Stochastic Hydrology

Prof. P.P. Mujumdar

Department of Civil Engineering

Indian Institute of Science, Bangalore

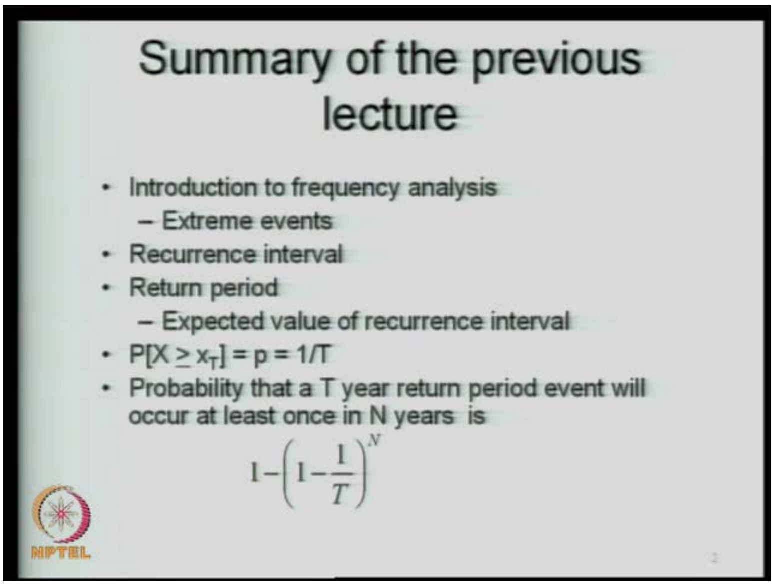
Module No. # 06

Lecture No. # 25

Frequency Analysis – II

Good morning and welcome to this (Refer Slide Time: 00:18), the lecture number 25 of the course Stochastic Hydrology. If you recall in the last lecture, we introduce the concept of frequency analysis.

(Refer Slide Time: 00:34)



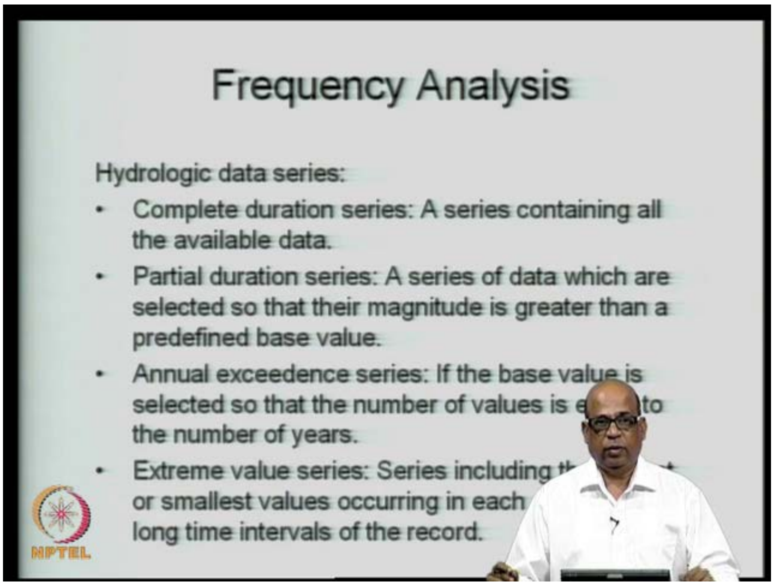
So, when we are dealing with extreme events, for example, high flows or high storm intensities or low flows, which result in droughts and low rainfall values and so on. So, essentially when we are dealing with extreme events, we would be interested in answering questions such as, what is the frequency of occurrences of a given magnitude of flow for example.

And that is when we define the recurrence interval, which is the time duration between occurrences of two extreme events of given magnitude. For example, the flow exceeding a value of 5000 cubic meters per second, what is the recurrence interval between such consecutive occurred nces? And we have also introduced the concept of return period, which is actually the expected value of the recurrence interval. When we say a return period of 20 years for a flood flow, it indicates that, once in above 20 years on an average, we may expect such an event to occur. Then, we went on to define the probability of such an event, which is probability of X being greater than or equal to X T, where X T is the magnitude of that particular flood let us say; and that is given by p which is which we showed it to be 1 over T, where T is the return period.

Then, we also answered the question that, the probability that a T year return period event will occur at least once in N years is, 1 minus 1 minus 1 over T over to the power N, what does this mean? This means that, we have a T year return period event, let us say 20 year flood, but we are interested in getting this, the probability that this 20 year return period flood will occur at least once, let us say next 10 years or next 15 years or next 50 years and so on. So, that probability is given by 1 minus 1 minus 1 over T to the power N. So, this is a concept that, we have introduced in the last lecture. We also solved a numerical example to demonstrate these, so various concepts that we have developed.

Now, we will proceed further and look at, what is a type of data that we use for the frequency analysis. Say for example, we have monthly data for last 20 years, collected at a particular location. We can use the entire data as it is and then, carry out the frequency analysis or we may put a threshold value and carry out the analysis on all those values, which appear above the threshold value and so on. So, there are different ways of constructing the hydrologic series on which, we carry out the frequency analysis.

(Refer Slide Time: 03:58)



Let us, look at some of this, for example, we may consider the complete series that is, whatever observed values that we have, we constitute a series out of that and then, call it as a complete duration series; and carry out the frequency analysis on the entire series, which has been observed. You may have a daily record for last 20, 25 years, you may consider all the daily record and then, carry out the frequency analysis, so such a thing is called as a complete duration series.

However, mostly we will be interested in the partial duration series, a partial duration series is a series of data which we select after putting a threshold value. Let us say, we are interested in flows above 2000 cubic meters per second, so we put a threshold of 2000 cubic meters per second; and then, pick up from the observed data, you pick up all those values, which are above 2000 cubic meters per second and then, construct a series out of that, that is called as a partial duration series.

We also have a commonly used series called as annual exceedence series. The threshold that we put in the case of partial duration series let us say, you adjust the threshold in such a way that, you get exactly those many number of values, as you have the number of years of data. Let us say, you had 40 years of data, you put the threshold such that, you get 40 such values above the threshold.

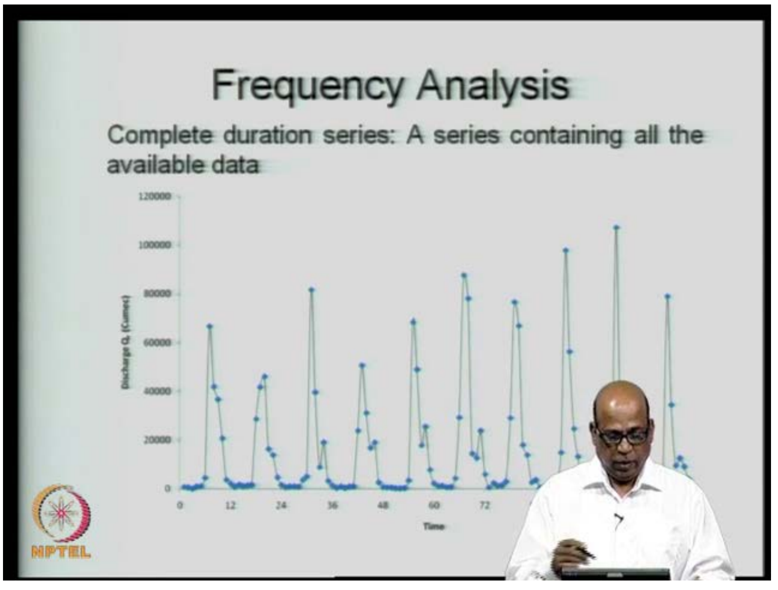
Note that, these 40 values need not correspond to 1 per year, simply 40 such values we get, may be somewhere or some years may have 3 such values or may or may not have any value and so on. So, we get 40 such values, when we have 40 years of data and then, we form what is called as a annual exceedence series. Annual exceedence series is also vey commonly used for frequency analysis.

But, the most commonly used series is the extreme value series: that is, it includes the largest or the smallest value occurring in each of the year. Like by we discussed in the last lecture, every year you pick up simply the maximum value, if you are interested in high extremes frequency analysis or if you are interested in the low extremes, you simply pick up the minimum value occurring in every year.

So let us say, you have every 15 minutes of data as I explained in the last class, 15 minutes you are recording the stream flow at a particular gage. So, every hour, you have 4 such values and every day you have 4 into 24 such values and therefore, every year you will have 4 into 24 into 365 such values, out of these number of values you pick up just 1 value namely, the maximum value that has occurred during that year; and use these maximum values for carrying out the frequency analysis.

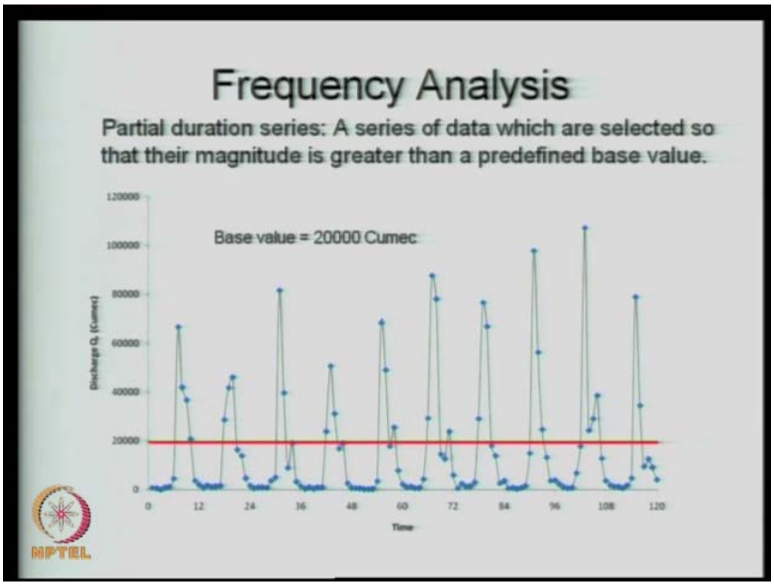
If you had 40 years of data, you pick up 1 value corresponding to each of these years which consisted of those 4 into 24 into 365 values, you pick up 1 value like this, you pick up 40 such values and then, carry out the frequency analysis on these 40 values. So, these are different ways of constructing the data series on which, we carry out the frequency analysis to make it more clear we will see one by one.

(Refer Slide Time: 07:35)



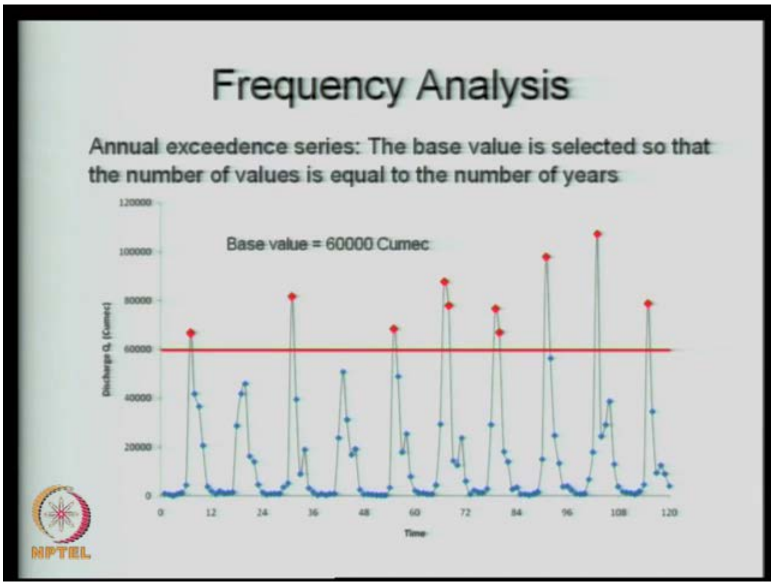
What I mean by complete duration series is? The entire time series that you have observed let us say, this is a example, where we have monthly data for 10 years making it a total of 120 values of course, 10 years is a small duration, this is I am just showing it as a example. When you have monthly data for 10 years, you have the time series like this, you consider the entire time series has as been observed for the frequency analysis and such a series is called as a complete duration series.

(Refer Slide Time: 08:15)



For the partial duration series we put a threshold value let us say, 20000 cumecs we we say that, anything higher than 20000 cumecs is considered for the series. So, some years may have 1 2 3 4 like this (Refer Slide Time: 08:31), and some years may not have any or may have just 1, so we pick up all those values, which either equal 20000 or exceed 20000 and construct the series, this is called as a partial duration series. So, in this case we will consider only these values, which are above this threshold value and this threshold value is called as a base value.

(Refer Slide Time: 09:00)



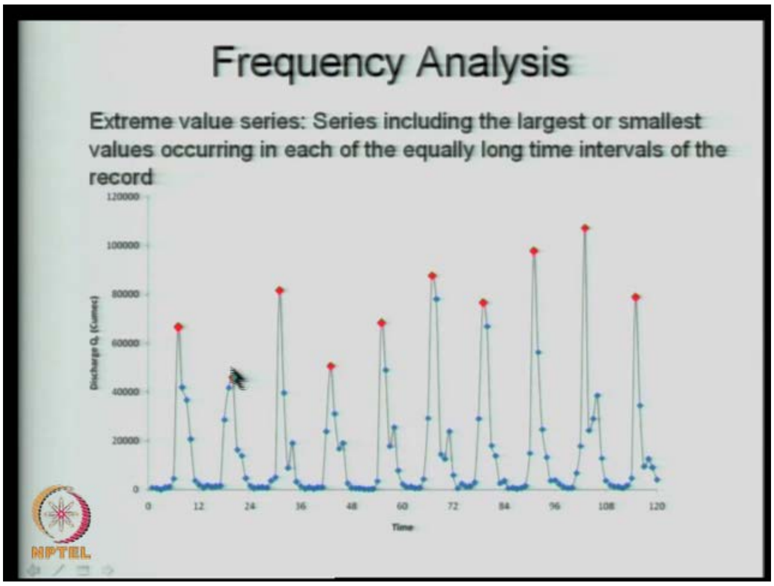
In the annual exceedence series, we still put a base value I have the threshold value as we did in the partial duration series, but we select the base value such that, we obtain these values which are above the threshold, the number of such values will be equal to the number of years of data.

In this case, we had 10 years of data, so we put such a threshold value such that, such a threshold value that we get exactly 10 values above the threshold value 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (Refer Slide Time: 09:31). Some of the years may have 2 such values and some of the years may not have any value above the threshold value, does not matter as long as you get, 10 such values totally, you call it as the annual exceedence series.

What happens in the annual exceedence series? For example, you consider these two events, the flood that has occurred or the maximum values that has occurred here, may have produced this maximum value also, in terms of its ability to saturate the soils; and therefore, a high intensity rainfall occurring immediately after that, may have caused this flood also and therefore, these need not to be completely independent events.

Similarly, these two need not be completely independent events, but for our frequency analysis we have assumed that, all of these events that you are picking up for the analysis are in fact, independent and identically distributed. And therefore, we must have enough justification to assume that, these events are in fact independent.

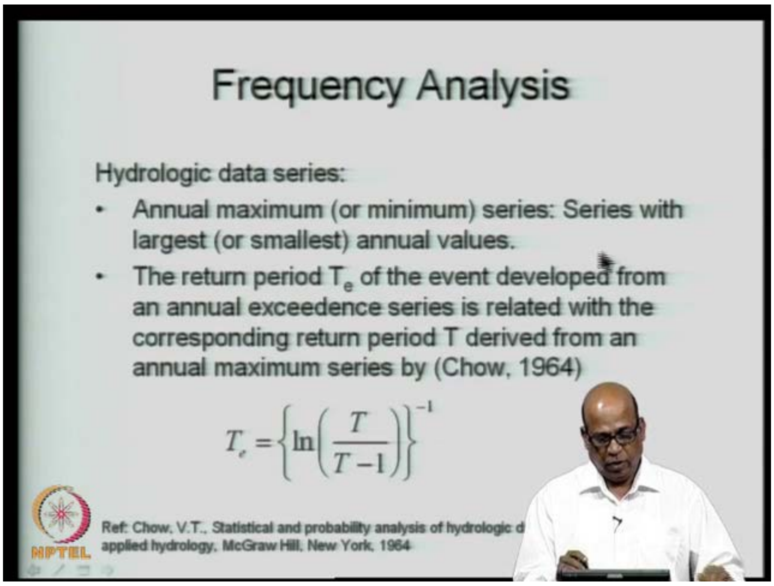
(Refer Slide Time: 10:41)



Then in the extreme value series all we do is, pick up the extreme values. Let us say, we are interested in the maximum values, simply pick up the maximum values occurring every year. So, if you have 10 years of data, you get 10 maximum values like this, pick up those 10 values and then, constitute the series for frequency analysis. So, these are several ways of constructing the series that we consider for the frequency analysis.

As I mentioned, the annual exceedence series and the extreme value series are the most commonly used series for carrying out the frequency analysis. The concept of the return period that we introduced in the last lecture was based on the annual maximum or the annual extreme series.

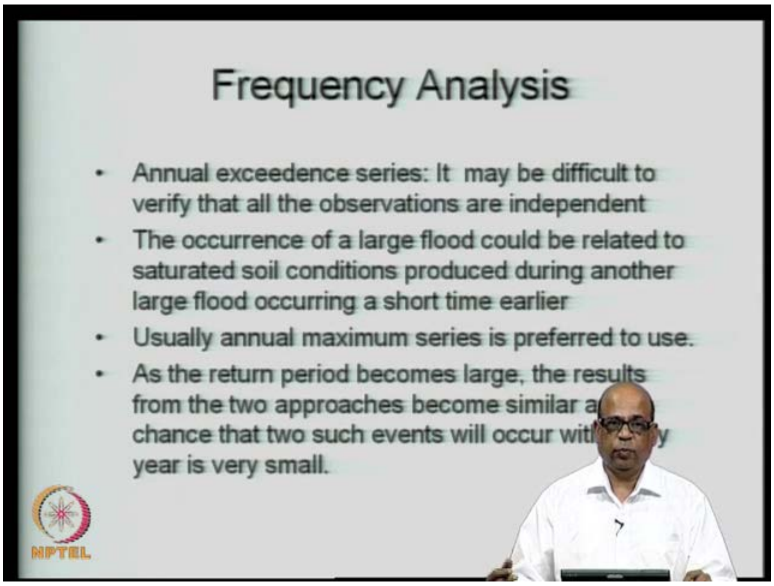
(Refer Slide Time: 11:37)



So, the annual maximum or minimum series with has the largest or the smallest annual values, we introduced the concept T return period from annual maximum series, but this can be related to the annual exceedence series, by this expression given by Chow in 1964 that is the return period T of the event from an annual exceedence series is equal to log T by T minus 1 raised to the power minus 1, where T is derived from the annual maximum series.

So, T is equal to 1 by p is what we derived in the last class, where p is a probability of x being greater than or equal to X T or probability of exceedence of that event, and from that you we have the T and from T, you can get T e for the annual exceedence value.

(Refer Slide Time: 12:43)



The annual exceedence series as I just mentioned, may have more than one event occurring in each of the years, in some of the years; and therefore, the series that you have just constructed as annual exceedence series, may not really be independent the values may not be really independent. So, it may be difficult for us to verify that all the observations are independent.

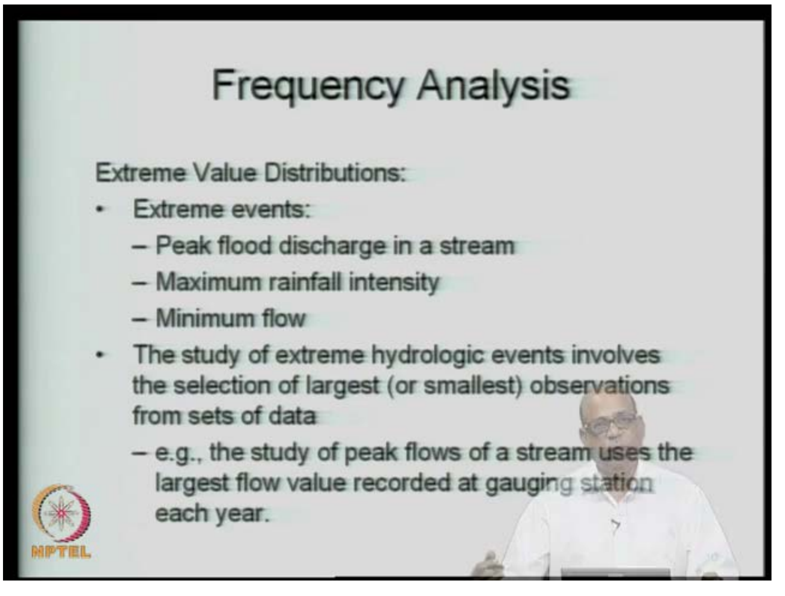
The occurrence of a large flood as I just mentioned in the annual exceedence series let us say, a flood occurred in the month of June and a flood also occurred in the month of July, the July month flood may not be completely independent of the June month flood, whereas, if you are looking at the annual maximum series, what we would have picked up? Let us say, between June and July, July was the highest occurring in that particular year, we would have only picked up the July month flood, for that particular year. In the next year, it may occur during the October month and therefore, the two floods may be totally unrelated.

And then, as the return period becomes large let us say, that from 10 year return period to 20 year return period to 100 year return period to 1000 year return period and so on. So, as your return period increases, what is happening to the probability of occurrence of the extreme events? Remember, p is equal to 1 by T when we said, the probability that we are talking about there is the probability of occurrence of that particular event in a given year.

So, probability of occurrence of a T year event in a given year is the probability that we are taking about. So, as your T increases, 1 by T decreases, so 1 by 100, 1 by 1000 and so on. So, as T increases, the probability of occurrence of that event in a given year decreases; and therefore, as the return period increases, the results that you get from the frequency analysis, from annual exceedence series and the annual maximum or minimum series, they tend to converge. So, there is not too much of a difference, whether you use the annual exceedence series or the annual maximum series for large T.

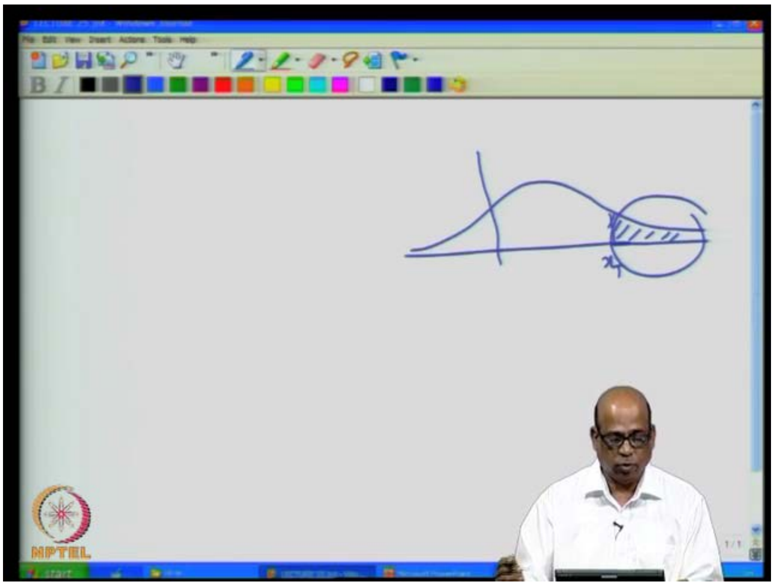
However, for small T for example, if you are looking at urban storm water drainage, where your return periods will be of the order of 5 years, 10 years and so on. Now, it is in such situations that, the choice of the particular type of series whether it is annual exceedence series or the annual maximum series that becomes important.

(Refer Slide Time: 15:47)



Now, we have also seen in earlier lectures, the extreme value distributions which deal with extreme events and by extreme events, we mean either flood peak flood discharge in a stream or maximum rainfall intensity or minimum flows an so on. So, given any time series, we may be interested in only the extreme values recorded. And we fit the distributions for such extreme values and these are called as the extreme value distributions. In fact, in the last lecture also I showed, what way would be interested in such situation sees really towards a tail of the distribution.

(Refer Slide Time: 16:30)

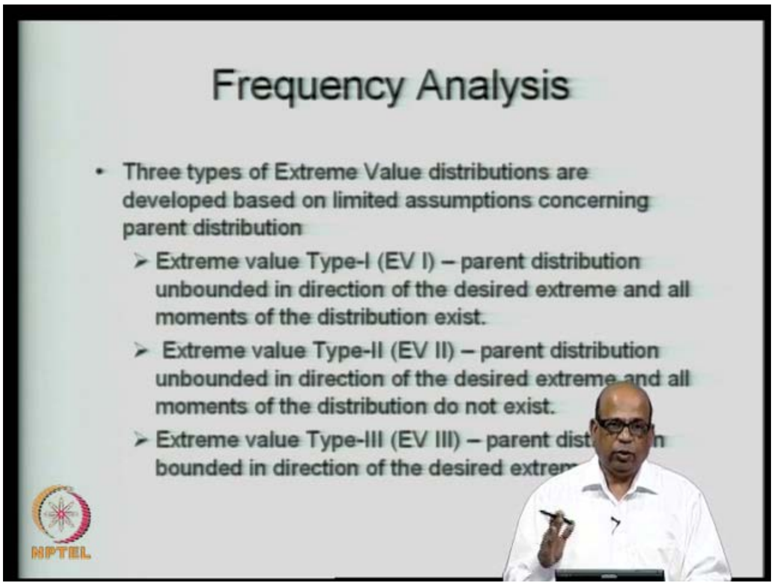


Let us say, you have the distribution like this of the random variable x and then, you are interested in high flows let us say that, this is x T that we are talking about. So, we are interested in this region, which is the tail of the distribution and this these values is what we would be considering in the extreme value distribution (Refer Slide Time: 16:50). So, we are essentially putting fitting distributions to the extreme values and then, talking about probability of x being greater than or equal to x T, which is actually this area and so on.

The extreme value distributions that we so formulate are derived from the parent distribution or or a function of the parent distribution. So, depending on whether the parent distribution is bounded in this direction or is unbounded in both the directions and so on, we get different types of extreme value distributions, as we have seen in one of a earlier lectures.

So, in the case of hydrologic extremes, we may be interested in let us say study of peak flows (Refer Slide Time: 17:46) where we use the largest flow values recorded data gauging station or if we are interested in low flows, we may simply pick the smallest observed flows in a particular stream and so on and then, formulate the the series corresponding to these.

(Refer Slide Time: 18:10)



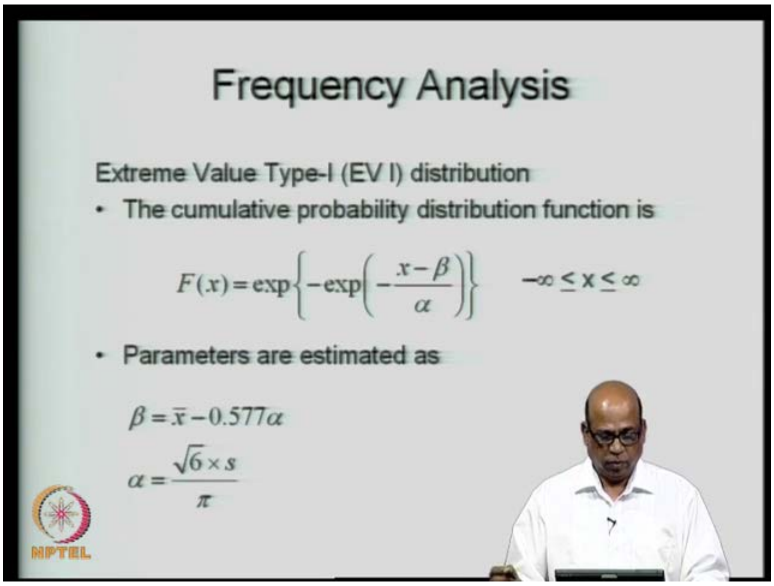
Depending on the parent distribution, there are three types of extreme value distributions that we normally use. One is the extreme value type I, E V I, it is called as it is called as E V I in which case, the parent distribution is unbounded in the direction of the desired extreme. Let us say, you are talking about maximum flows or the maximum values, so the distribution is unbounded in the direction of the maximum flows; and all the moments of the distribution exists, because we will be interested in distribution being unbounded as well as, the moments being moments existing in such a case, we call it as extreme value type I distribution.

In the case of type II, the parent distribution is unbounded in the direction of the desire desired extreme, but all the moments of the distribution do not exist. In the case of type III, normally we use extreme value type I and extreme value type III. In the type III, what happens? The parent distribution is bounded in the direction of the desired extreme.

Let us say, you are looking at low flows, low flows generally you know in hydrology, most of the variables that we considered are all nonnegative and therefore, they are bounded by 0. So, when you have a distribution which is bounded by 0 and you are considering the distribution of the low values, then you use extreme value type III.

So, typically the extreme value type III is used for low extreme values, minimum values. Whereas, extreme value type I is typically used for maximum flows or maximum values, although it can also be used for in some situations, it can be used for low values as well.

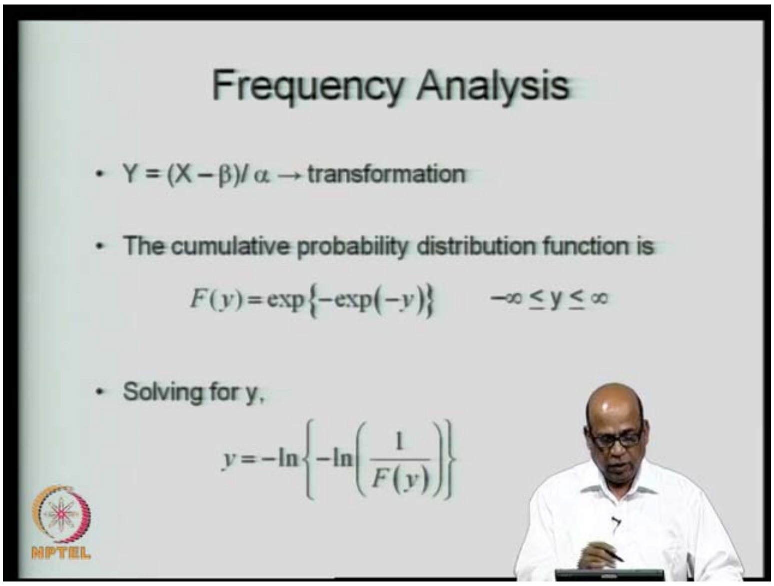
(Refer Slide Time: 20:25)



Let us look at the extreme value type I distribution. The type I distribution, the C D F or the Cumulative Probability Distribution, cumulative distribution function F of x is given by this we have seen earlier just for completeness sake I will put it again, exponential minus exponential minus x minus beta or alpha, where beta and alpha are the parameters of these and it is defined for the entire region minus infinity to plus infinity.

These parameters alpha and beta are estimated thus, alpha is equal to root 6 s divided by pi, where s is the standard deviation, sample estimate of the standard deviation; and beta is equal to x bar minus 0.577 alpha. Alpha given by this (Refer Slide Time: 21:16), x bar is the sample mean, so this is how we estimate alpha and beta, once alpha and beta are given, the F of x is completely defined. This is the extreme value type I distribution as I said, it is unbounded on both the directions, so it is minus infinity to plus infinity.

(Refer Slide Time: 21:38)

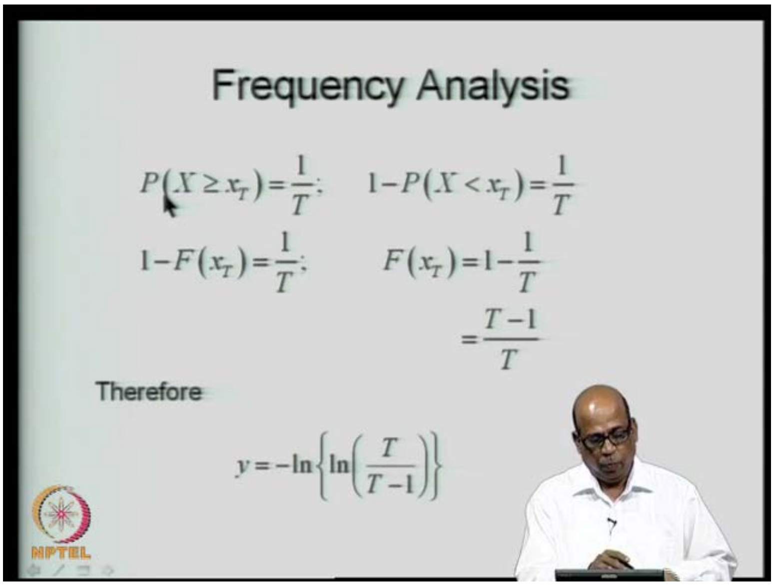


For convenience look at this expression (Refer Slide Time: 21:41), minus x minus beta over alpha, so we use the transformation, y is equal to x minus beta over alpha, we use this transformation (Refer Slide Time: 21:48), and then we express the c d f, these are linear transformations. So, we write F of y is equal to exponential minus exponential minus y, and y varying from minus infinity to plus infinity from this. This is commonly known as the double exponential distribution, you have exponential minus exponential minus y, double exponential distribution.

It is possible for us to solve for y from this expression and then, we do this to get y is equal to minus l n minus l n 1 divided by F of y. Such type of c d f's are called as, such type of functions are called as invertible functions; that means, given F of y, I am able to express y in terms of F of y, but not all c d f's have this property of invertibility. Because, the extreme value type I distribution, the transformation of that into the double exponential distribution is invertible we use this property and then, carry out the frequency analysis.

The frequency analysis that we carry out directly on based on the c d f's, using the property of the invertibility are actually analytical methods, which means analytically we should be able to x determine y T for a given return period T. Because, we will know from a given return period T, we know the probability of exceedence probability of let us say, y being greater than or equal to y T and from which we know F of y, from F of y we get y. So, this is just a straightforward analytical method, which is possible only when your c d f's are invertible in the sense that, we will be able to express x in terms of F of x or y in terms of F of y in this particular case.

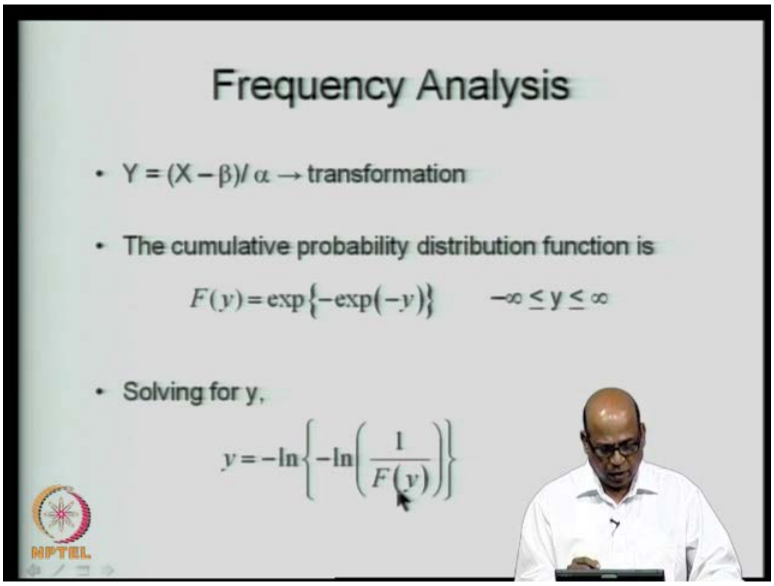
(Refer Slide Time: 23:50)



Because, our probability small p we defined as probability of X being greater than or equal to x T, this is the magnitude corresponding to the return period T and this we have shown it to be 1 over T. So, small p which is equal to probability of x being greater than or equal to x T is 1 by T. What is probability of x being greater than or equal to x T? We can write this as, 1 minus probability of X being less than or equal to x T, because it is a continuous variable, it does not matter if you write less than or less than or equal to, so this will be equal to 1 over T, this is our F of x.

So, we write this as 1 minus F of x T, this is F of x T (Refer Slide Time: 24:31), so 1 minus F of x T is equal to 1 over T. So, from this we write F of x T is equal to 1 minus 1 over T or this will be equal to T minus 1 over T.

(Refer Slide Time: 24:51)

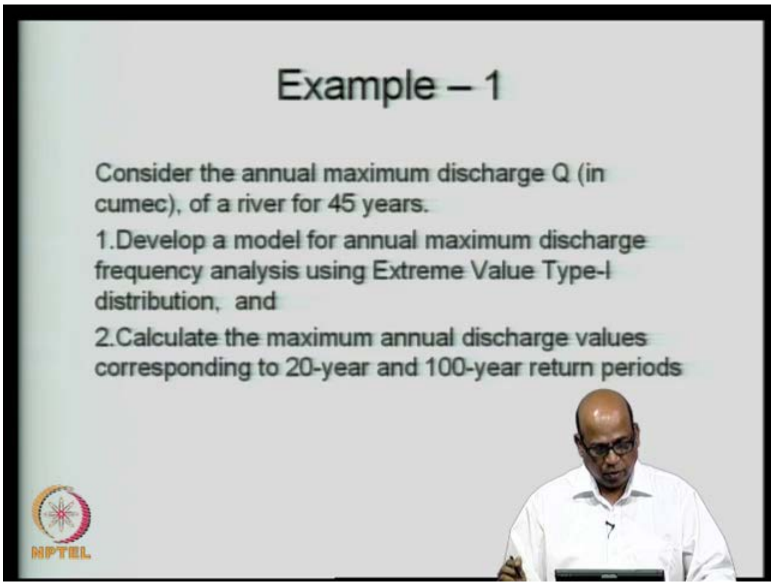


So, from here y is equal to minus l n minus l n 1 over F of y, now this F of y using the expression for F of y here, F of x T in this particular case I write it as, y is equal to minus l n l n t over T minus 1, because 1 over F of y. So, given the return period, straightaway we will be able to estimate the corresponding y, once you know y you know the transformation that we have used that is X minus beta over alpha, beta and alpha are estimated from the sample; and therefore, we should be able to get the corresponding X value.

What I mean by that is, that let us say you have observed stream flow data which are X observed maximum annual maximum stream flow data, which are given by X; and then, from the observed series you would have estimated alpha and beta from these expressions. So, the parameters are estimated and therefore, you have used this particular transformation and fix the return period for which you want the corresponding value.

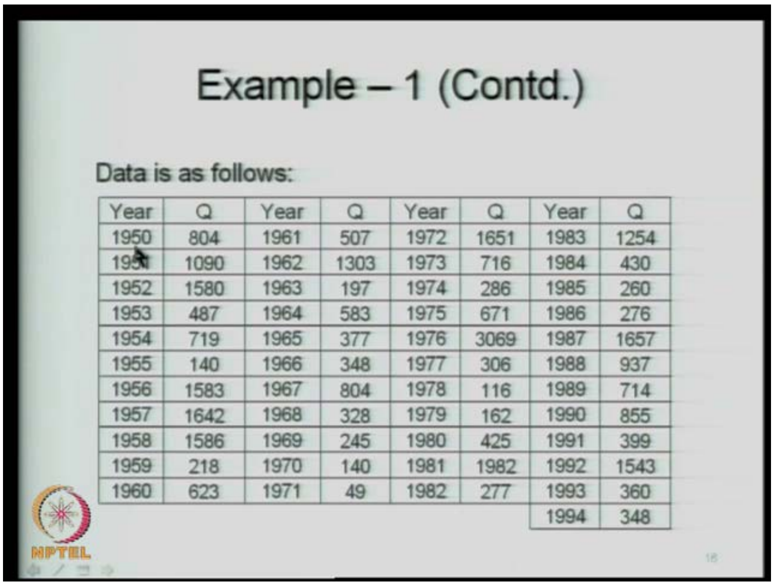
Let us say, you want a value corresponding to 100 year return period, then corresponding to the 100 year return period, you get y, this is T is fixed as 100 year and therefore, you get y, once you get y you can get x from this, because alpha and beta are known. So, this is how we obtain the magnitudes corresponding to a given return period, when we are using the extreme value type I distribution.

(Refer Slide Time: 26:25)



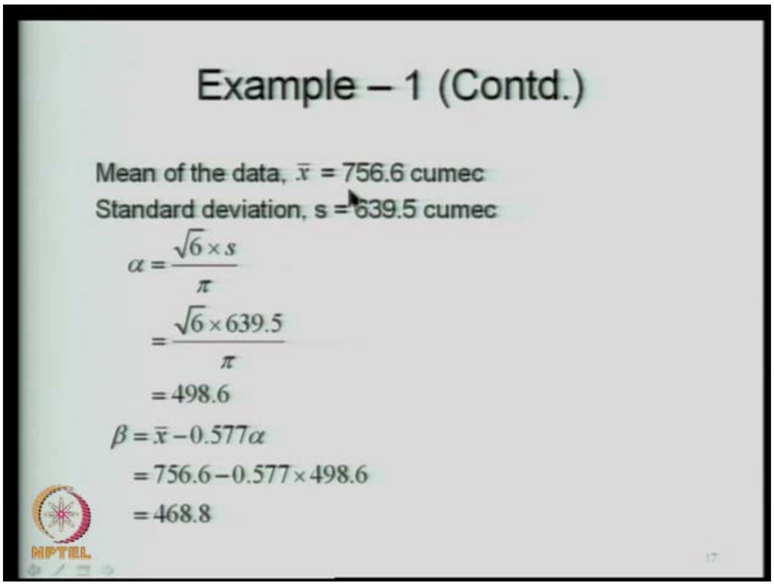
Let us, carry out a simple example now, the same 45 years of maximum discharges that I considered in the last lecture will take it now and then, we will first formulate a model in the sense, we will formulate the c d f expression for the c d f, based on the parameters; and then from that, we will calculate the 20 year and 100 year return period discharge values, that is the discharge values corresponding to return periods of 20 years and 100 years.

(Refer Slide Time: 27:02)



So, this is the data from 1950 to 1994, there are 45 values here. The Q that is shown here is the annual maximum discharge in cumecs, all these values are in cumecs. Then, what we do is? We first calculate the mean and the standard deviation of this data.

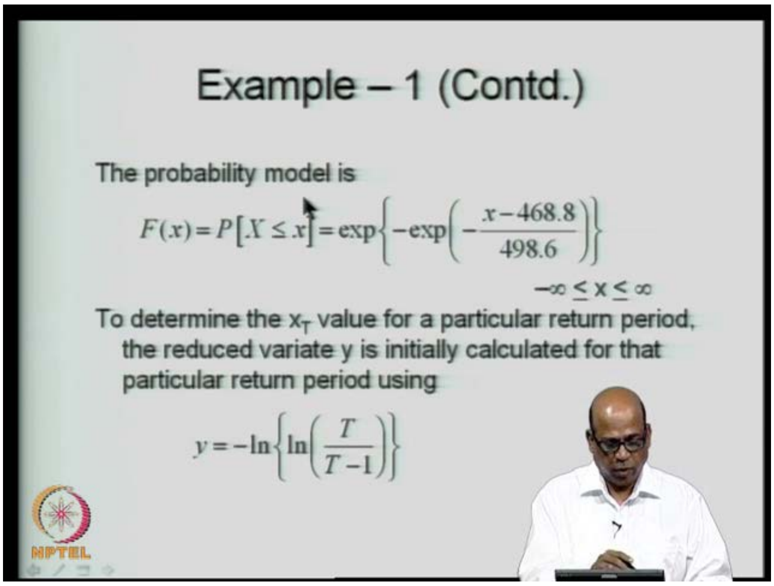
(Refer Slide Time: 27:35)



So, the mean comes to be 756.6 cubic meters per square second and the standard deviation is 639.5 cubic meters per square second with these, we calculate alpha, which is root 6 s by pi; and s is 639.5 therefore, alpha comes out to be 498.6. Then, we calculate beta x bar is 756.6, s is 639.5 and this beta is given by x bar minus 0.577 alpha and that is what you get here (Refer Slide Time: 28:14), alpha is 498.6, so you get beta as 468.8.

So, once you have fixed the parameters, once you have determined the parameters alpha and beta, the F of x or the c d f of the extreme value type I distribution can be directly written and that becomes a probability model for us.

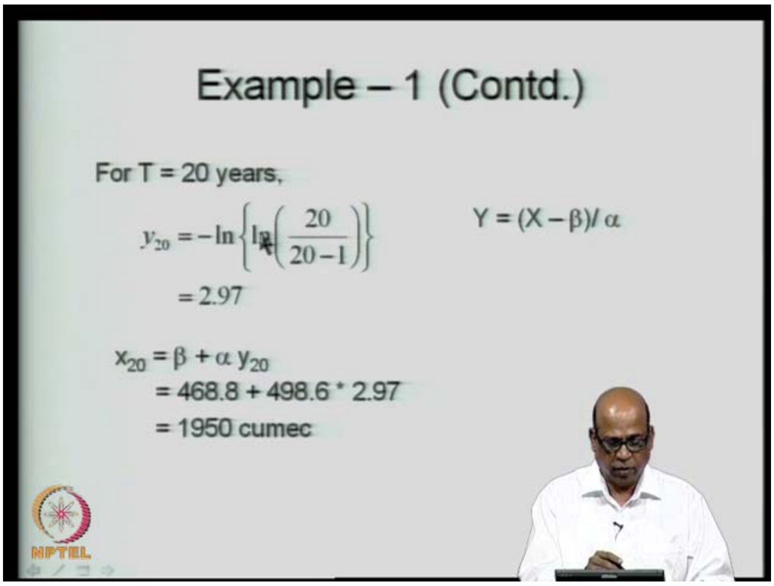
(Refer Slide Time: 28:36)



So, the probability model is F of x which is given which gives probability of X being less than or equal to x is equal to exponential minus exponential minus x minus 468.8, which is beta, x minus beta over alpha and this is defined for x ranging from minus infinity to plus infinity. Then, we go to the transformation y, so this is a probability model.

Remember, this is probability of X being less than or equal to x and what we would be generally interested in when you are using extreme value type I distribution is, probability of X being greater than or equal to x and that is given by 1 minus F of x; and using that, we have obtained this relationship, y is equal to minus l n l n T by T minus 1, where y is the transformation X minus beta over alpha. So, for a given T we calculate y. Let us say, you are interested in in this particular example we are interested in getting the stream flow corresponding to the return period of 20 years.

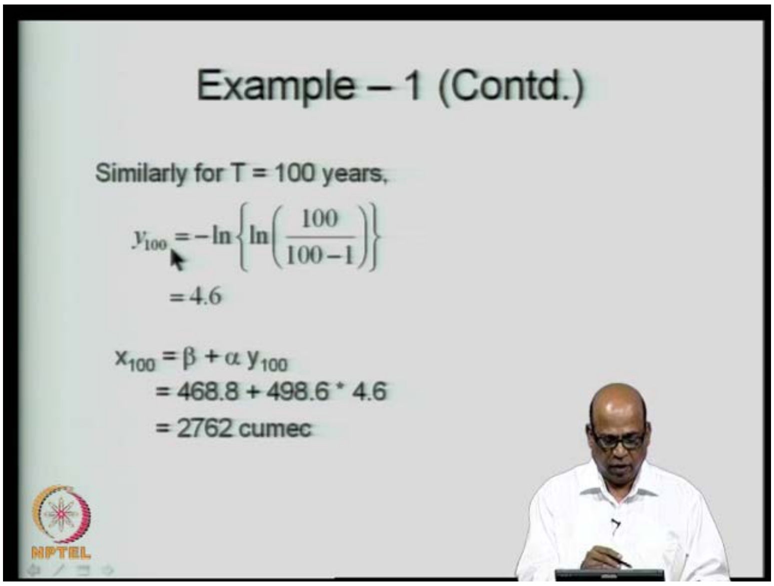
(Refer Slide Time: 29:53)



So, for T is equal to 20 years we just substitute T is equal to 20 years here, 20 here. So, y 20 is minus l n l n 20 by 20 minus 1, which is 2.97, where y is X minus beta over alpha. Once you get y, we use this expression to get x 20. So, x 20 will be simply equal to beta plus alpha into y 20, this case that becomes 1950 cubic meters per second.

So, for the data that we have here you just look at this (Refer Slide Time: 30:28), we are getting a value of 1950, 1950 for a return period of 20 years, there is 1 value which is more than that (Refer Slide Time: 30:38), other than that, you do not see there is another here (Refer Slide Time: 30:44), other than that you do not see values, that have exceeded this in the record 1950. So, 1950 cumecs is a value that you can expect at this particular sight to occur, once in above 20 years in a long period of time. So, this is a interpretation of the flow corresponding to a return period of 20 years.

(Refer Slide Time: 31:21)

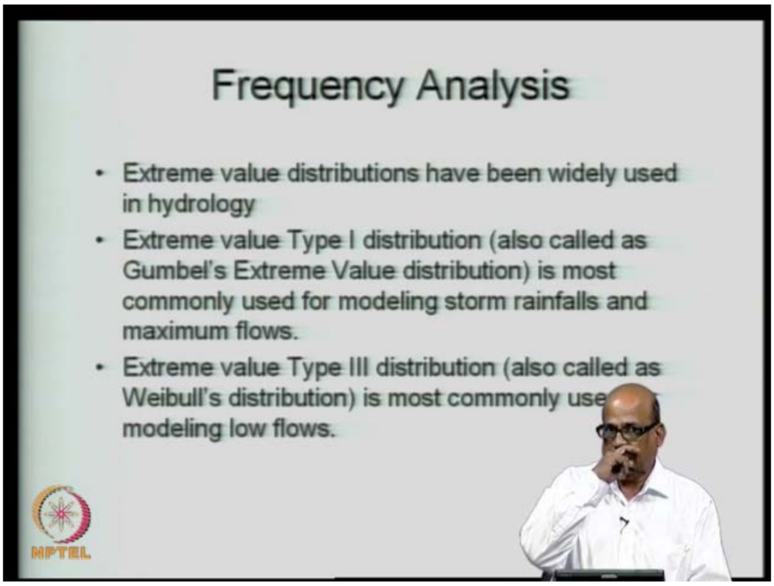


Similarly, for T is equal to 100 years, we use a same method and get y 100 as 4.6 by putting T by T minus 1 and taking the logarithms double logarithms, you get y 100 is equal to 4.6. And the corresponding x 100 will be equal to beta plus alpha into y 100, beta and alpha remain the same and you get this as, 2762 cubic meters per second. Between 20 year return period and 100 year return period, there is a significant difference, 1950 per second is what you get for 20 years; and for 100 years, and you get 2762 cubic meters per second.

From the data of 45 years now, we are now able to say using the extreme value type I distribution we are now able to say that, if we are using that particular data to construct a reservoir which has to which has to pass over a 100 year return period flow; then, this is the value that you have to use for the design of the spillway; that means, the spillway should be designed to pass a discharge corresponding to 2762 cubic meters per second; which may not have occurred in the data at all, although in one case it has occurred.

But, even if it has not occurred in the data at all, from the extreme value type I distribution, we should be able to assess or estimate the value for which you have to design the spillway. If it has to pass the 100 year return period flow and that is the idea of the frequency analysis.

(Refer Slide Time: 33:09)

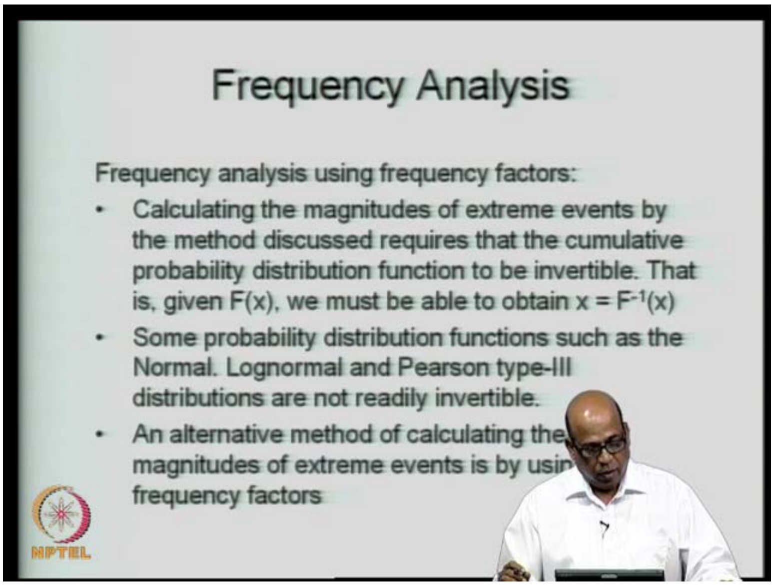


Extreme value distributions as I said, I have been widely used in hydrology, especially for floods and low flow distributions and so on. Typically, we use extreme value type I or the type III distributions. The type I distribution as I just demonstrated are commonly used for high flows, flood peaks maximum discharges, maximum temperatures and so on, maximum annual temperatures and so on.

Whereas, the extreme value type III distribution, which is also called as the Weibull's distribution is most commonly used for modeling low flows or low values. And low flows as you know are important for drought analysis or in the case of water quality stream water quality and so on, we will be using the low flow analysis. So, it is important for us to assess low flows corresponding to certain return periods. The extreme value type I distribution is also called as a Gumbel's extreme value distribution.

A concept of general extreme value distribution also exists, where you have an expression, which encompasses all the extreme value distributions that we have, we generally deal with, that is called as a general extreme value distribution I will not recovering that in the in this particular course. But, remember that, this different types of extreme value distributions that we talked about, type I, type II, type III etcetera, they can all be combined into some general expression to give a general extreme value distribution.

(Refer Slide Time: 34:48)

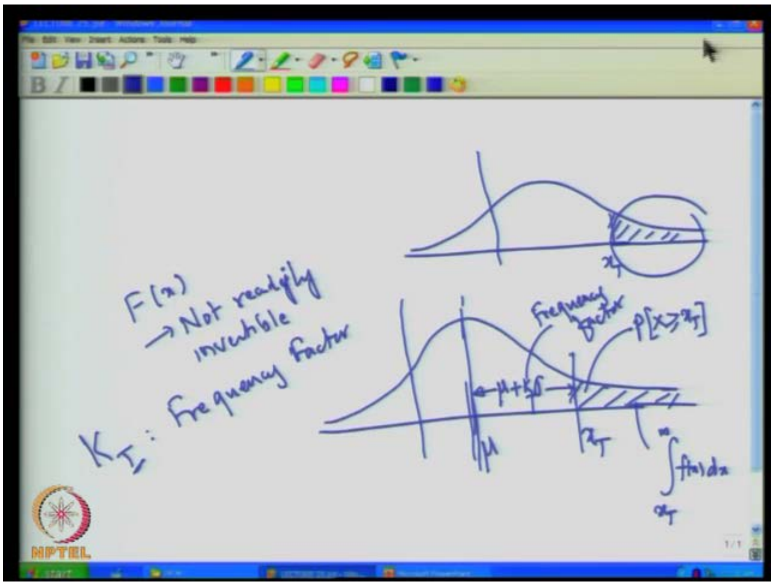


Now so far, what we did in the frequency analysis? We knew the distribution and from the distribution, we could invert in the sense that, given F of x we we would be we would be able to express x as F inverse F of x and then, carried out the frequency distributions frequency analysis.

However, it is not always possible that, the c d f's are invertible in the sense that, you may not be able to express always x as F inverse of x. The form of F of x may be such that, you would not be able to readily invert the F of x to obtain x. Say for example, probability distribution such as normal distribution, log normal distribution, Pearson type III, log Pearson type III distributions, these are not invertible; however, they are most commonly used in many of the hydrologic analysis for frequency analysis.

And therefore, we developed an alternate method for calculating the maximum magnitudes of extreme events, and this method is by using frequency factors, what we call as frequency factors what we call as frequency factors.

(Refer Slide Time: 36:33)



What is the logic in this? Essentially what we do there is, that we will use the distribution the parent distribution like this, for example. Let us say, we are interested in x T let us say, we are talking about normal distribution and then, you are looking at high extreme values x T and because, we cannot get x as a inverted value from F of x; that means, because capital F of x in the case of normal distribution, is not invertible not readily investable I will say, not analytically invertible in in any case, not readily invertible.

What we do is, that, we take this extreme value as a deviation from the mean. So, we write this as, mu plus k into sigma, (( )) k time sigma, this is mu here, this is a mean and this is x T (Refer Slide Time: 37:31). And we are interested in getting to this value that is we are defining this as an extreme, so we write this as mu plus k into sigma. Now, x T depends on T which is a return period, so we write this as, mu plus k t into sigma.

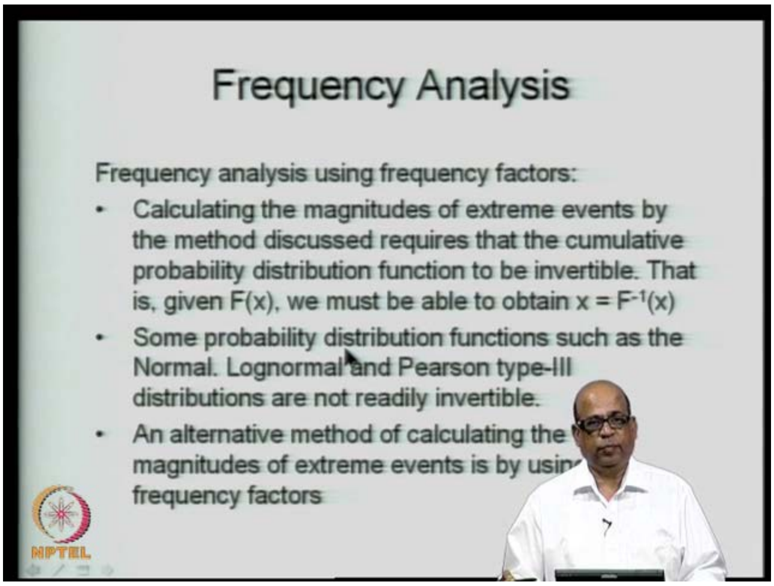
And we are interested in the probability that X being greater than or equal to x T. So, essentially we are interested in this area of the parent distribution. What we write this as, this is probability of X being greater than or equal to x T; and this is given by from our fundamentals x T to infinity F of x d x, that is the area under the curve from x T to infinity.

So, essentially we are writing this extreme value as equal to mu plus sum factor k T into sigma and these factors are called as the frequency factors. So, k T we denote it as a frequency factor. As it is obvious as your T changes or the return period changes, your x

T changes, mu and sigma remain the same, your x T changes therefore, your frequency factors changes, because you are talking about the deviation from the mean. So, k T is a function of the return period, and that is why we put this subscript T, it is also noteworthy that, k T is also a function of the probability distribution itself.

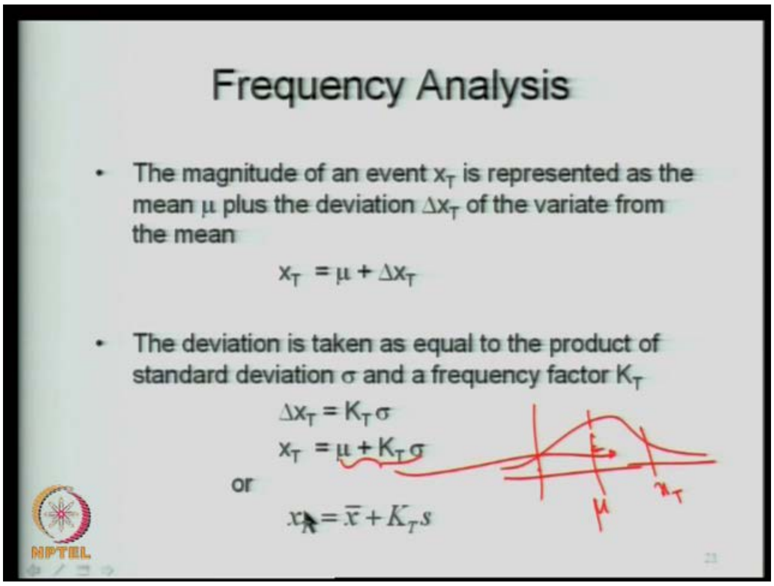
For the same x T value let us say, if your mu and sigma changes, k T will also change. So, k T is the function both of the return period and therefore, we put T and of the probability distribution itself. So, for a given probability distribution, you should have the factors k T for different return periods and we use this concept of frequency factors, to get the magnitudes corresponding to given return periods.

(Refer Slide Time: 40:10)



So, that is what we do in the method of frequency factors. So, we use the frequency factors and determined the magnitudes.

(Refer Slide Time: 40:28)

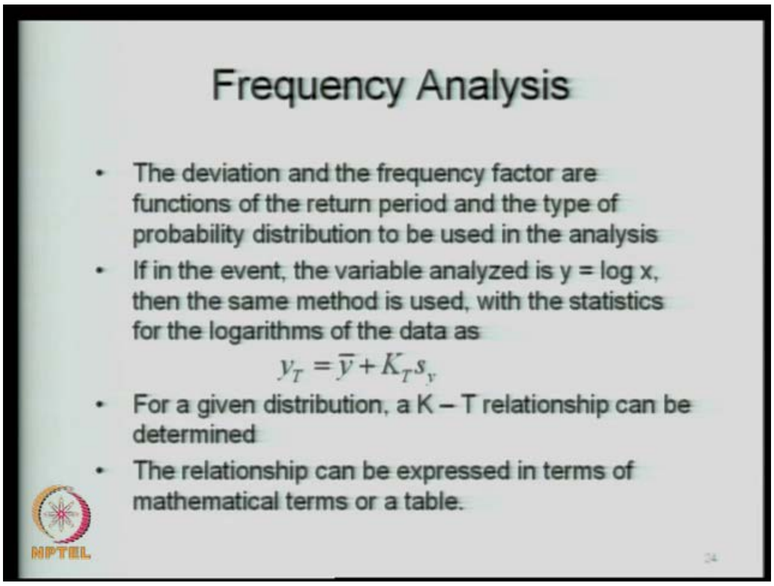


So like I said, we essentially express x T as mu plus sum delta x T, this is the deviation and this delta x T we write it as, k T into sigma I will just explain it again for completeness sake. So, what we are doing is, for any given distribution not necessarily normal distribution, we are writing x T here and your mean may be somewhere here mu. So, we write this as k T into sigma, so this is k T into sigma.

So, x T we write it as, mu plus k T into sigma and then, for sample values you use x T is equal to x bar plus k T into s. So, knowing x bar and s and if you know k T, we should be able to get x T. Now, how we determine k T for different distributions is a question that we will address now. So, essentially the method of frequency analysis using the frequency factors, (( )) round to this expression x T is equal to x bar plus k T into s, s and x bar are estimated from your data; and k T is known or can be derived or is available for a given distribution, let us say normal distribution we have the k T for a given value of T. Log normal distribution, we can derive k T for a given value of T and so on.

So, given the distribution you know k T values and therefore, we know we will be able to get x T, what is x T again I have to repeat, x T is the magnitude of the particular variable corresponding to a return period of T.

(Refer Slide Time: 42:27)



So, the derivation I am sorry the frequency factor is as I said is a function of both the probability distribution the parent distribution as well as, of the return period. So, as the return period changes, your k the frequency factor changes, as the probability distribution changes, the frequency factor changes.

Now, let us say from your original data from stream flow data let us say, you are converting it into logarithmic flows. If you are interested in the log transformed data, you can still use the same expression in terms of y T now, y T is equal to y bar plus K T into s y, where y is the log transformed data series. So, you get from the log transformed data series, the K T values remain the same and you get, s y and y bar from the log transformed series; and you obtain y T, from y T you should be able to get K T again back.

As I mentioned for a given distribution, the k values the frequency factor values should be determined corresponding to several desirable T values. So, the K T relationship can be determined for a given distribution, there are methodologies available we will not go into those methodologies, but we will use the results from such methodologies for some specific distributions.

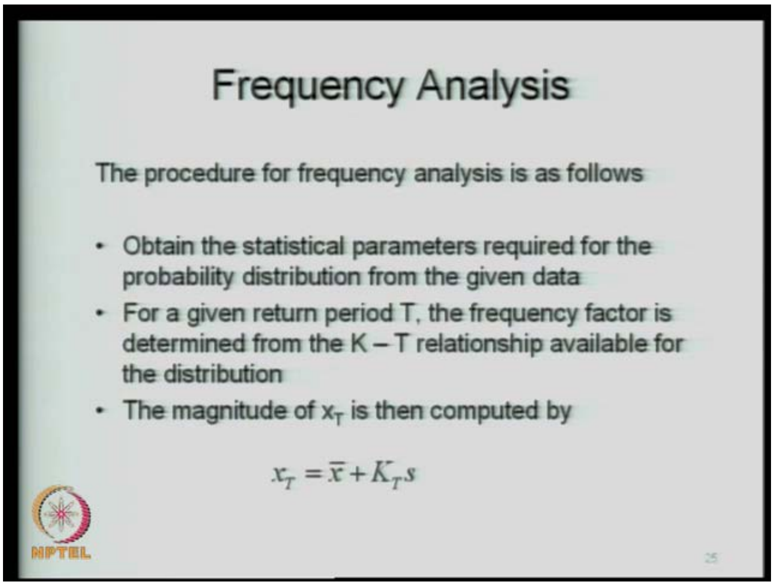
Now, this relationship the K T relationship is fundamental to the frequency analysis using the frequency factor. So, for any distribution, we should have this relationship; that means, given a T, given a return period for that particular distribution, we should be able to get the corresponding K value. We will see, how this can be done for the normal distribution presently.

So, essentially what we do is, from the historical data we obtain the moments that is a mean and a standard deviation; and from the available K T relationship for that particular distribution, we obtain the K T value and we use these K T values to get the x T values that is that is all that is there to the frequency analysis.

A major question then is, which distribution to use? Do we use extreme value type I distribution, in which case we know that, it can be done analytically. Do we use the normal distribution? Do we use the log normal distribution, log Pearson type III distribution and so on. It is there that some judgment is involved and also there are analytical methods of estimating, which distribution best fits the particular data, which will cover in subsequent classes.

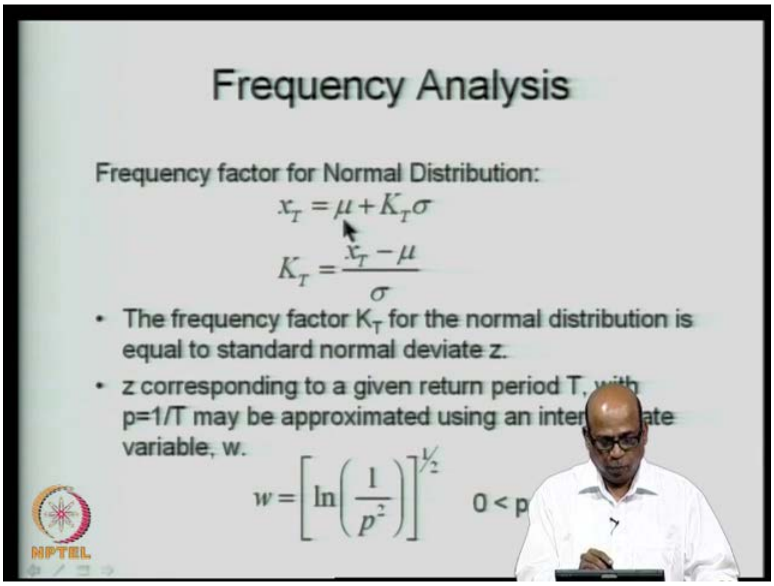
But typically, if we want to use a given distribution, we should be able to get the frequency factors, if that distribution is not invertible; if it is invertible, we can always use the analytical method.

(Refer Slide Time: 45:35)



So, we will just se now, how we obtain for example, the frequency factor for normal distribution. The expression that we are using is, x T is equal to mu plus K T into sigma.

(Refer Slide Time: 45:50)



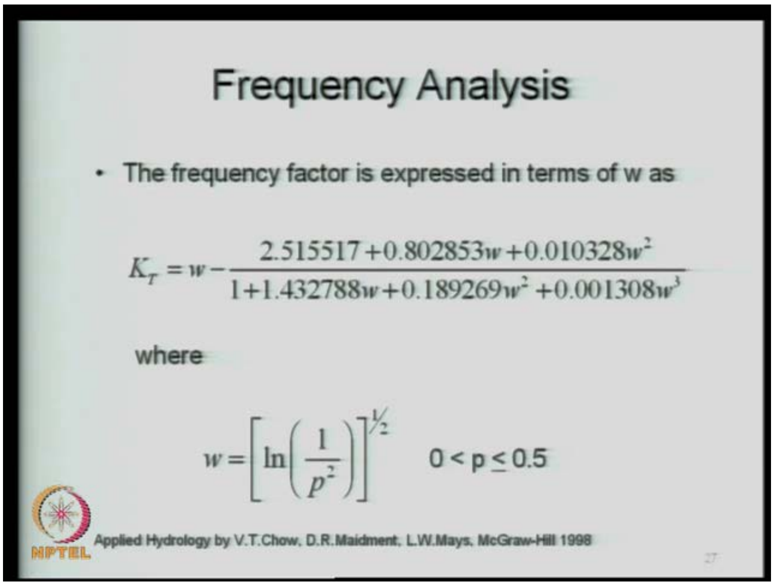
So, x T is equal to mu plus K T into sigma, is our frequency analysis expression. From this, we can write x T is equal to K T is equal to x T minus mu over sigma. And we identify this as, simply the standard normal deviate z, recall that we wrote that z is equal to x x minus mu over sigma, so this is simply the z value.

So, you have the option of simply going to the table and picking up the corresponding z value, given x T minus mu over sigma, you can obtain the z value corresponding to those particular probabilities. What I mean by that is? Given the return period T, you know the probability p is equal to 1 over T; and from the probability, you know the z value and you know the F of z value; and from the F of z value you enter the table of normal distribution, and obtain the z value.

As we have done in our normal distribution calculations you can refer to one of the earlier lectures, where we discuss the normal distribution. However, there are also empirical expressions available for obtaining the z value corresponding to a given probability p we will use that here, but I again repeat that, this is not necessary you can always use the tables normal normal value tables, once you get the p values.

So, once you fix the return period T, you can get p, which is p is equal to 1 by T. The z corresponding to a given return period, with p is equal to 1 by T may be approximated using an intermediate variable, w. So, we define w as log 1 over p square the whole raised to the power half for values of p between 0 and 0.5.

(Refer Slide Time: 48:03)

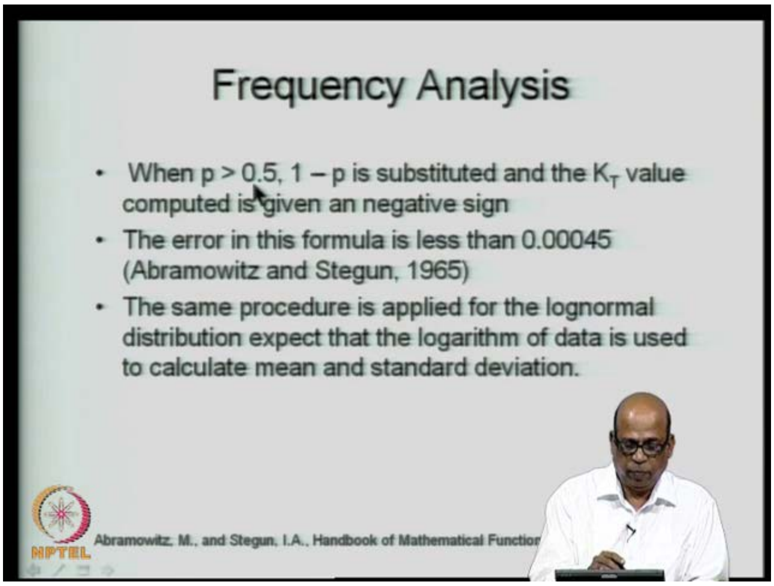


We use this intermediate value w to obtain the K T value in fact, this is an expression for z directly, for the normal distribution this expression that I am writing it for z value here (Refer Slide Time: 48:14), and z is the same as K T in this case therefore, I write it here for K T.

So, we define first w, knowing p, how do we know p? Because, we have fixed the return period T, from p is equal to 1 over T, we get p; once we know p, we can get w. As long as p is within the range 0 to 0.5 in most of the cases it will be. If it is not, we will see what to do, so once you get w, you can get K T. Now, these are slightly cumbersome looking numbers, but they are obtained from some empirical relationships. So, you get from w, you can get K T, once you get K T, you can get x T, because s and x bar are known.

We said, this is valid for p is equal to 0 to 0.5, if p is more than 0.5 then what we do is? Instead of of p, we substitute 1 minus p here in this expression (Refer Slide Time: 49:23), let us say p is equal to 0.6, then we use 0.4 here 1 minus p and then, still use w and then from w, we get K T; and this K T we use it as in negative value with a negative value.

(Refer Slide Time: 49:40)



So, when p is greater than 0.5, 1 minus p is substituted here for w, we use that w resulting from 1 minus p and obtain K T, but in the expression for x T namely, x T is equal to mu plus K T into sigma we use K T as a negative value. And therefore, we write it as x T minus K T into sigma, when p is greater than 0.5. When will p be greater than 0.5, when T is very small, because as T increases, p decreases. So, when T is small, you may get a situation where p is greater than 0.5. So, that is how we address the case, where p is greater than 0.5.

The error in this formula is very small of the order of zero 0.00045. So, this is the error in z that we are talking about, that is the standard normal deviate as obtained from this expression, that error is extremely small as has been shown by these authors here and that is the reference.

So for the normal distribution, estimate x bar and s from the data and then, for a given return period T, you estimate you obtain the probability p is equal to 1 over T; and from that, you obtain the w value which is the intermediate value, from w value you can get the K T value. And once you get K T, you know x T is equal to x bar plus K T into sigma, K T into s, if you are dealing with samples and therefore, you can get x T.

The same procedure we can also use for log normal distribution. So, the straight a straightforward way of doing it is, simply convert all your values into logarithm transformed values and then, use the same principle, as we have done for the normal distribution.

So, before we go into the example now, we will just try to summarize what we covered in this particular lecture. So, essentially we started with the introduction of frequency analysis and we saw that, the data series that we consider for the frequency analysis can be constructed in different ways; one is to consider the entire series or the other one is, to put a threshold value and take pick up only those values, which are above the threshold value, if you are looking at high extremes and the third one is, you put the threshold value such that, you will get those many values above that particular threshold, as the number of years of data itself, that is called as annual exceedence series.

The annual exceedence series is commonly used for frequency analysis, but the annual exceedence series as I mentioned earlier on (( )) from the fact that, the independence of the data cannot be ensured always or you you may not have a way of ensuring that, the data series that you have constructed consists of values, which are all independent of each other.

The most commonly used series is the annual maximum or minimum series, where for every year you pick up exactly one value, which may be the maximum value or the minimum value, depending on your need, depending on the analysis that you are carrying out. Then, once you know once you have conducted the series, then you can go with the extreme value distributions, extreme value type I distribution, type II or type III distribution in hydrology most commonly we use extreme value type I or the extreme value type III distributions.

If you are looking at the maximum values or the high extreme values, you typically use extreme value type I distribution, which is also called as the Gumbel distribution. From the c d f of the extreme value type I distribution, it is possible for you to obtain x T, which is the magnitude of that particular variable, corresponding to a return period of T, because extreme value type I distribution is invertible. Similarly, we use the extreme value type III distribution, when you are looking at low extremes for example, low flows low rainfall values and so on. The extreme value type III distribution is also called as the Weibull's distribution.

There are certain distributions, which are not invertible and which are also most commonly used in hydrology. For example, the normal distribution, the log normal distribution, log Pearson type III distribution etcetera. Now, for these distributions we use the method by method of frequency factors and that is why what we introduced in this particular lecture.

The frequency factors, the basis for the frequency factors is that we take x T as a deviation from the mean. So, we write x T is equal to mean plus sum factor k into sigma. So, it is deviating from the mean from by an amount of k into sigma, and k are called as k the factors k are called as the frequency factors. Frequency factors are a function of the return period T as well as, of the parent distribution itself.

So, for a given distribution we should have a K T relationship, which means for a given return period, we should be able to get K. And we saw that, for the normal distribution the K T just corresponds to the standard normal deviate z. So, once you fix the T you are actually fixing the p and there is probability of X being greater than equal to x T; and therefore, from this you can go to the normal distribution tables and pick up the z values. There is also an empirical relationship that, I just introduced from which you can get K T value, so once you get the K T value, you can get the corresponding x T value.

The whole idea of the frequency analysis is to obtain x T values, x capital T for a given capital T, which is the return period we are interested in getting the magnitude corresponding to that particular return period return period. And for different distributions, we obtain the K T values and obtain x T values.

So, we will continue this discussion in the next lecture, we will start off with an example of the normal distribution, numerical example which will drive home the points that I have been discussing in this particular lecture, thank you very much for your attention.