Advanced Hydrology
Prof. Dr. Ashu Jain
Department of Civil Engineering
Indian Institute of Technology, Kanpur

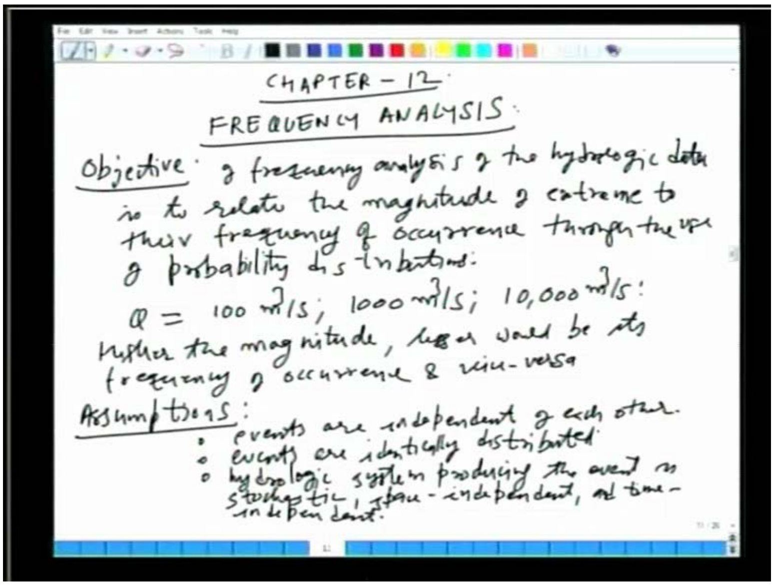
Lecture – 33

Good morning and welcome to the post graduate video course on advanced hydrology. We are into stochastic hydrology module and we looked at various simple concepts of probability etcetera and how to fit probability distributions, that is what we were doing in the last class. We also looked at what is called the testing the goodness of fit, once we have fitted a particular distribution how do we know, how do we find out or how do we quantify weather the fitted distribution is good or bad. We looked at a test called chi square test. In the last class we looked at the procedure, step by step procedure of how we can carry out this chi square test.

Then we looked at one numerical example of carrying out this or implementing the chi square test. We looked at the example of the exponential distribution in which the random variable was the inter arrival times for which ten different data were given. We found the average the parameter lambda, which is 1 over x bar and then we carried the various steps which are required to determine whether the fit is good or not. And then in the final analysis we had found out that since the computed chi square statistic was less than the standard one that is why the fit was good at 5 percent significance level.

So, similarly we can you know test this goodness of fit of more than 1 probability distribution fit to the same data and then we should be able to say that this particular distribution is better than the other one and so on. So, with that actually we complete chapter 11 from the Ven Te Chow's book and what we would do today is we will start chapter 12 on frequency analysis.

(Refer Slide Time: 02:09)



So, here we go we have chapter 12 within this stochastic hydrology it is on. So, we are looking into this chapter 12, which is frequency analysis. So, what is frequency analysis and what do we do in this is basically the objective of a frequency analysis is to determine a relationship between the magnitude of a particular event or a variable that is random variable and its frequency of occurrence. So this is the main objective in which we want to be able to say that this magnitude will occur with a frequency of so and so right.

So, using the data historical data about that particular hydrologic random variable we would be able to establish a relationship between the magnitude and the frequency of occurrence. And there are many methods available for that and that is what we are going to study in this particular chapter. So, if I want to define it here, what we will do is if objective of frequency analysis of the hydrologic detail is as I said to relate or to establish a relationship between the magnitude of extreme events.

Normally, we are more interested in floods and draughts and other extreme events to their frequency of occurrence how well through the use of probability distributions. So, what we are saying here is that we will establish the relationship or we will relate the magnitude of an extreme event such as a flood with its frequency of occurrence all right. How we will do it we will use certain probability distributions, all right? A particular type of hydrologic variable will follow a particular probability distribution.

So, we will fit that distribution and using that distribution we will establish that relationship. So, we will look at some of those things, now looking at some basic concepts about this frequency of occurrence. For example if I say that your Q is at a particular location in a river is 100 meter cube per second or 1,000 metre cube per second or 10,000 meter cube per second, which of these events do you think will be occurring with a higher frequency and which 1 will be occurring with a less frequency.

By intuition you can say that the lower is the magnitude so, out of these let us say 100 meter cube per second will occur more frequently, will occur with more than once in a year all right or it will occur a lot more frequently than the other event, which is Q is equal to 10,000 meter cube per second. So, higher the magnitude of any event lesser will be its frequency or it will be more rare or it will occur very rarely, all right? So, from intuition we know that. But we need to quantify the relationship and will do that using certain probability distributions.

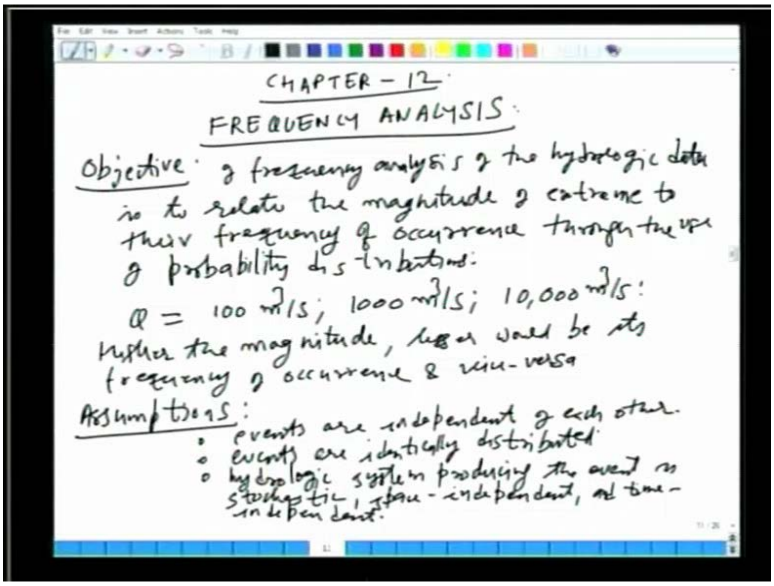
So, before I go to that let me say that higher the magnitude lesser would be its frequency of occurrence and vice versa. So, the opposite is also true if the magnitude is less, it will be occurring more frequently. Now, moving on before we actually look at the various methods that are available for carrying out this frequency analysis. Let us look at some of the assumptions which are made while carrying out this analysis, what is assumed is that all the events, which we are using or the data are independent of each other.

What do we mean by that when we say that the events are independent of each other let us say we are looking at the annual peak flow series. We have 50 years of data so, what was the value of the annual peak discharge let us say 1970 has no bearing on what will be the annual peak discharge in 1990 or what was the annual peak discharge in 2001 has no relationship with that which is going to occur in 2002.

So, that is what we mean when we say the events are independent of each other there should be any correlation. So, that is the basic assumption the next one we say that events are identically distributed. Again what do we mean by this when we say that the events are identically distributed, means let us say we have 50 years of data or 50 data points all of those data or all of those observations follow the sane distribution. The characteristics in the catchment or the the manner or the pattern in which this random events are generated are governed by the same probability distribution.

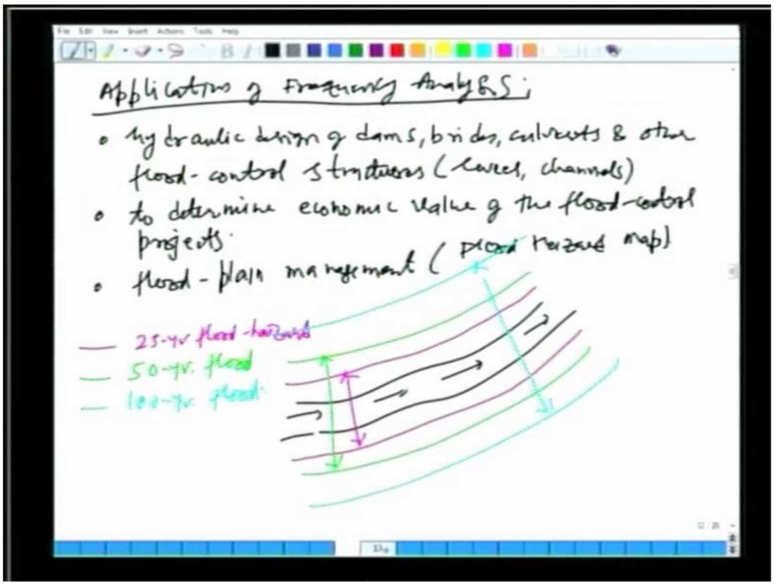
So, there should not be any major changes in the characteristics of the catchment such that we can assume that the data come from the same population and that population follows a particular probability distribution. So, these are the assumptions basically so that we can apply, you know the the principles which we have. The next one is the hydrologic system which is producing the event is what is stochastic, so it is not deterministic in nature all right so we are saying that it is uncertain or stochastic is space independent space independent and time independent this is a, this one is a crucial assumption actually because we know that this assumption you know will be violated more often than not.

(Refer Slide Time: 02:09)



Because the hydrologic system we know there is a deterministic this conceptual modules also available, but because of the fact that the complexity is such that we are not able to understand that so we can assume that it is stochastic. Also we are saying that these events are specially uncorrelated and also uncorrelated with respect to time which again we know that it is not the case all right, but we carry out the frequency analysis under these assumptions, okay?

(Refer Slide Time: 11:15)



So, let us move on the next thing we are going to look at is why we need this or what is the application of your frequency analysis any idea why this needed. Well I will just try to list some of them first and four most is the hydraulic design of your major hydraulic structures such as dam's bridges culverts and other flood control structures. Flood control structures such as levees channels etcetera. So, what do we mean by this when we say that the frequency analysis is useful in the hydraulic design. You know that, how do we design a hydraulic structures? For example, a dam or a bridge let us say you want to build a dam or you want to build a bridge across a river.

So, the the bridge actually is normally designed for some magnitude of the peak discharge which we want to it to pass through. Safely the magnitude of that peak discharge for which we design that bridge all right which is going to be correlated with what with its frequency of occurrence, all right let us say we are going to design it for a recurrence interval of of 100 year all right or a return period of 100 year, all right? So, its frequency of occurrence is once in 100 years, but the data that are available to us is let us say only 50 or 55 years, okay?

So, what we need is, we need to carry out some kind of extrapolation. So, we establish a relationship between the magnitude and the return period and then we can extrapolate find the magnitude for 100 year for which we can design the structure. Number two is to determine the economic value the economic value of the flood control projects. What do we mean by this one, is to determine the economic value of a water resources project or a flood control project. When we carry out the flood frequency analysis what we are able to get is the probability of exceedance of a particular magnitude flood or some particular magnitude event. Using that, we can determine or find out what is called the risk associated with a particular project. We would like to be able to find out what is the probability that the design flood for which this particular project is design will be equalled or exceeded lifetime of a particular project.

Let us say you are building a bridge, the lifetime of a bridge is let us say 100 years in that 100 years what is the chance, what is the probability that these magnitude for which it is design will be exceeded or will occur. So, that probability will actually determine the reliability of that particular water resources project and that reliability and risk associated can, we can determine using the frequency analysis. So, to determine the economic value and the economic value of that particular water resources project will depend upon the risk associated with it if the risk is less, all right? We may want to go head and build the project or we can rake the water resources decision depending upon that discuss associated.

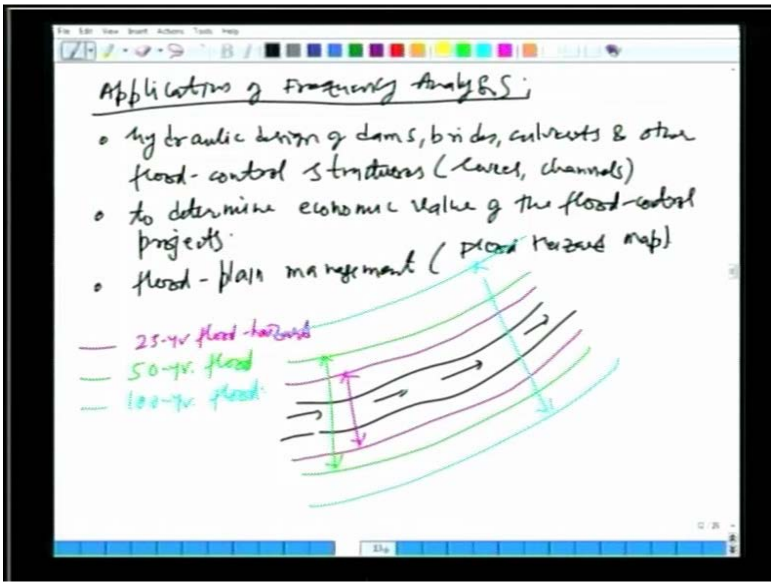
So, this kind of stochastic information or statistical probability in terms of reliability is very useful in determining the economic value of a particular water resources project. Moving on the third advantage or application of this frequency analysis is in the flood plain management. Flood plain management, what do we understand by what is called a flood plain management? We we have two types of, you know flood control management kind of things one is the structural measures other is the non structural measures.

So, when we are talking about the flood plain management we have these two types of methods to protect the people and the property from the flood. For the structural measures, again as i said we need to design those things for which the frequency analysis is useful. For the non structural measures, we what we do is we study the flood plain or the extent to which the waters will spread out during a flooding event corresponding to a particular magnitude, to demonstrate this let me show you that, let us say this is a river which passes through a city and towns.

So, this is a river flowing in this direction, you have the cities and towns on either side of it, now you want to delineate the magnitude of the flood hazard or you want to develop what is called the flood hazard along this river. So, that the people who are living in the in the flood plains will be able to buy the insurance or will be able to take a decision depending upon the magnitude of the flood. So, let us say I am going to delineate different boundaries in hydrology, but in all the different fields other branches of engineering and even sciences and so on wherever the the problem is such that this particular distribution can be applicable or can be useful.

So, what we will do is we will look at the binomial distribution first. Its corollary is this in the pink colour. So, you should not have any construction or any activity in this area. Similarly, we can carry out further analysis and we can find out the magnitude of an event which is higher than this right let us say the green represents a 50 year event 15 years flood. 50 years flood obviously will have higher magnitude than the 25 years flood. So, if you delineate that it will be bigger than this.

(Refer Slide Time: 11:15)

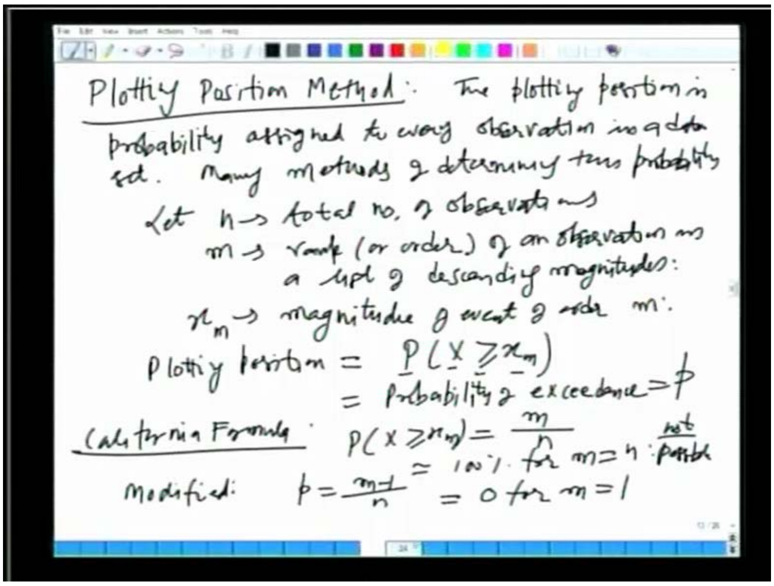


So, let us say this is your 50 years flood which will be bigger than this on either side I am drawing it very nice and easy, but you know that depending upon the topography and the and the slope and everything along the channel you will have, you know different types of boundaries which of course, can be delineated. Similarly, if you want to use or draw what is called the let us say 100 year event, it may be higher than this. So, this correspond the the blue lines correspond to what, the 100 year event or 100 year flood.

So, you see that we can manage the flood by minimising the damages, which can be caused due to the different magnitude floods. How can we do that, well we need to know the magnitudes of those floods depending upon the frequency of occurrence. In the United States, they have a national flood insurance program according to which what they have done is they have mapped all the major rivers in the whole country. So, you have this flood maps available for any ordinary citizen or for the businesses and so on so you can get these maps for free from the government.

Then buy out your own insurance in the US the responsibility of the damages due to flood is not on the government. It it is your own responsibility of the citizens, but in our country in India the responsibility is a, lies with the government. So, there is a, you know a major difference in how we actually manage the floods so, flood management is a very important topic in which we need the help of this frequency analysis.

(Refer Slide Time: 20:59)



So, let us move on what we are going to do is we will look at, before we move to the frequency analysis method. We will look at a very simple method of relating the magnitude and the probability this method is what is called the plotting position method. I am sure you may have seen this in your earlier classes, may be it is called something else the plotting position. What is it? It is nothing, but the probability assigned to every observation in a data set or in data. How it is done, well there are many methods available of determining this probability so, what we are saying is that we we will carry out very simple method, which is called the plotting position method in which what we do is we calculate or estimate what is called the plotting position. What is a plotting position? It is nothing but the probability which is assigned to what to each observation. For example, let us say you have 50 years of annual peak flood data.

So, you have 50 values of the peak discharges, what we will do in the plotting position method? We will assign a probability of occurrence or the frequency of occurrence or the probability of accidence. Actually to each observation or each magnitude of the annual peak flood all right how do we do that? One is a very simple method, I am sure you have seen it in your earlier classes, we arrange the data in the descending order the highest magnitude flood or highest magnitude event we put on the top. We assign a order of 1 to that then the next highest is the second one, the order will be 2 and so on. So, we arrange all the observations in the descending order of magnitude and then we use a particular equation to calculate the probability of accidence or the plotting position. So, let us look at that.

So, if we have N as the total number of observations with us in the data set and M is what M is the rank or order of what, of the particular observation or data point in what in a list of descending magnitudes. So, what we are doing is we are arranging all the observations in the descending order. We assign an order of 1 to the top value 2, 3 and so on and that this column let us say we call M. So, m can assume any value from 1 to N. N is the total number of data points and then X subscript M, we say is the magnitude magnitude of event of order M.

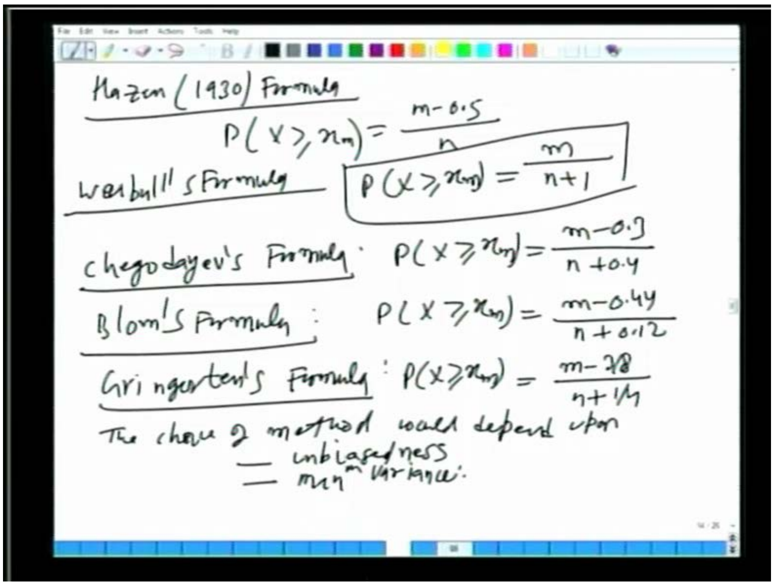
Now, what we are going to do is let us say we have the data and different data points observations. We arrange them so we determine what is M all right M is the order which will vary from 1 to N. So, there are many formulas available to calculate what is called the probability of accidence or what is the plotting position. So, what I am going to do is I will write down all of these, may be discuss very briefly their merits and demerits and one of them which is the viable formula I am sure all of you may have seen in your earlier classes.

So, let us start looking at what is called the plotting position formula which basically gives you what the probability of accidence what is this denoting this is the probability of random variable X having a magnitude greater than or equal to what, X M. X M is the magnitude of that rank M and this is called the probability of accidence and let us say we denoted as small P. So, let us start looking at many if these formula California formula is the first one in which we say that the p is or the probability of accidence is given by simply m over n. So, when when value of your m is one the probability is one by N. Let us say N is 100 for M is equal to 2 it is 2 by 100, 3 by 100 and so on.

So, what will be the probability of accidence for m is equal to N well it will be 100 percent for m is equal to N, but whenever we talk about the probabilities, can we be 100 percent sure. Can you say that the flood magnitude next year will be equalled or acceded will be this much with 100 percent accuracy or with 100 percent confidence? No, right? So, we need to modify this formula, so what people have done is they have modified it. Being able to say every, you know something about probability with 100 percent confidence is not possible.

So, what they say is that then P, later on this is modified as M minus one over N. However this formula also has you know slight problem why because the value of will be the probability of accidence will be 0 for what for your M is equal to 1. Well then what do we do? We go to some other formula so different people have proposed a different formulas.

(Refer Slide Time: 28:10)

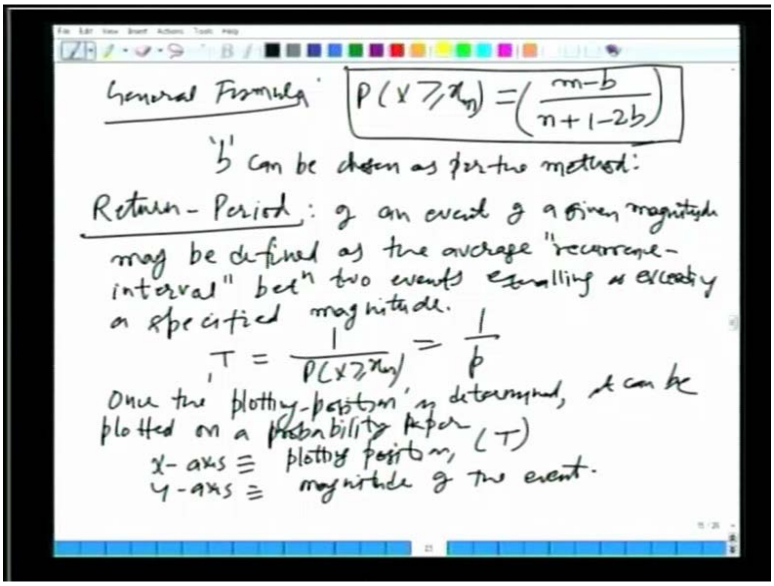


This is proposed by Hazen in 1930 according to which it is given by this, as M minus point 5 divided by M. Another formula is the Weibull's formula, which is the most commonly used. This is e and this is Weibull's formula in which the probability of accidence is given by what let us write it is M over N plus 1. I am sure you have seen this in your earlier classes then another 1 is Chegodaev formula, it is a Russian scientist according to which the probability of accidence is given by this m minus 0 point 3 over N plus 0.4.

So, and there is 1 what is called the Blam's formula which is given by M minus 0.4 over N plus point 1 2 and then there is Gringorten's formula. Gringorten's formula, which says that this probability of accidence is given as M minus 3 X over N plus 1 fourth, so you see that there is a. So, many formulas available to determine this plotting position all right now the question that arises is which one am I going to use. So, this depends actually upon the characteristic or the pattern that are involved in aparticular data set or in a catchment. So, different people tend to use a different formula, but the criteria is that any formula which we use should be unbiased all right and it should given estimate which reduces the variants in our data.

So, the choice of method would depend upon, upon 2 things first 1 is the un biasness means what you see that some of these equations will be bias towards either the higher magnitude flow or the lower magnitude flow or, or or and so on. So, you need to be able to use something which is bias towards any magnitude flow and the other one is the minimum variants which minimises the variants from the fitted line which we are going to have.

(Refer Slide Time: 31:49)



So, all of these formulas then what we can do is we can write a general equation general formula, which is what your probability of x greater than equal to X M, then can be described as m minus B right divided by N plus 1 minus 2 B. So, this is a general plotting position equation. And then B can be chosen as per your method or 1 can use actually 2, 3 different methods and then determine which 1 gives you the best results.

So, this is a, what is called a very preliminary or a very approximate method of carrying out the frequency analysis now what we will do is, we will define or move into what is called the very general method. But before we do that let us look at what is a return period because the frequency of occurrence or the probability accidence which we have just seen can be represented by another term which is called the return period. I am sure all of you have seen this term before and know what it is, so will just define it quickly.

What is a return period? It is a return period of an event of a given magnitude, may be defined as may be defined as the average recurrence interval between 2 events equalling or exceeding a specified magnitude. So, what is a return period basically is that it is again the inter arrival time on an average over a very large period between 2 events.

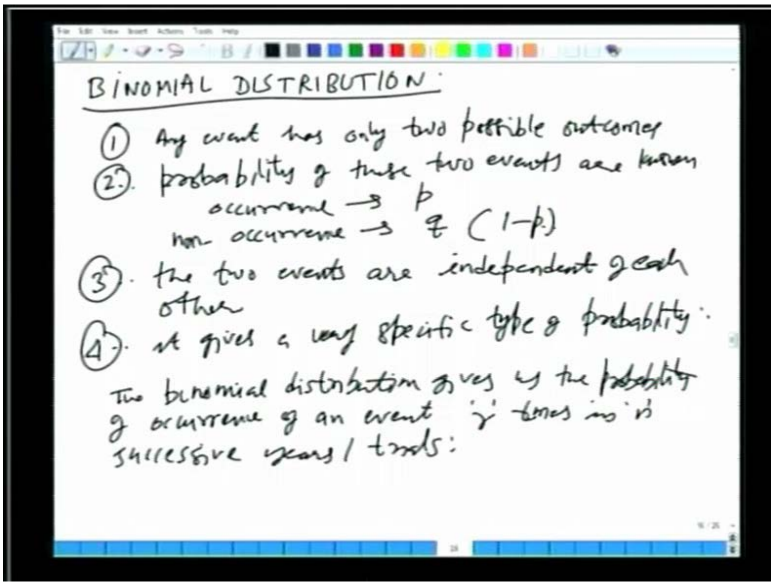
So, on an average what is the time interval elapse or the recurrence interval between 2 events which will, which are of a magnitude, which will be equalled or exceeded once in every likes T years and thus T is what is called the return period. Now this t is given as the reciprocal of the probability of accidence X greater than equal to X M or this is one over P. So, you see that we just looked at what is plotting position method in which we determine the P or the probability of accidence

So, if we have the data quickly what we can do is we can just arrange them in the descending order we can find out the probability of accidence we take the reciprocal of that and then that will give us what the return period. So, quickly in an excel sheet you can do that actually so you have the magnitude of the flood or magnitude of an event it may be rain fall it may be wind velocity any hydrologic variable. So, you have the magnitude and associated with each magnitude you have what the return period.

So, you have established a relationship between the magnitude and the frequency of occurrence. So, you can plot that and then that will give the equation and then you can fit line line or an equation to those data points. So, that you will be able to get what is called the frequency analysis or the fitted equation so, that you can do interpolation extrapolation or do some kind of stochastic analysis. So, moving on what we do is once the probability or once the plotting position is determined it can be plotted it can be plotted on a probability paper were your X axis represents what the plotting position or the probability or the return period it can be in terms of the return period and Y axis is what Y axis actually represents your magnitude of the event.

Now, we will move on and look at something very interesting and something very important we have seen the plotting position and actually before we move on to the other methods of a frequency analysis. We will look at a discreet probability distribution called the binomial distribution right and this binomial distribution is very important not only in hydrology, but in all the different fields other branches of engineering and even sciences and so, on wherever the the problem is such that this particular distribution can be applicable or can be useful. So, what will do is we will look at the binomial distribution first, its corollary is you know some, you know results which can be deduced and then I would like to look at an example based upon the binomial distribution.

(Refer Slide Time: 38:24)



So, let us first look at binomial distribution as I said that binomial distribution is a very special type of distribution which is applicable under very unique or different kind of a situation. So, let me define those things when it is applicable and what are the criteria under which we can use this. First one if them is that any event has only two possible outcomes, what do we mean by that? All we can apply the binomial distribution to a situation in which you define an event or you have an event of which there can be only two possibilities of outcome and the classic example that is tossing the a coin when you toss a coin you will either get a head or tail there is no other possibility.

Similarly, how can we apply this this kind of concept in hydrology, well you you can define that the magnitude a particular magnitude of peak discharge in a river constitutes flood. For example, let us say you have river Ganga in which the in a particular region let us say close to Varanasi or somewhere the flood will occur when the magnitude or of the discharge of the you know river exceeds let us say 1,00,000 meter cube per second.

So, once we have define that you would be able to say that the next year weather it will flood or the area around that you know particular river location weather it will flood or not flood. There will there will be only two possible outcomes, so depending upon the magnitude of the particular event the outcomes will only be 2. So, we can actually define our event in that manner. So, this is one particular characteristic of your binomial distribution when it has to be applied other 1 is that the probability of these two events either this or that are node before hand and this is called what occurrence probability of occurrence is. Let us say defined as P and the non occurrence is defined as what let us say Q and Q will be equal to what based on our law of total probability it will be one minus P.

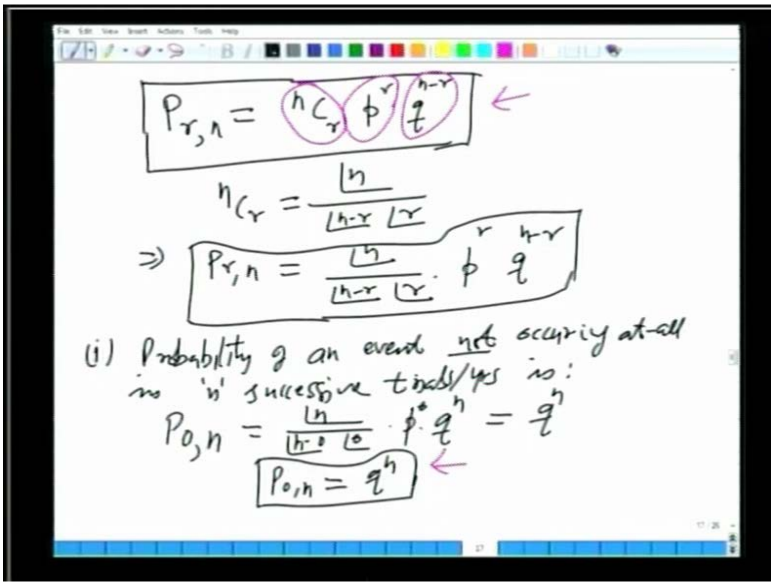
So, you see that when you're tossing a coin you will have a head. The probability of head is if it is point 5 what is probability of tale it is 1 minus 0.5 which is also 0.5. Or in case of hydrology if the probability of occurrence of flood is let us say 20 percent, then non occurrence will be 80 percent. So, they are complementary to each other third 1 is that the two events are, what are independent of each other? Well again I think you know this concept when we say these two in two events are independent of each other. When do we say that? When the occurrence of one does not depend upon the occurrence of the other one, when you toss a coin, let us say 10 different times, all right? So, does the probability of occurrence of head ninth time depends upon what happened during the second time or third time or fifth time? No. So, all these events are independent of each other, similarly in case of hydrology we assume that the occurrence of flood during any given year is independent of what happened in the past. If a flood occurred in 1971, it has no bearing on what is going to happen in 2000 or the next year and so on, all though there may be little correlation, but we ignore that.

So, this, these are some of the conditions and then the last thing I would like to say about binomial distribution is that it gives a very specific type of probability to us what does it give. We are saying that it gives us a very specific type of probability of occurrence and what is that we are going to look at in the definitions of this binomial distribution, all right? So, let me say then the binomial distribution gives us, gives us what? The probability of occurrence the probability of occurrence of what kind of event? Of an event, now this is interesting and important to understand R times in n successive years or trials is this clear also what does the binomial distribution give us, it gives us the probability of occurrence of an event R times in n successive years or N successive trials. For example, let us say you flip a coin 20 times.

So, N is equal to 20, so you are carrying out 20 trials out of those 20 trials you want to find out what is the probability that the head will occur exactly 5 times that is R is equal to 5. So, you are trying to find out a probability of getting a head 5 times out of 20 trials, which are carried out independently and successively 1 after the other there should not be any gap. Similarly, we can how is it useful in hydrology well we may be interested in finding out in hydrology that what is the probability that a flood will occur 2 times in the lifespan of a dam the lifespan of a dam may be let us say 75 years.

So, we may be interested in finding out what is the probability that the flood will be equalled exceeded the design flood for which the hydraulic structure been design will be equalled or exceeded exactly 2 times or exactly 5 times in the lifespan and that probability will be very useful in finding out and design, in the design of the structure and in determining what is the risk and reliability associated with that particular structure. So, this is the kind of probability gives and there is an expression or formula which we will write directly I would not like to go in to the mathematical proof or the details of that because this is an applied course.

(Refer Slide Time: 46:50)



So, this probability is given by this, this is denoted as P r comma n that is how I am going to write, different books will follow different notations, but you understand what is P r comma Nn It is basically giving you what? It is giving you the probability of occurrence of occurrence of an event what R times in R successive trials or in case of hydrology. We can say R times in N successive years. So, this is equal to n C r P to the power R and Q to the power N minus R. So this is the final actually result I do not want to go into its derivation.

So, the probability of occurrence of an event R times in an N successive years is given by this formula were what is N C R. It is a combination from your knowledge of a permutations and combinations your N C R L will give you is directly given as the N factorial divided by N minus R factorial divided by R factorial. So, that your P R comma N is going to be equal to this thing. Times P to the power R N Q to the power N minus R were R is the number of times exactly the event will occur out of N different trials again. I will take a couple corollaries of this or results which can be deduced out of this the probability of an event not occurring at all, not occurring at all can you find it, in N successive trials or years can you find it out from the binomial distribution not occurring at all means R is equal to what R is equal to 0.

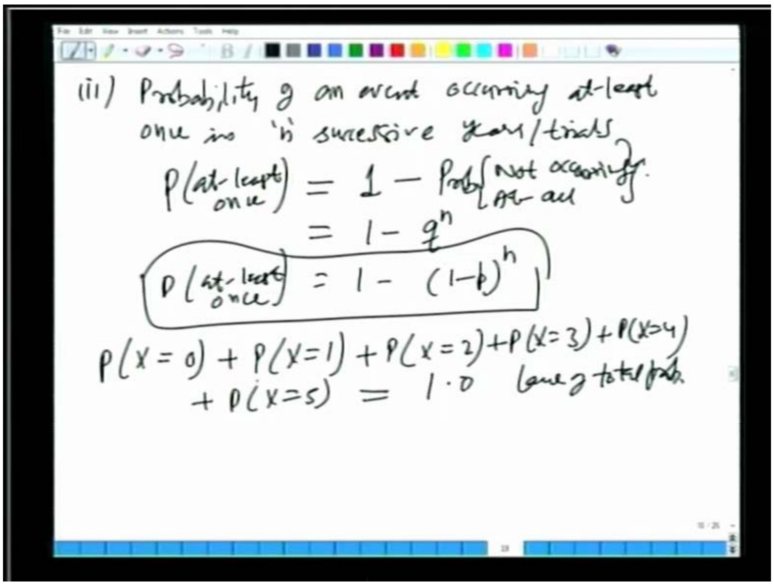
So, all you do see you put R is equal to 0 in this that is occurring 0 times and not occurring N minus R or N times all right So, this you can say is what P 0 comma N were R is equal to 0 not occurring at all means occurring 0 times. Right so, you put R is equal to 0 you will have the result so if you did that you will have N factorial divided by N minus 0 factorial times R is 0. So, 0 factorial times what P to the power 0 times Q to the power N right N minus R. R is 0 N factorial N factorial will cancel out 0 factorial is one anything raise to be power 0 is 1.

So, what would this be this is nothing but, Q to the power N so P 0 comma N is what is Q to the power N what is this saying this equation is saying that if you have N trials all right in each of these. N trials the probability is let us say Q what is the joint probability joint probability is Q times Q times Q times Q and times we are seen that earlier when we did the that method of maximum likelihood. It is a similar thing here so, if you apply this law of a probabilities intuitively you will get the same result. I will actually like to go back to this this equation, this one which gives you the probability of occurrence of an event R times in N successive trials, right? What is P raise to the power R representing? It is actually representing the probability of occurrence of the event r times exactly r times.

So, P times P times P times P up to R times, that is what it is the joint probability of occurrence what is this Q raise to the power N minus R is what if you have 20 trials and out of those 20 trials you have 5 times. Let us say R is equal to 5 you have this event is occurring 5 times means, what it is not occurring N minus R times or 20 minus 5 which is 15 times. So, the total probability then has to be P to the power R multiplied by what, its non occurrence which is Q to the power N minus R.

So, this is the log of actually your joint probability of occurrence we are trying to apply. Now, what is this N C R doing here? Why do we have this? If you think about it this combination can occur in many different ways. We are saying exactly 5 times, but those 5 times can be first 5 years, last 5 years somewhere in between or there can be many such combinations. How many such combinations would be there, well N C R...

(Refer Slide Time: 52:57)



So, we can derive actually this binomial distribution using our basic law of or basic probability, so this was the first corollary in which we found out or we said that the probability of non occurrence of an event in N trials is Q to the power N. What is the second one? The second result is this were in we say that the probability of an event the probability of an event occurring at least once in N successive years or trials is given by this probability of at least once and this is also very important will be given by what it will be 1 minus of probability of what event of not occurring at all which we have just seen.

So, if you have if you are carrying out N trials then what you are saying is that the probability that a particular event will occur at least once. At least once will be equal to what 1 minus it will not occur at all, which is equal to what one minus Q to the power N or 1 minus 1 minus p to the power N. So, this is your at least once this formula actually can be derived or using the law of the total probability. So, let us say you try an event 5 different times so, let us say n value of N is equal to 5.

So what are the possible combinations of what can happen. An event can occur let us say during the first trial and it will not occur in the remaining four similarly, if r is equal to 1. then it can occur in the second one and not in the first one and the remaining one and so on. So, this way you can have that the probability that your X will be equal to either 0 that is it will not occur at all plus…