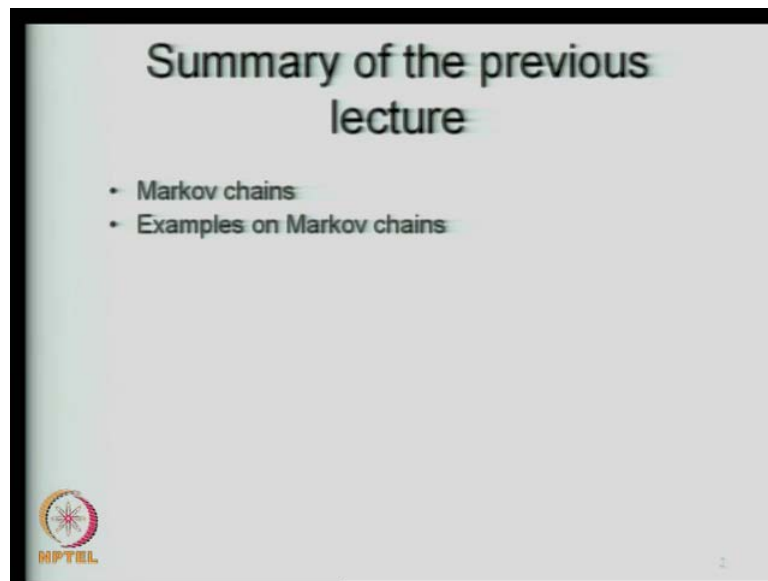


Stochastic Hydrology
Prof. P.P.Majumder
Department of Civil Engineering
Indian Institute of Science, Bangalore

Module No # 06
Lecture No # 24
Frequency Analysis- I

Good morning, and welcome to this the 24th lecture of the course on, “Stochastic Hydrology”. In the last class that is the lecture number 23, we essentially discussed the Markov chains.

(Refer Slide Time: 00:33)



And in the previous class to that we introduced the Markov chains, the transition probabilities and then we also discussed steady state Markov chains and understood how to derive the steady state probabilities of the Markov chain. Then in the previous lecture we also discussed a few examples on Markov chains.

So far, the journey has been on discussing the fundamentals on stochastic processes with a few applications through numerical examples and so on. For example, we also discussed in earlier classes the probability distributions, joint probabilities, conditional

probabilities and then data generation, data forecasting, analysis in the frequency domain, spectral analysis, and then the time series analysis, where we discussed in detail the ARIMA type of models and then their applications, tool data generation, hydrologic data generation as well as hydrologic forecasting.

Now onwards, the focus of the course will slightly shift in the sense that we will take specific applications of many of these techniques that we have learned, for various problems that we encounter in hydrology. Be it the flood forecasting or determination of design storms or run links of drought analysis and so forth. So, now onwards, we will specifically pick up few topics of direct application, relevant to hydrology, and then discuss how we use the theory that we have learned so far, on their applications; on the applications that we would be interested in.

Now, some of these may have been covered also in your applied hydrology course or the basic hydrology course. For example, we may start with flood frequency analysis today, which may have been covered in your basic hydrology course, but still, because in as much as these topics, use the theory that you have covered in the stochastic process. Of course, we will still cover even at the risk of repetition of some of these topics, so that the fundamentals of the theory that we have covered in their applications will be clear.

We will start today **with the frequency**, with the topic frequency analysis. Right at the outside, I would like to mention that we will be talking as far as this course is concerned of at site frequency analysis. In the sense, that we are not talking about special distribution of the data or the regional frequency analysis, which is a different topic all together, this is slightly advanced. But, we will be discussing in today's class and may be subsequent two to three classes, will be discussing, the frequency analysis of the data that is gathered at a particular location.

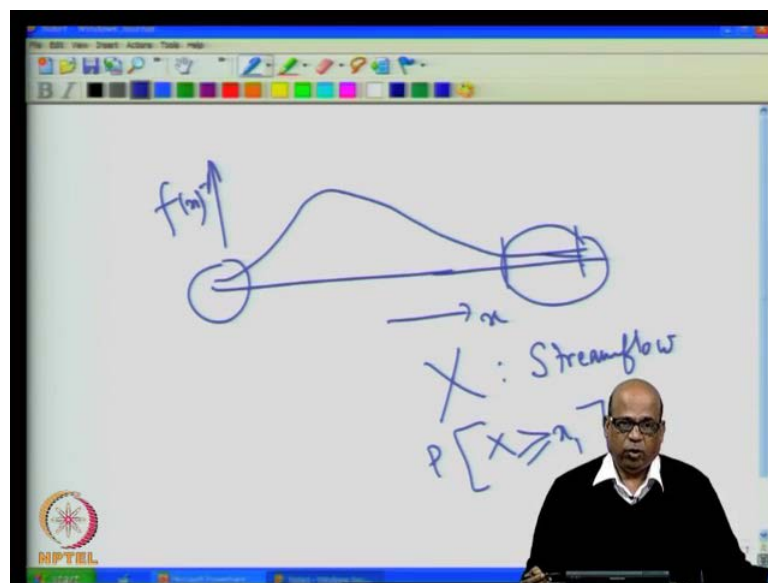
This we call it as at site frequency analysis. Say, for example, we may have a rain gauge and we have been continuously recording the rainfall at a particular location. That type of data will be using for frequency analysis or the stream gauge data and are recorded at a particular stream location the stream gauge location.

What do we mean by frequency analysis? Most of the hydrologic problems are characterized by occurrence of extreme events. For example, we may be discussing about floods, which are really high extreme flows or maybe we are interested in really

low flows, which define either your hydrologic draughts or which may be critical for your water quality concentrations. We may be interested in really higher intensities of rainfall, which are necessary for design of storm water drainage in a city and so on and so forth.

So, we will be interested really in the extreme values, either the high extremes or on the low extremes. So, for a while, now we will depart from our normal processes and then focus on really the extreme processes.

(Refer Slide Time: 05:40)



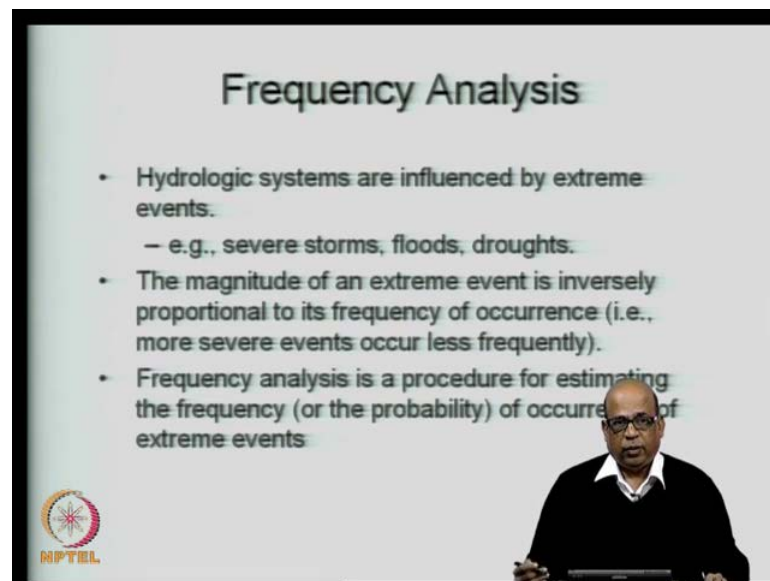
If you take any probability distribution that we have been seeing, so far, let us say we look at some probability distribution like this, either a completely symmetrical or a non symmetrical distribution.

If you look at the extreme high values, let say this is your random variable, the values at the random variable takes and this is your f of x , in our notation. There is a small tail here towards the right, which describes the extreme values. The probability associated with these extreme values are all small, however, if you look at the distribution of these extremes themselves, either on this side or in this side, these distributions will have some interesting characteristics. We will be interested in really looking at the distribution at the lower extreme or the higher extreme of the distribution of the parent population. This we call as a parent population.

So, we are deriving the extreme values or the distribution of the extreme values from a given parent distribution. What I mean by that is? Let us say your random variable X is your stream flows at a particular location. And we may be interested in particular values of stream flows or we may be interested in the event that X is greater than or equal to a given value of x , let us say X_1 or some such thing.

And we may be interested in probability of X , being greater than and equal to x_1 , for the flood flows. In which case, we would be interested in getting the distributions of the really extreme flows, which are derived from the parent distribution of the stream flow random variable X . So, we derive the extreme value distributions from a given parent distribution. In other words, the frequency analysis that we will be doing, will be a function of, and will be dependent on the parent distribution from which we are deriving the extreme value distributions. So, with that background, let us see what we really mean by frequency analysis and where it is useful.

(Refer Slide Time: 08:23)



The slide is titled "Frequency Analysis" and contains the following text:

- Hydrologic systems are influenced by extreme events.
 - e.g., severe storms, floods, droughts.
- The magnitude of an extreme event is inversely proportional to its frequency of occurrence (i.e., more severe events occur less frequently).
- Frequency analysis is a procedure for estimating the frequency (or the probability) of occurrence of extreme events.

In the bottom right corner of the slide, there is a small inset image of a man with glasses, wearing a dark sweater over a white collared shirt, who appears to be the presenter. In the bottom left corner of the slide, there is a circular logo with a compass rose and the text "NPTEL" below it.

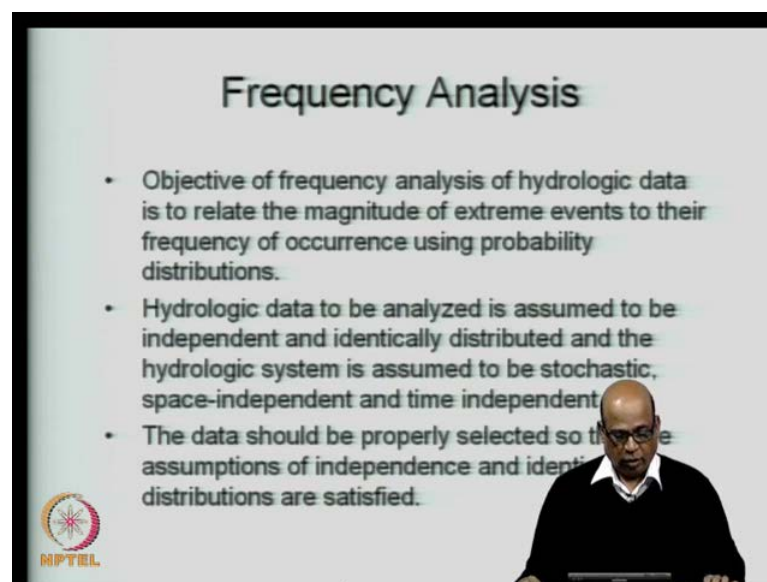
As I mentioned, we may be frequently encountering severe storms, floods, droughts, low water quality situations, high demand situations, low flow situations, low rainfall situations and so on. These are really extreme events, in the sense that they do not normally occur. In fact, the more the severity of the event, the less frequent it is to occur. For example, we may be interested in the highest level of flood that can occur at a

particular location. This may be least frequent compare to other events that we may be interested in that location.

So, the magnitude of an extreme event is inversely proportional to its frequency of occurrence. What do I mean by that? Let us say that we are interested in the magnitude of a flood, which on an average occurs every 10 years. This magnitude will be smaller than a flood event that occurs once in about 100 years. So, as the frequency increases, the magnitude decreases. Once in 10 years is the higher frequency compare to once in 100 years and therefore, the flood magnitude that may occur once in about 10 years will be smaller than the flood magnitude that may occurs in about 100 years.

With this information in place, we carry out what is called as the frequency analysis. Now, the frequency analysis is a procedure of estimating the frequency or the probability of occurrence of extreme events. That is we may be interested in answering the question like in how many years a flood of a given magnitude is likely to occur on an average or what is a probability that in a given year a flood of a given magnitude will occur. These are the types of questions that we would be interest in answering to frequency analysis.

(Refer Slide Time: 11:00)



The slide is titled "Frequency Analysis" and contains the following text:

- Objective of frequency analysis of hydrologic data is to relate the magnitude of extreme events to their frequency of occurrence using probability distributions.
- Hydrologic data to be analyzed is assumed to be independent and identically distributed and the hydrologic system is assumed to be stochastic, space-independent and time independent.
- The data should be properly selected so that the assumptions of independence and identical distributions are satisfied.

In the bottom right corner of the slide, there is a small inset image of a man with glasses, wearing a dark sweater over a white collared shirt, sitting at a desk with a laptop. In the bottom left corner of the slide, there is a circular logo with a star-like pattern and the text "NPTEL" below it.

So, the objective of the frequency analysis is to relate the magnitude of the extreme events to the frequency of occurrences. Let say a magnitude of 50000 cubic meters per second **of flood** of flood discharge. We would be interested in answering once in about how many such an event may occur. So, that is how we relate the magnitude of the

extreme events to the frequency of occurrence. By frequency of occurrence, I mean, once in how many years the particular event is likely to occur.

As I mentioned earlier, we will be doing the frequency analysis at a given site. This is different from the regional frequency analysis where we are interested in the frequencies of these events over a large region consisting of several data points. Let say you may be interested in south India, peninsular region, what kind of flood frequencies are likely to be present in that particular region. When we are doing that we may consider the flow data from several stream gauges located at the entire region. That is the regional frequency analysis. But, in this course, for the time being, we are focusing on at site frequency analysis, where we would be collecting data from given stream gauge or given rain gauge. Then we will be carrying out the frequency analysis for that particular stream gauge or the rain gauge.

A basic assumption in doing this frequency analysis is that the data that we have, let say the stream flow data at a particular stream gauge. This is assumed to be stochastic, it is space independent and it is time independent. Now, for the same rain gauge, we are essentially looking at **these two independent** these two assumptions. Mainly, that the data are independent and they are identically distributed. So, that the data you want to analyze for frequency analysis must satisfy these two conditions for the procedures that we are going to introduce now. This is called as IID. I also used this term earlier, in the case of time series analysis, IID is independent and identically distributed.

Now, how do we ensure independence? Let say, we are looking at really high flows that have occurred in a year. Now, you may be gauging the data or you may be observing the data for every 15 minutes let say, in a day. And you have a continuous monitoring of data for at the intervals of 15 minutes. So, you will have 4 such data per hour and 24 hours and then in a year you will have 4 into 24 hours into 365 days. So, those many data points you have in a given year. Now, if you put simply a threshold. Let say that flood values exceeding a certain amount or stream flow values exceeding a certain threshold value of stream flow.

Let say 10000 cusecs or some such things. Then in a given year you may have several such values, which all exceed a given threshold of 10000 cusecs. If you collect such values then it not is really independent, they may not be independent. For example, a

flood that occurs in a given year may have saturated the soil completely and therefore, any rainfall that occurs subsequently may cause a flood of a high magnitude, any high rainfall that occurs subsequently may cause a flood of higher high magnitude. Therefore, both of these floods that I am talking about may cross the threshold value that we have specified and therefore, they are not independent events.

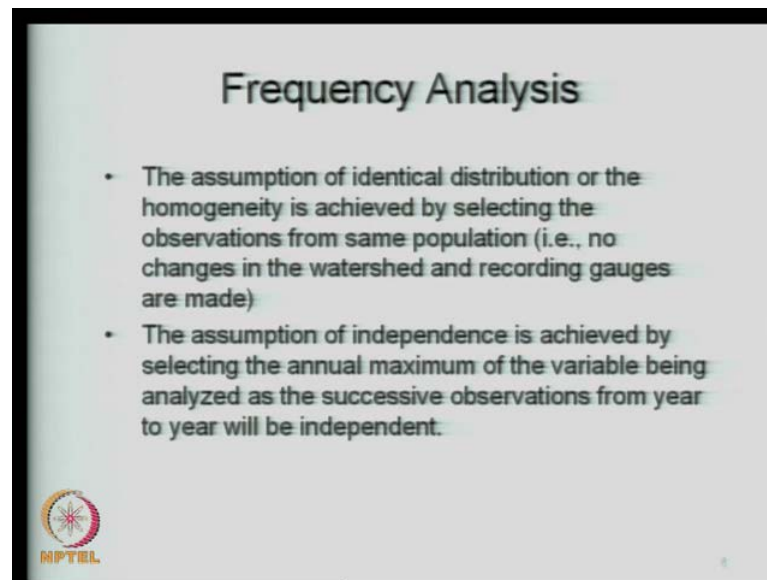
However, if you choose the maximum value that has occurred, maximum flood that has occur in a year, which means out of all these values that you collected, 4 values per hour into 24 hours per day into 365 days. Those many values and out of these values you pick up just one value, which is the highest value that have occurred in that particular year. Like this, for every year, you have one value, which is the maximum value that has occurred in that particular year.

If you choose the annual maximum values, then this can be reasonably assumed to be independent of each other. For example, in one year the maximum value may have occurred during July. In another year, it may have occurred during October, and these two may not have any relationship at all. And therefore, they can be safely assumed to be independent of each other. Therefore, when we are choosing the data for the frequency analysis it is important for us to choose that particular type of data for which the assumption of independence can be made reasonable.

Then what do we need by identical distribution. Remember, we are talking of the same random variable here. We are not talking about x and y being independent of each other. We are talking about, let say flow at a particular location, and therefore, we have a time series, time series of flow. So, in that sense, identical distribution would mean time homogeneity; that means, you have a time series and then this should be homogeneous among time. Contrast this with space homogeneity, which will talk about on regional frequency analysis. We are not interested in space homogeneity in as per as frequency analysis of at site data is concerned. we are talking about time homogeneity.

So, we select that particular data such that the assumptions of independence and identical distributions are satisfied. These are the two i most important assumptions that we make in frequency analysis.

(Refer Slide Time: 18:22)



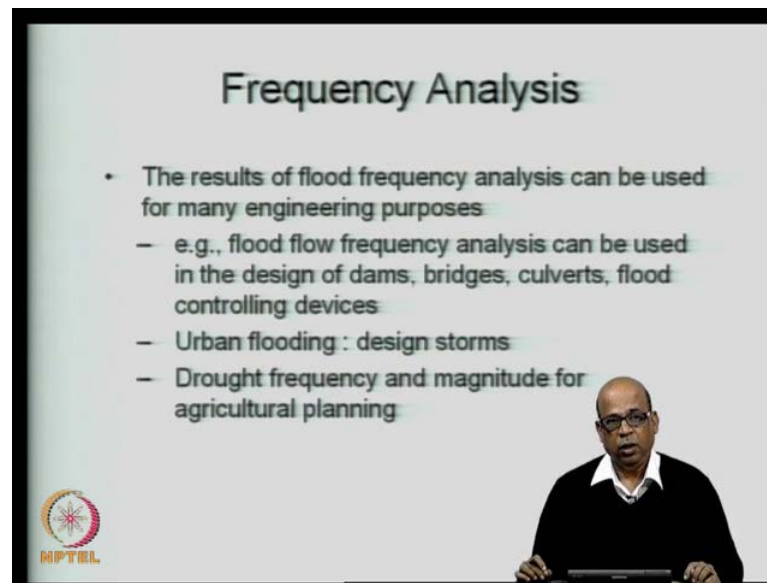
The slide is titled "Frequency Analysis" and contains two bullet points. The first bullet point states: "The assumption of identical distribution or the homogeneity is achieved by selecting the observations from same population (i.e., no changes in the watershed and recording gauges are made)". The second bullet point states: "The assumption of independence is achieved by selecting the annual maximum of the variable being analyzed as the successive observations from year to year will be independent." In the bottom left corner, there is a circular logo with a star-like pattern and the text "NPTEL" below it.

As I said the assumption of independence is achieved by selecting the annual maximum or minimum of the variable being analyzed, as successive observations from year to year can be assumed to be independent. The assumption of identical distribution or time homogeneity is achieved by selecting observation from the same population. What I mean by that same population is that the data that you are analyzing, must have come from the same water shed, which has not gone through any drastic changes in recent years or during the period of the data.

For example, sudden urbanization has not taken place or the characteristics of the watershed have essentially remained unchanged during the period of the data. Then also the gauge that you have used for the data, they have not undergone any changes. For example, the type of gauge has not changed or the gauge has not been shifted from one place to another place in the same watershed and so on.

So, we are looking at a same type of rain gauge, the same rain or stream gauge for the entire period of data and the watershed has not gone any specific changes, during the period of data. This will ensure that the time homogeneity is maintained for the data.

(Refer Slide Time: 19:58)



The slide is titled "Frequency Analysis" and contains the following text:

- The results of flood frequency analysis can be used for many engineering purposes
 - e.g., flood flow frequency analysis can be used in the design of dams, bridges, culverts, flood controlling devices
 - Urban flooding : design storms
 - Drought frequency and magnitude for agricultural planning

In the bottom left corner, there is a circular logo with a star-like pattern and the text "MPTEL" below it. In the bottom right corner, a man in a dark sweater and glasses is visible, appearing to be presenting the slide.

So, once you have ensured this, we will look at the purposes, for which we will use frequency analysis results. As I mentioned, we would be getting answers to questions such as what is the return period or the average recurrence time that elapses between a critical events occurring in two successions, occurring successively. Let say that these year a flood of a given magnitude have occurred, then we will be interested in what is the probability of it will occur again next year or will be interested in the average time that may elapse between occurrence of this event successively. Does it occur once in five years, once in ten years? and so on.

Now, these kinds of questions are of extreme relevance or high relevance in designs of dams, bridges, culverts and flood control devices and so on. When we build a dam, we may want to design it for a flood of a return period of 100 years; that means you want to build a capacity, which can pass the discharge that corresponds to a flood, which returns once in about 100 years or once in about 200 years and so on.

So, right at the design stage, you will make a decision on what is a type of flood discharge that you would like to pass through this particular spill way and that discharge corresponds to your design decision of the return period. And the decision on the return period itself will depend on the type of designs that you are making.

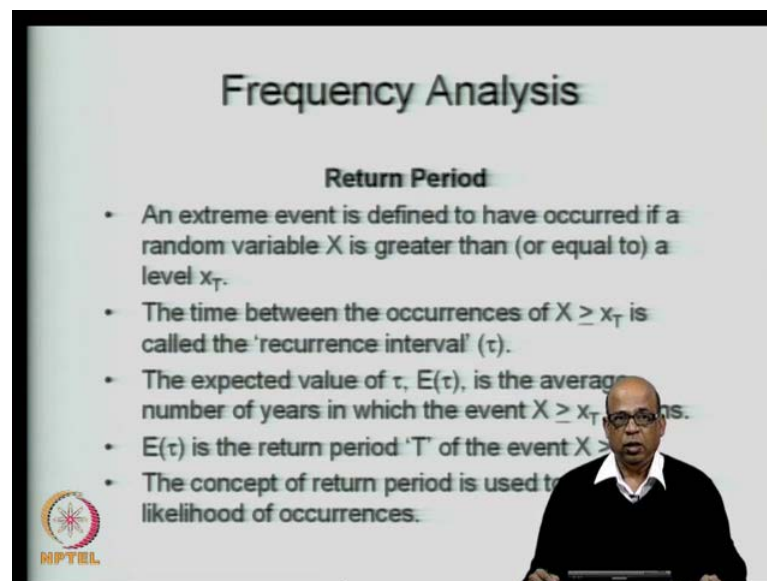
For example, for the dams, bridges, culverts, etcetera you may be interested in 100 years return period, whereas, the flood controlling devices, you may be interested in more

frequent events, for example, once in 10 years, once in 15 year and so on. And when we come to urban flooding, your designs are based on much more frequent events and therefore, lesser discharges, smaller discharges. For example, urban storm water drainage, we may be interested in return periods of once in 5 years, once in 10 years and so on.

So, these are the return period, in the sense that the magnitudes of the discharges or the rainfall intensities that we talking about will appear once in about those many years on an average. We may be also interested in drought frequency and magnitude or we may be talking about meterologic drought or hydrologic drought. Meterologic draught are related to rainfall essentially and then the hydrologic draught are related with the soil moisture and water availability, in terms of the stream flow and the reservoir levels and so on.

So, we may be interested in the lower end of the distribution time; that means, the low stream flows, low rain fall and low soil moisture levels and so on. This will be useful in agricultural planning. So, the frequency analysis really provides useful information for planning either for the floods or for the droughts. And in most of the hydrologic designs we use the results provided by the frequency analysis.

(Refer Slide Time: 23:54)



Frequency Analysis

Return Period

- An extreme event is defined to have occurred if a random variable X is greater than (or equal to) a level x_T .
- The time between the occurrences of $X \geq x_T$ is called the 'recurrence interval' (τ).
- The expected value of τ , $E(\tau)$, is the average number of years in which the event $X \geq x_T$ occurs.
- $E(\tau)$ is the return period 'T' of the event $X \geq x_T$.
- The concept of return period is used to estimate the likelihood of occurrences.

MPTEL

So, how do we go about this? I just talked about the return period. Let us go through the concept of return period more formally now. Let say that we are interested in an extreme

event, which we define as the random variable, the event that the random variable X takes on a value greater than or equal to a threshold level x_T . Random variable of stream flow being greater than or equal to a value of 5000 cubic meters per second, discharged being greater than or equal to a value of 5000 cubic meters per second. Now, that is the event we call it as extreme event; that means we put a threshold and then say that any time the threshold is crossed call that as an extreme event.

Then we define the time between the occurrences of X greater than or equal to x_T . That is the event X greater than or equal to x_T has occurred now, and then a time has elapsed and then again it occurs, another time has elapsed and then again it occurs and so on. So, the elapsed time between the two consecutive occurrences of X greater than equal to x_T is called as the recurrence interval. And we denote this as by τ . We denote the recurrence interval by τ .

Then in a long period, there may be several recurrences of this event. X greater than equal to x_T has occurred in let say year number one, it occurred again year number 14, again in year number 15, occurred again year number 29, and so on. So, the recurrence interval keeps on changing in a record of the time series. The expected value of τ is denoted by $E\tau$ is therefore, the average number of years in which the events X greater than or equal to x_T returns. Let say you had a record of 50 years and then you are interested in getting the event or the return period corresponding to an event X being greater than or equal to 5000 cusecs, when we are talking about the discharge. And then in these 50 years of record, the average recurrence interval or the expected value of the recurrence interval may turn out to be, let say 10 years. And that is what we call it as a return period.

So, expected value of τ is the return period T of the event X greater than equal to x_T . This concept of the return period is used to describe the likelihood of occurrences of extreme events. In fact, in many times, we confuse the concept of return period with the exact time at which the event occurs. For example, I may when I say an event has a return period of five years. It does not mean that in the next five years it has to happen or it has to happen once. Let say that five year return period event have occurred this year.

It does not mean that same five year return period will not occur in the next year. In fact, every year has the same probability of this five year return period occurrence, which will



show it to be one by five in the subsequent slides. So, do not confuse the concept of return period with the time that it takes for that particular event to return exactly. It only indicates an average recurrence time; that means, over a long period of time on an average this event will return once in about so many years, which is the return period of the event.

(Refer Slide Time: 28:34)

Frequency Analysis

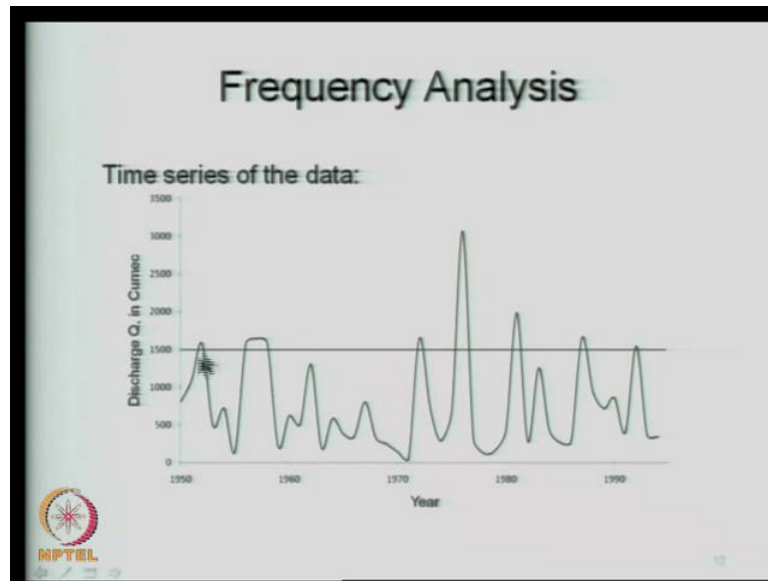
Example – 1: Consider the annual maximum discharge Q, (in cumec) of a river for 45 years

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	116
1956	1583	1967	804	1978	116	1989	162
1957	1642	1968	328	1979	162	1990	425
1958	1586	1969	245	1980	425	1991	1982
1959	218	1970	140	1981	1982	1992	277
1960	623	1971	49	1982	277		

Let say, to drive home that point, we will take annual maximum discharge at a particular river for 45 years, so from 1950 to 1994, 45 years are there, and the Q that we indicating here is the maximum discharge; that means, out of all the values that we have collected over here, we have picked up one value per year. That is our maximum discharge and the units are in cumec, which is cubic meter per second. So, you have 45 such values. Let say that we want to put a threshold of 1500; that means, we say that an extreme event has occurred whenever it has crossed a 1500 cubic meters per second.

(Refer Slide Time: 29:28)



So, we draw the time series first on the x axis you show the year and y axis is the discharge Q. We draw the time series and then draw a line 1500 cusec, I am sorry 1500 cumec; cubic meters per second. Any flow that exceeds this Q, 1500 cumec, we count it as a success, in as far as our definition of extremes values concerned; that means, whenever that event occurs we count it as a success, whenever it has not occur we count it as a failure.

(Refer Slide Time: 30:07)

Frequency Analysis

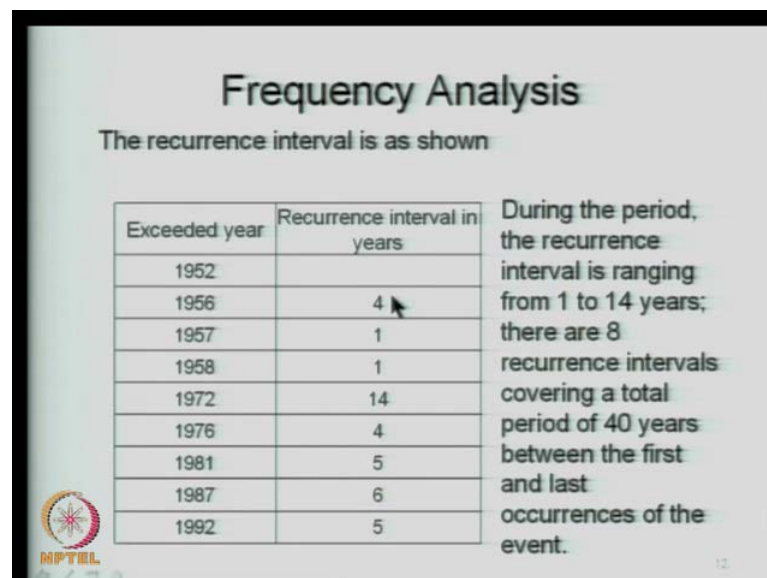
If $x_T = 1500\text{cumec}$, during the period, annual maximum discharge Q, exceeded x_T nine times

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	
1957	1642	1968	328	1979	162	1990	
1958	1586	1969	245	1980	425	1991	
1959	218	1970	140	1981	1982		
1960	623	1971	49	1982	277		

So, there are nine such events as we can see here. From the data, whenever the value 1500 has exceeded, we call it as a success. So, 1, 2, 3, 4, 5, 6, 7, 8, 9; 9 times in the 45 years record, the value of 1500 has been exceeded. What do we do with this information?

We will now calculate the recurrence interval; that means, in 1952 it occurred once. When did it occur next? It occurred next in 1956. So, there is a lapse of four years. Then when did it occur next. It occurred in 1957. So, there is a lapse of one year, again a lapse of one year. So, this event of X being greater than or equal to 1500 had a recurrence interval of 4 years here, had a recurrence interval of one year here, recurrence interval of one year here, then it again occurred in 61. So, 3 years and so on. So, this is what we count here.

(Refer Slide Time: 31:21)



The recurrence interval is as shown

Exceeded year	Recurrence interval in years
1952	
1956	4
1957	1
1958	1
1972	14
1976	4
1981	5
1987	6
1992	5

During the period, the recurrence interval is ranging from 1 to 14 years; there are 8 recurrence intervals covering a total period of 40 years between the first and last occurrences of the event.

MPTEL

So, there is a recurrence interval of four years between years 1952 to 1956. Then one year, one year, then 58 to 72 **sorry** it was 1958 after that it occurred in 72, not in 61. It occurred in 72. So, there is a 14 years gap there. Then it again occurred after four years, in 1976. So, there is a four year gap here. So, we calculate, we count the recurrence interval in years like this and then list it, which means what, in the record that we had seen the recurrence interval is varying between 1 to 14 years.

The event returned sometimes in one year, sometimes as frequently as in one year, as soon as in one year and as distantly as 14 years. So, it valid between one to 14 years and there are 8 such intervals, because we talked about 9 events between them is 8 such

intervals and there is a 40 year period in this recurrence interval. So, what is an average recurrence interval? Average recurrence interval is in this 40 years, eight times, it has eight intervals are there. Therefore, the average recurrence interval is 40 divided by 8 which are five years. This is an average recurrence interval.

This is eight recurrence intervals covering a period of 40 years between the first and last occurrence. So, that is the area of obtaining the recurrence interval. We leave out the first one. So, after the first one has occurred, how many times it has recurred? And what is the average recurrence interval. That is what we are talking about. And therefore, the average recurrence interval comes out to be five years.

(Refer Slide Time: 33:29)

Frequency Analysis

The return period for 1500 cumec annual maximum discharge on the river is equal to the average recurrence interval

$$\bar{r} = \frac{40}{8} = 5 \text{ years}$$

The return period of a given magnitude is defined as the average recurrence interval between events equaling or exceeding a specified magnitude

NPTEL

The return period of a given magnitude is defined as the average recurrence interval between events equaling or exceeding a specified magnitude. So, we specified a magnitude of 1500 cubic meters per second and then we recon the recurrence intervals of such an event occurring; that means, X being greater than or equal to 1500. How frequently it occurred in the record? Having occurred once, how much time elapse between its occurrences again? That is what a recurrence interval is and then we computed the average recurrence interval.

So, in the record of 45 years the average recurrence interval turns out to be 5 years, and this 40 that we have used is the number of years between the first occurrence and the last

occurrence. So, after the first occurrence has been noticed, we start counting the recurrence interval.

This average recurrence interval is in fact the return period of that particular event. So, for this example the return period of the event that the flow magnitude equals or exceeds the magnitude of 1500 cumec, cubic meter per second, is Q_{max} , is 5 years, which means once in about 5 years, you can expect this particular magnitude to be equal or exceeded. We put this more formally through the probability now.

(Refer Slide Time: 35:16)

The slide is titled "Frequency Analysis". It contains the following text:

The probability $p = P(X \geq x_T)$ of occurrence of the event $X \geq x_T$ in any observation is related with return period as follows:

For an observation, two outcomes are possible: "success" or "failure"

Success is the probability 'p' i.e., $X \geq x_T$; Failure probability $(1 - p)$ i.e., $X < x_T$

The slide also features the NPTEL logo in the bottom left corner and a small inset image of a man in a dark sweater and glasses in the bottom right corner.

So, this is the event we interested in, X being greater than or equal to x_T . The subscript T here denotes the return period. As your return period changes your x_T changes or as your flow magnitude here x_T changes your return period changes. And therefore, we denote this with a subscript of T , indicating that the flow here or the event here is dependent on the return period T .

So, we denote this probability p as probability that X is greater than or equal to x_T . Then what we do is whenever this event has occurred, X being greater than or equal to x_T in the record, whenever it has happened, we count it as a success. If that event has not happened, has not occurred then we count it as failure. So, any event X is less than x_T will be reckoned as a failure. Now, because we define p is equal to probability of X being greater than or equal to x_T , the success occurs with the probability of p , and then the failure occurs with the probability of one minus p . So, the event X is greater than or

equal to $x \leq T$ has a probability of p . The event $X < T$ has the probability of $1 - p$.

(Refer Slide Time: 37:05)

Frequency Analysis

Since the observations are independent, the probability of a recurrence interval of duration τ is the product of probabilities of $\tau - 1$ failures followed by a success i.e., $(1 - p)^{\tau - 1} p$.

1 2 ... $\tau - 1$ τ

$(1 - p)(1 - p)$... p

NPTEL

Because the events are all independent, all the observations are all independent. The probability of the recurrence interval of duration is τ . Now, what we are saying that once the event has occurred here, what is the probability that it occurs again only after τ years, because there is a concept of return period. What is the probability that it returns after τ years? What is that mean? That means, in first year it does not occur, second year it does not occur, third year it does not occur and so on, the τ minus one year it does not occur, but the τ th year, it occurs.

That particular event occurs. And this probability is the probability that the event does not occur is $1 - p$, does not occur here is $1 - p$ etcetera up to τ minus one, and on the τ th year it occurs with a probability of p . Because the events are independent, the probability of such a sequence occurring is $(1 - p)^{\tau - 1} p$. So, that is what the probability here is. So, the recurrence interval of duration τ is the product of probabilities of τ minus one failures, up to this point, followed by a success in the τ th time period. So, that will be $(1 - p)^{\tau - 1} p$.

(Refer Slide Time: 38:57)

The slide is titled "Frequency Analysis". It contains the following text and equations:

The expected value of τ is

$$E(\tau) = \sum_{\tau=1}^{\infty} \tau (1-p)^{\tau-1} p$$

Expanding the equation,

$$E(\tau) = p + 2(1-p)p + 3(1-p)^2 p + 4(1-p)^3 p + \dots$$
$$= p \{ 1 + 2(1-p) + 3(1-p)^2 + 4(1-p)^3 + \dots \}$$

The slide also features the NPTEL logo in the bottom left corner and a photograph of a man in a dark sweater and glasses in the bottom right corner.

So, this is the probability of the event recurring once in tau years. Now, we have defined the return period to be the expected value of the recurrence interval. Now, this is the probability of recurrence interval. The expected value of the recurrence interval is what we are interested in.

So, let us recall how we define the expected value? Remember, we wrote a expected value of random variable, when the variable is discrete, we wrote this as $\sum_{x=1}^{\infty} x p^x$. Same way we are now defining it for the random variable tau, expected value of tau, the value of tau itself is multiplied by the probability. So, this is the probability of occurrence, this is the probability. So, τ into p of x that gives the expected value. So, this is how we are getting the expected value of tau.

If we expand the equation; that means, tau is equal to 1 we put. Then 1 into 1 minus p to the power 0 into p, that is p. Then tau is equal to 2, we put 2 into 1 minus p to the power 2 minus 1, which is 1 into p, plus 3 into 1 minus p whole square into p and so on. So, this is how we expand the expected value of tau and this we write is as, p you take it common, 1 plus 2 into 1 minus p into 1 minus p whole square and so on. So, this is the expected value of tau. Now, the terms within the bracket here, namely 1 plus 2 into 1 minus p plus 3 into 1 minus p whole square plus 4 into etcetera, now, this can be equated to a series.


(Refer Slide Time: 41:03)

Frequency Analysis

The expression within the brackets has form of power series expansion:

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \dots$$

With $x = -(1-p)$ and $n = -2$.
Then equation (a) becomes:

$$E(\tau) = p(1 - \{1-p\})^{-2}$$




There is an expression within the bracket. We can write this as, this particular expression as one plus x to the power n is equal to and this is expanded as one plus n x plus n into n minus one by two factorial x square plus etcetera, where we are putting x is equal to minus one minus p and n is equal to minus two. So, in this if you put x is equal to minus one minus p and n is equal to minus two, you can verify that you will get the same expression as it is in the bracket here. With this now, what we are writing for example, one plus x to the power n the terms in the bracket. So, one minus, X is minus one by minus p, to the power n is minus 2.

(Refer Slide Time: 42:01)

Frequency Analysis

$$E(\tau) = \frac{p}{(1 - \{1-p\})^2}$$
$$= \frac{p}{(p)^2}$$
$$= \frac{1}{p}$$

$E(\tau)$ is the return period 'T'

$$T = \frac{1}{p}$$


Also, we write is as expected value of tau is equal to p that is this p divided by 1 minus of 1 minus p to the power of two. That will be p by p square, this becomes p square and therefore, is equal to 1 by p. Therefore, expected value of tau, we are showing it as one by p. Recall that p is a probability of X being greater than equal to x T or a probability of success. So, because of expected value of recurrence interval tau is in fact, the return period, we write it as T is equal to one by p.

Now this is an important and very useful result. That is the return period is equal to one divided by probability of X being greater than equal to x T.

(Refer Slide Time: 43:01)

The slide is titled "Frequency Analysis". It contains the following text and equations:

or $p = \frac{1}{T}$

i.e., the probability of occurrence of an event in any observation is the inverse of its return period

$$P(X \geq x_T) = \frac{1}{T}$$

Therefore the T year return period event is $X \geq x_T$ and it occurs on an average once in T years.

In the bottom left corner, there is a logo for NPTEL (National Programme on Technology Enhanced Learning) featuring a stylized sun or starburst design.

In the bottom right corner, there is a small inset image of a man in a dark sweater and glasses, likely the presenter, standing in front of a screen.

And mostly we write it on the other way round; we write p is equal to one by T. That is the probability of occurrence of the event is equal to one divided by T. So, specify the return period, immediately you will get the probability, and that probability we use for all over design and so on. So, we are saying here the probability of occurrence of an event in any observation is the inverse of its return period, that is probability of X being greater than or equal to x T is equal to one by T. Therefore, the T year return period event is X is greater than equal to x T and it occurs on as average once in T years.

(Refer Slide Time: 43:56)

Frequency Analysis

Example – 2: What is the probability that the annual maximum discharge in the river will exceed or equal 1500 cumec for the data in Example-1

Sol: Return period of the event $[X \geq 1500]$ is 5 years

$$\begin{aligned} P(X \geq 1500) &= \frac{1}{T} \\ &= \frac{1}{5} \\ &= 0.2 \end{aligned}$$

MPTEL

Say for example, we look at this a same example one that we considered, the same data 45 years data and the same threshold value **1500 years** 1500 cubic meters per second and we know that the return period of event is five years. We just obtained this return period to be five years as I showed here (Refer Slide Time: 44:27). So, this is average return interval or return period. So, this is five years.

The return period is associated with the particular event. So, we are saying X is greater than equal to 1500 cumec has a return period of five years. The event X is greater than equal to 1500 cumecs has a return period of five years. Now, we will use that and answer the question what is the probability that the annual maximum discharge in the river will exceed or equal 1500 cumec for the data in example one.

So, we simply put p is equal p of X greater than equal to 1500, probability of X greater than equal to 1500 is one by T, which is equal to 1 by 5, which is 0.2. So, this is how we obtain the probabilities for a given return period. Let say our return period is forty then; obviously, this 1500 will be different. As the return period increases, your magnitude also will increase. That is a 40 year flood will be greater than a five year flood, a 100 year flood will be greater than a 40 year flood and so on. So, this 1500 we have written as $x T$, which is associated with the return period T. So, from the return period you can obtain probability.

(Refer Slide Time: 45:54)

Frequency Analysis

A useful question to be answered is: what is the probability that a T year return period event will occur at least once in N years?

- Consider the situation that in N years, the event $X \geq x_T$ does not occur at all.
- The probability is $(1 - p) \times (1 - p) \times (1 - p) \dots \dots N \text{ times} = (1 - p)^N$

$(1 - p)$ is the probability that in a year the event $X \geq x_T$ does not occur

NPTEL

Now, related to this, an interesting and useful question that we need to answer is what is the probability that a T year return period event will occur at least once in ten years? There is a certain understanding that is necessary here. Let say we are designing a critical structure or let say a nuclear reactor or some such things and you want to protect this structure against flooding. And this nuclear reactor may last for let say 1000 years or some such things. **and the flooding that you want to** the flood discharge that you want to prevent it from may be a 1500 year flood or 2000 year flood or so on or a PMF, a probable maximum flood.

So, we may fix the return period for the flood, which may be smaller than or may be higher than the time at which it is required, the time for which it is required. So, we may be interested in getting the probability that in the life span of this particular structure that we are planning. What is a probability that a T year return period event will occur? To give you another simpler example let us say that you are building a dam. The dam is meant to last for 100 years, but, the structure of the dam you may have designed for a 1000 year flood event or the forces that correspond to 1000 year flood, which is different from spillway discharge capacity. Spillway discharge capacity may be a 100 year flood capacity, but the structure itself may be designed to be standing the forces that are caused by a thousand year flood, the dynamic forces and so on.

And therefore, you will be interested in answering the question. What is the probability that a 1000 year flood event will occur in the span of 100 year life time of the reservoir? Therefore, this question is relevant. What is a probability of a T year return period will occur at least once in N years? So, let us try to answer this question.

We will take the complementary event; that means, at least once in N years, the complementary event to that is that it has never, it never happens in the N year period. So, we will consider the complementary event that in N years the event X greater than equal to x T does not occur at all. If it occur even once, we are saying that it has occurs at least once, and therefore, it should not occur at all for the event to be complemented.

So, we first consider the situation where the event X greater than equal to x T does not occur in the next, in the N year period. What is the probability of this? That is in the first year it does not occur. In the second year it does not occur. Third year it does not occur and so on. For all the N years this event does not occur. What is the probability that it does not occur in a given year. It is one minus p, because p is the probability that it occurs in a given year, and therefore, the events that are all independent. Therefore, the probability that X greater than equal to x T does not occur in any of the N year is one minus p into one minus p etcetera N times. Because one minus p is the probability that in a year the X greater than equal to x T does not occur.


(Refer Slide Time: 50:15)

Frequency Analysis

The complimentary event is that the T year event occurs at least once in N year period; and hence the probability is $1 - (1 - p)^N$

Since $p = 1/T$,

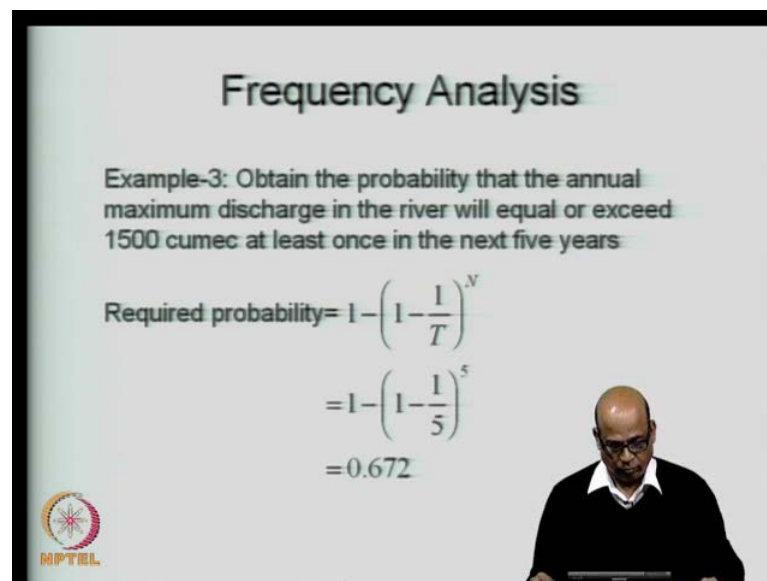
Required probability is $1 - \left(1 - \frac{1}{T}\right)^N$

 NPTEL 22

So, this will be equal to 1 minus p to the power N. The event that the situation that it occurs at least once in the N year period is simply one minus this particular probability. That is 1 minus the probability that it does not occur in any of the N years. So, that will be 1 minus 1 minus p to the power N and what is p? So, p is the probability of X being greater than equal to x T, which is simply equal to one by T as we have seen. So, the required probability is one minus one minus p to the power N, which one minus one minus one by T to the power N.

So, this is the probability that a T year return period event occurs at least once in N years and this is a important result, because we will be using this kind of results when we want to assess what is the risk that we are taking in building hydrologic structures, hydrologic designs. That is we may have built it for 1000 year return period, but we would also be interested in getting the probability at that 1000 year return period occurs. In fact, in the next 100 years and that will give you the measure of the risk that you are taking in designing such a structure.

(Refer Slide Time: 51:41)



The slide is titled "Frequency Analysis". It contains the following text and equations:

Example-3: Obtain the probability that the annual maximum discharge in the river will equal or exceed 1500 cumec at least once in the next five years

$$\begin{aligned}\text{Required probability} &= 1 - \left(1 - \frac{1}{T}\right)^N \\ &= 1 - \left(1 - \frac{1}{5}\right)^5 \\ &= 0.672\end{aligned}$$

The slide also features the NPTEL logo in the bottom left corner and a small image of a person in the bottom right corner.

So, we will use a simple example to drive home this point. let say we have got a five year return period that is the return period corresponding to a discharge of 1500 cumec is five years. We will answer the question what is the probability that this five year return period occurs at least once in the next five years? Intuitively, you may feel that this probability is one, because we are saying that this is a return period of five years and

therefore, in the next five years it will occur once. But, as you can see here, the probability is one minus one minus one over T to the power N. Therefore, this turns out to be this T is five years, return period is five years and this is N we are interested in answering the question at least once in the next five years therefore, this turns out to be 0.672.

So, the fact that the return period is five years does not mean that it will occur at least once in the next five years. The probability that it will occur at least once in next five years will is 0.672. So, to summarize them we have introduced the concept of frequency analysis and we are talking about at station frequency analysis in the sense that we are not concerned about the spatial or the regional frequency analysis here. We are talking about a frequency analysis with respect to a given data, given data time series, the data that we use for frequency analysis should be independent and identically distributed.

The requirement that they are identically distributed may be ensured by using the data from the same population, because we are talking about time homogeneity here. By same population, I mean the watershed should not be disturbed during the period of record nor should the gauge either the rain gauge or the stream gauge that you are interested in be shifted or changed or the major changes have taken place in the way of recording of the data.

The requirement of independence may be ensured by looking at annual maximum series or annual minimum series, where every year you pick up one value out of the large number of values that have been recorded; be it the maximum or the minimum. Then from year to year you can assume them to be independent. Once we have the correct data we use do the frequency analysis, we introduce the concept of average recurrence interval.

We call an event X greater than or equal to x T, as an extreme event and then we look at the recurrence or the time that elapses between the occurrences between consecutive occurrence of such event and recon it as a recurrence interval. We compute the average recurrence interval and then call it as the return period. The return period indicates that on an average such an event will occur once in about T years, where T is the recurrence interval, T is the return period. Then we also examine the relationship between the probability and the return period and we showed that the probability of X being greater

than equal to x_T is in fact one by T , where T is the return period. Then we answer the question of what is a probability that in a given N year record the T year event will occur at least once and that is the solution that we have obtained here.

So, in the next class, we will continue this discussion further and use this frequency factor to carry out the recurrence in turn to carry out the frequency analysis. We will also see how the frequency analysis is different or uses different value for different parent distributions. As I mentioned in the beginning of the class, the frequency analysis is dependent on how we define their extreme values and this extreme values are derived from a parent distributions. We are trying to look at extreme value distribution, which are derived from a given parent distribution. So, thank you for your attention. We will continue the discussion in the next class.