used for most of the distribution.

(Refer Slide Time: 37:04)

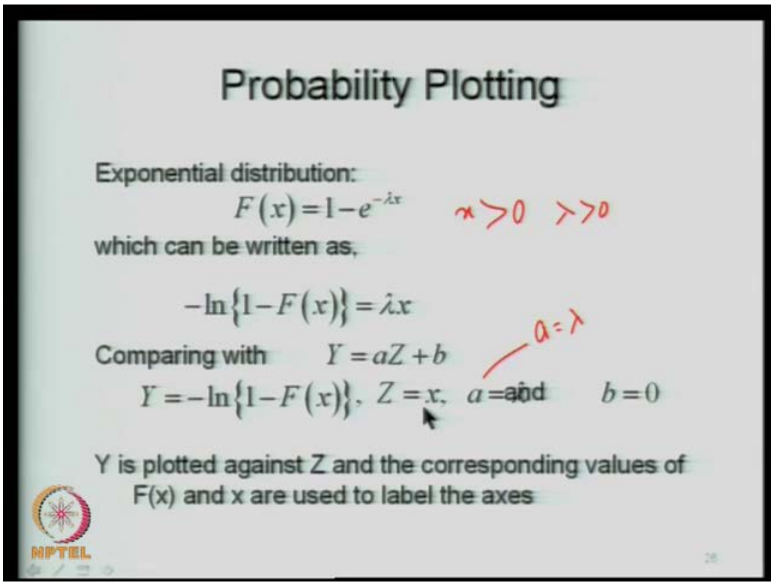


So, as I said for some probability distributions, the probability papers can be constructed analytically, what is it that we want? We want the cumulative distribution function to be of the form Y is equal to a Z plus b. Where Y is a function of the parameters and the cumulative distribution function and Z is a function of parameters and x which is your random variable and a and b are functions of the parameters.

Let us, say you have exponential distribution, then exponential distribution has a parameter lambda. So, Y can be a function of the C D F as well as lambda C D F means it also includes lambda and Z is a function of your original sample, original data x as well as the parameter and a and b are functions of the parameter. So, if you convert your C D F into such a form, then you can plot Y versus Z that plots as a straight line this is a equation for straight line.

So, your idea is that the non-linear F of x you want to transform in such a way that, you will get a linear expression of some function of F of x and then you prepare the graph by transforming the axis.

(Refer Slide Time: 38:34)

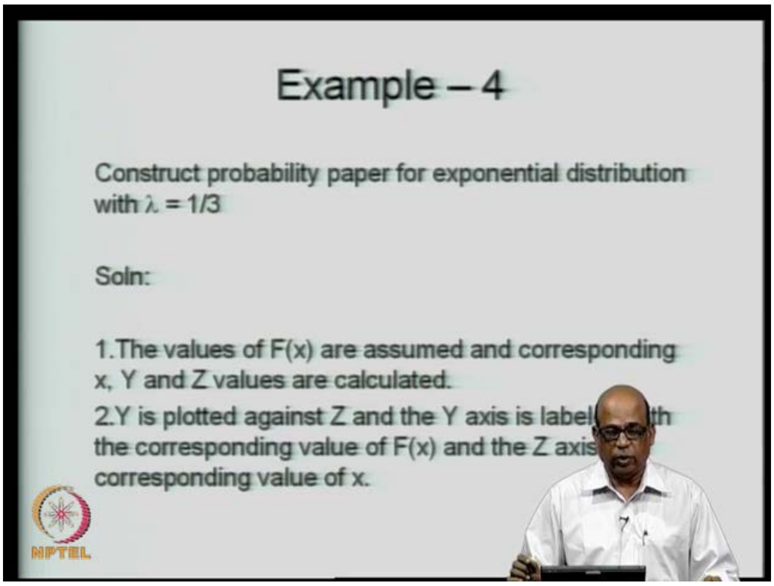


So, exponential distribution, you have F of x is equal to one minus e to the power minus lambda x and this is valid for x greater than 0 lambda greater than 0. Now this I will write it as, I will take lambda x on the right hand side and write this as minus l n 1 minus F of x is equal to lambda x.

What is it that I want? I want an expression of the type Y is equal to a x plus b and therefore, I will write this as Y is minus l n 1 minus F of x this side and Z is equal to x and a is equal to lambda and b is equal to 0. This is a is equal to lambda here and b is equal to 0. So, form that I can write this as a linear expression Y is equal to a Z plus b, so given F of x you know Y now and Z is equal to x and therefore, I can plot Y against Z.

So, what I will now do I will simply plot Y against Z and then label them, F of x and x because, Y is a function of F of x and Z is a function of x Z is equal to x. In fact, in this particular case, so I plot Y against Z and then simply change the labels that is all. Let us see, what how we obtain this.

(Refer Slide Time: 40:19)

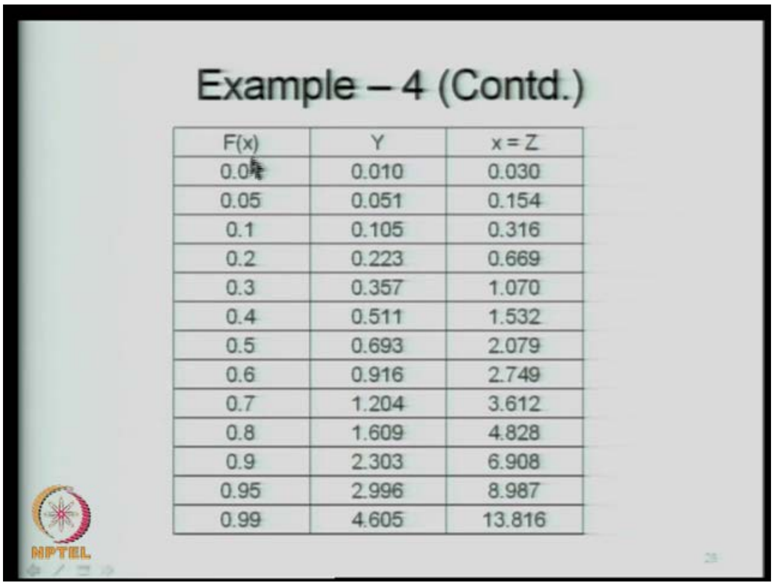


So, we will take an example, for the exponential distribution with a parameter lambda as 1 by 3, with the value of lambda as 1 by 3. So, what do we do here (Refer Slide Time: 40:39), we want the values of Y is equal to a Z plus b, so we want to plot Y against Z, so we want Y values corresponding to a particular x value and the associated Z value. So, that we can plot Y versus Z, that is the whole idea. So, we start with some values of F of x we assume values of F of x, F of x varies from 0 to 1.

So, you assume convenient values of F of x and then calculate corresponding value of x, Y and Z, x you can calculate because, it is a exponential distribution it is invertible therefore, given F of x you can calculate x and Z you can calculate, because Z is equal to x and Y you can calculate because you are fixing the value F of x.

So, fix the value of f of x get x, Y and Z and once you get Y and Z plot Y and Z and label the associated x and Y axis by Z as well x as well as F of x as I will demonstrate. So, Y is plotted against Z and the Y axis is labeled with the corresponding value of F of x, Y axis will be F of x and the Z axis with corresponding value of x , so let see what we do here.

(Refer Slide Time: 42:06)

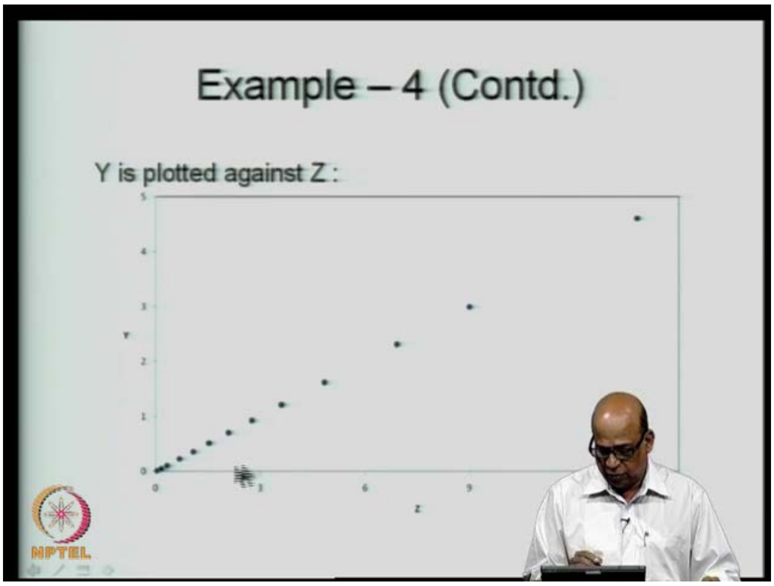


F(x)	Y	x = Z
0.01	0.010	0.030
0.05	0.051	0.154
0.1	0.105	0.316
0.2	0.223	0.669
0.3	0.357	1.070
0.4	0.511	1.532
0.5	0.693	2.079
0.6	0.916	2.749
0.7	1.204	3.612
0.8	1.609	4.828
0.9	2.303	6.908
0.95	2.996	8.987
0.99	4.605	13.816

So, we are assuming f of x here and (Refer Slide Time: 42:11), from F of x we use this expression to get Y which is minus l n 1 minus f of x. So, this is minus l n minus 1 minus F of x that is how you get Y and x is equal to Z itself. So, you get Z here is equal to x and associated x value you get from here. So, you will get x is equal to Z, so you are getting F of x fixing F of x getting Y and form F of x you are getting x and then that itself is equal to Z.

The way to get x from F of x is simply by using (Refer Slide Time: 42:52), this F of x you are fixing therefore, lambda is known in this particular case it is 1 by 3 the example, that I am solving. So, lambda is known therefore, you can get x, that is how you get x and that is equal to Z itself; so you tabulate F of x Y and Z, now we plot Y versus Z now.

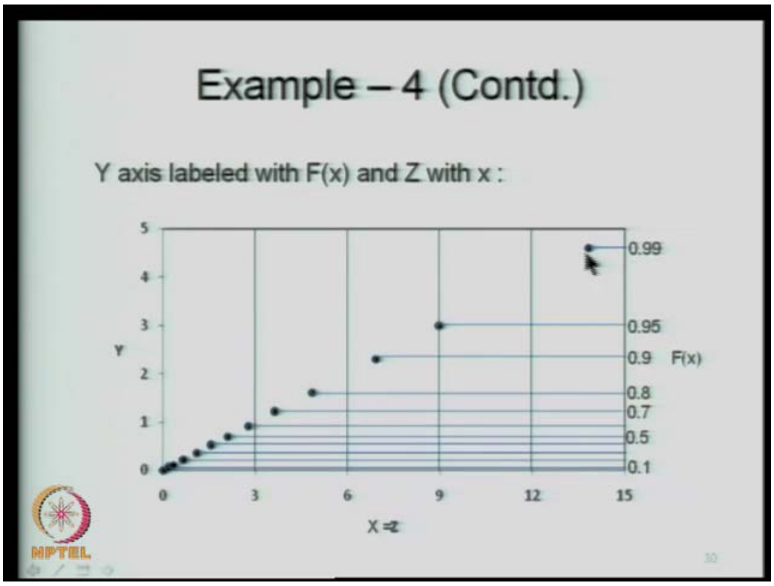
(Refer Slide Time: 43:19)



So, we plot Y here on the Y axis and Z here on the Z axis this is a arithmetic state or we then do is we change the labels. So, Y was on arithmetic scale, this was Y I simply put Y is equal to the corresponding value of F of x and x is equal to corresponding value of Z that is all. So, I have plotted Y versus Z and I simply change the labels here and put F of x here, with associated values of F of x and Z here, with associated values x that is x here, with associated value of Z in this particular case x is equal to Z.

So, it does not matter, what is it that we achieved now, we make sure that this plot f of that is Y versus Z is plotting as a straight line here, if you join all these points its plots as a straight line, because Y is a linear function of Z. So, it plots as a straight line and then, we simply change this labels, what I mean by that is that this point here this point corresponded to some 4. something that comes to somewhere around here 4.605, that is 0.99.

(Refer Slide Time: 44:39)

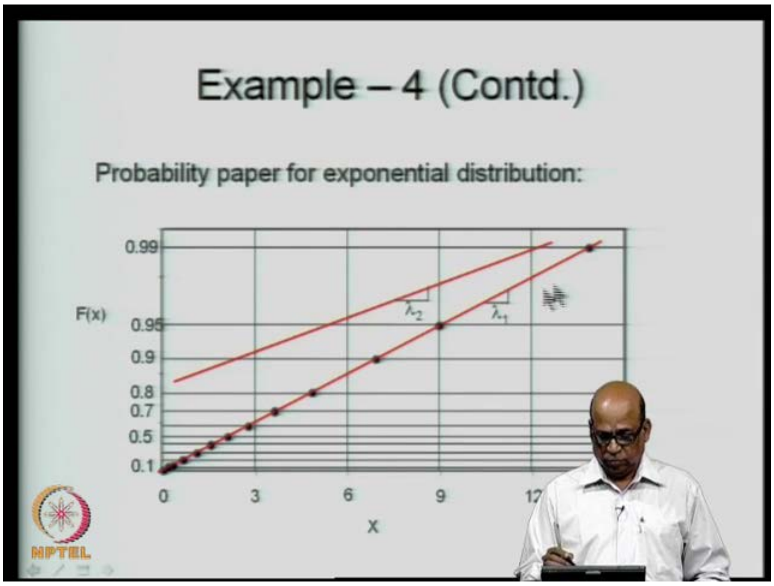


So, that point that I name it as 0.99, so I am just translating the Y axis into a F of x axis, so the value of Y that you had I will rename or re-label it, with the associated F of x value. So, this comes to 1.99, next comes to 0.95, next comes to 0.9 and so on. So, this is what I am doing now (Refer Slide Time: 45:07), this was 0.99 corresponding to 4.60, 5.95 corresponding to 2.996 and so on.

So, these are the values that I will replace for these values of Y, so I will get 0.9, 0.95 etcetera like this and then we draw lines here, corresponding to these and get the probability scale. So now, you get rid of Y get rid of Z, retain x, retain f of x that is it and that gives you a probability paper it will appear like this (Refer Slide Time: 45:47), so all I did is, that I plotted Y against Z, which comes to be a straight line and then corresponding to each of the points that I have obtained I will put the associated F of x values, from here and then construct another scale here, which is the probability scale, which is F of x in fact.

And then, rename or re-label all this points like this and then extend the line and obtain the probability paper and then for convenience you forget about this Y, put F of x on the left side and x on the Y side, x on the x axis.

(Refer Slide Time: 46:24)



So, this is the probability paper that you get for exponential distribution 0.99, you have here 0.95, 0.9, 0.8, etcetera and x value is here. Now, a question arises, we did this example for a given value of lambda, namely lambda is equal to 1 by 3. Remember this x axis is still arithmetic scale, all the that has change is only the y axis, if we change lambda, can we use the same probability paper.

In the case of exponential distribution s you can still use the same probability paper, that you have constructed, let us say you constructed lambda is equal to 1 by 3, so if your data. In fact, follows exponential distribution with parameter value, lambda is equal to one by three it will follow exactly the same straight line that I showed just now.

However, if your data follows exponential distribution with some other parameter, let us say lambda is equal to some other value, you can still use the same probability paper but, it will plot as a different straight line on the same probability paper, I am showing it here. So, the data that you had for lambda 1 is equal to 1 by 3, if you use the exponential probability paper would plot as this straight line, if you have another exponential distribution with another lambda value, you may get another straight line like this.

Remember, this x axis is a arithmetic line and therefore, for a given distribution you can vary these values it need not be 3, 6, 9, 12 and so on. So, you may get another straight line as long as your F of x versus x plots as a straight line on the exponential, exponential probability paper then you can assume that it follows a exponential distribution.

Of course, it need not be exactly a straight line, it may approximately follow a straight line then, you are reasonably sure that you can use exponential distribution. So, this is how we construct mathematically the probability paper for a given distribution.

So, when the distribution is invertible and you can express F of x in terms of that is z you can express in terms of y and x as I said (Refer Slide Time: 49:17), you what you did here is F of x you assume and then calculated x Y and Z, this need not be possible this may not be possible for many distribution, because from F of x you may not be able to get x and so on.

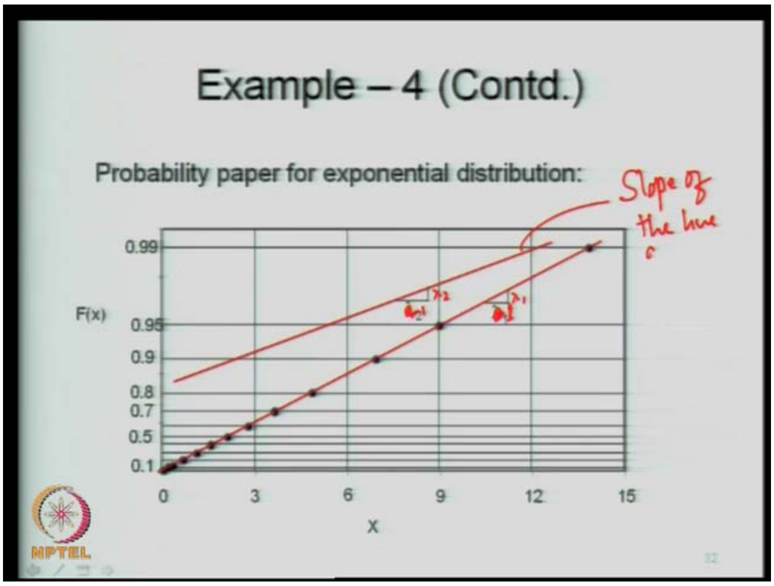
So, distribution such as exponential distribution, gamma distribution, etcetera such analytical treatment is possible; for the gamma distribution you may use the analytical procedure, because you have gamma function there and then once you fix the parameters you can obtain the gamma distribution and then associated x values you can obtain and so on. However, for gamma distribution you may have to you will have to construct different probability papers corresponding to different sets of parameters.

So, the probability paper construction is a function of the parameters, that you consider for the gamma distribution whereas, in the case of exponential distribution, we saw that you use some parameter value. And then construct a exponential probability paper and that probability paper is valid for any exponential distribution with different parameter values.

So, even if your parameter values defer, you can still use the same exponential probability paper. The exponential probability paper that I just demonstrated on that probability paper your data will plot as a straight line if it, in fact, follows a exponential distribution, the slope of that particular straight line will give you the lambda value exponential distribution has only one parameter namely lambda.

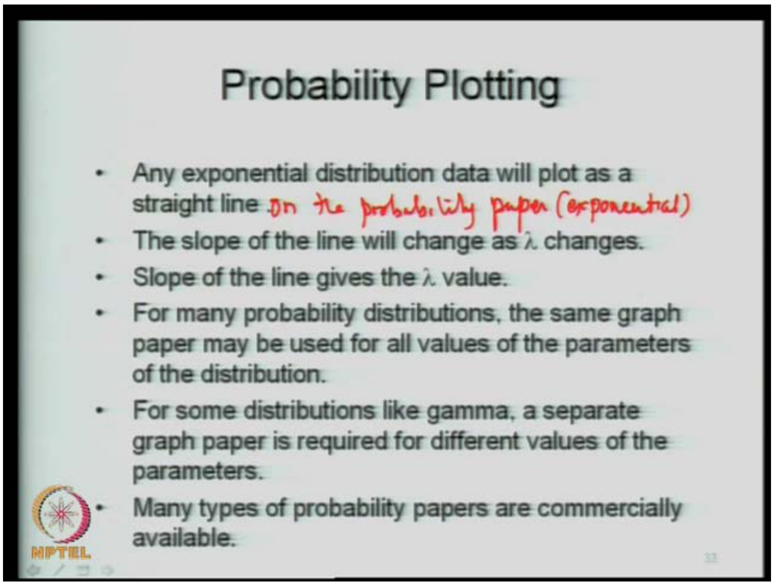
So, the slope of the line straight line that you obtain by plotting F of x versus x on the exponential distribution will give you the parameter value like I showed earlier, if you have two different exponential distribution they may plot like this.

(Refer Slide Time: 51:52)



So, the slope of this line this has to be lambda 1 here and lambda 2 this is a slope actually, so this is lambda 221 and this is lambda 121. So, the slope of this line gives you the parameters, so I will write here slope of the line gives lambda.

(Refer Slide Time: 51:58)



So, as long as your data plots as a straight line on the exponential probability paper, it follows the exponential distribution with the slop of that particular line giving the parameter lambda. So, we will just go through, what we just did now, any exponential distribution data we plot as a straight line on the exponential probability paper, that is on the probability paper prepare for exponential distribution, then the slope of the line will change as lambda changes.

So, if you have three, four different samples all following exponential distribution all will plot as different straight lines on the same probability paper, then the slope of the line gives the lambda value. Then for as I mentioned for many probability distribution such as exponential distribution, the same probability paper can be used for different parameter sets. However, for distribution like gamma distribution a separate graph paper is required for different values of parameters.

You know the probability papers are commercially available, much like your arithmetic graph paper, log paper, semi log paper, etcetera you can also obtain probability papers, on the internet, there is a good site that is the wible site, where you can download the probability paper for several distribution. So, you can construct or download the probability papers and then plot your data on the probability paper, to see whether that data follows a particular distribution or not.

So, today in the lecture essentially we started off with the frequency factors and we introduced the frequency factors associated with the normal distribution and we solved an example corresponding to that. And then remember that the normal distribution frequency factor is, in fact, the same as the z value which is the standard normal deviate associated with the return period T from the return period T you get P which is 1 by T. And from that you can always get z value and the z value is the same as K T value for the normal distribution, then we also introduced the K T value, which is the frequency factor for the extreme value type one distribution and the log Pearson type III distribution.

Then we address the question of which distribution fits a particular data set; that means, how are we sure that the data set, that we have can be approximated to follow a normal distribution, log normal distribution and so on. So, we introduced the concept of the probability plotting and constructed the probability paper, the probability paper is essentially constructed by transforming the arithmetic scale into a probability scale. So, there are two ways of doing it I introduced today only the analytical way, analytical method, where we express the F of x or the C D F as a linear function, so we construct Z as Z is a function of F of x.

And then plot Z versus Y, Y is that is we are writing Z is equal to a Y plus b or some such thing and then plot Z versus Y, which plots as a linear function which plots as a line straight line on the arithmetic paper. And then we translate or transform the Y axis into the F of x scale and we demonstrated this with the exponential distribution and for the exponential distribution you obtain the probability paper for a given parameter. And the same probability paper can be used for any other parameter as long as the exponential the data follows exponential distribution, it plots as a straight line on the exponential probability paper.

For other distributions like gamma distributions, gamma distribution, etcetera you may have to you will have to construct the probability papers for different sets of parameters gamma distribution, I repeat is a family of distributions. So, associated with a set of parameters, you will get a different shape and size of the the distribution itself.

So, in the next lecture, we will continue the discussion and then introduce the graphical method probability construction, a probability paper construction. And then we will also look at some statistical methods available statistical test available to examine whether a particular data fits a given distribution or not; specifically we will introduce the chi square distribution and the kolmogorov-smirnov test and so on, chi square test and k s test and so on.

So, thank you for your attention, we will continue other discussion in the next class.