Water Resources Engineering

Prof. R. Srivastava

Department of Civil Engineering

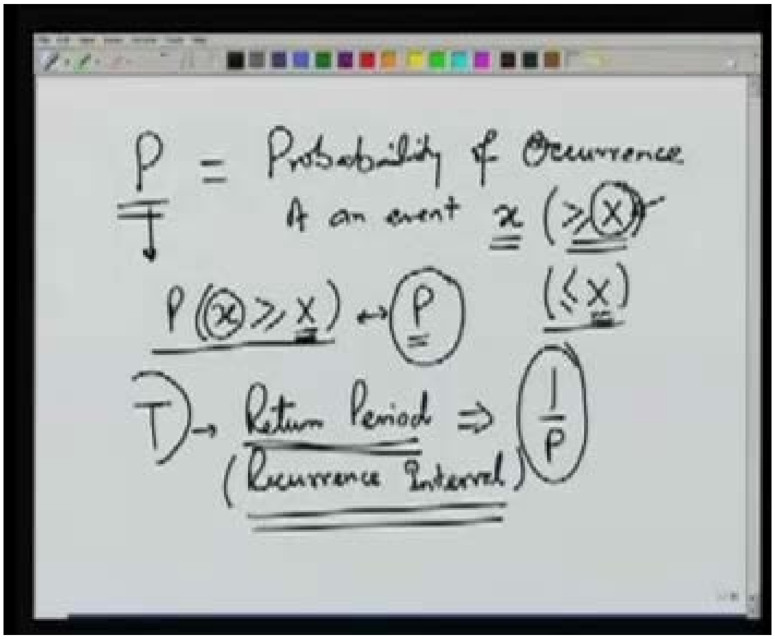
Indian Institute of Technology, Kanpur

Lecture No. # 14

Whenever we are designing a water resources project, we must decide on what should be the design values. For example if we are designing against that control then we should know the maximum flood that can possibly occur. These depend on chance. Example, we do not know when the floods would occur the next year, or next to next year. But based from the previous data, we can estimate some of the values. Similarly if we are making a storage reservoir to protect against drought we should know the minimum flow that can occur in that area. Since all of these are based on probability and chance, we should now look at some of the things which depend on a probability and these are known as risk reliability and safety factor. All these factors partly depend on what is the probability of occurring of an event.

Most of the hydrological events for example, rainfall or floods would be dependent on chance and nobody can predict what will happen next year or next to next year. That is why we have to analyze them in a probabilistic frame. We will know the probability of the flood occurring the next year or the probability of the minimum flow in the river. We will look at some probability analysis in this lecture, for the above reasons. We will look at some of the probability theories and based on that we can assign a risk and choose a design value.

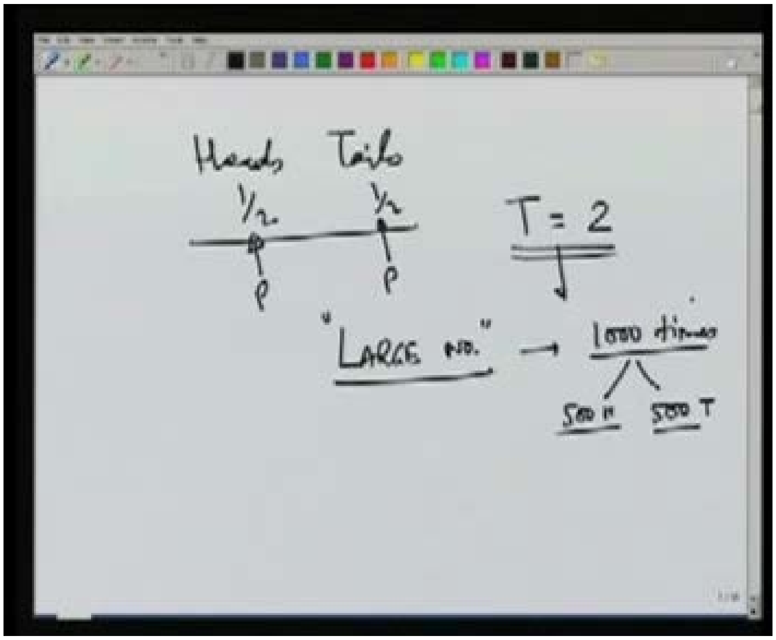
(Refer Slide Time: 01:57)



So let us write the probability as probability of occurrence of an event. Let us say x is an event which may be rainfall, flood in a river, it may be water level in the river and we want probability that this x would be greater than or equal to some specified value x. If you are dealing with floods, we will be using greater than or equal to. If we are concerned with the drought, then we will want a probability that x is less than or equal to a certain given value x, which may be the minimum desired flow in a river. If we use this greater than or equal to x, it means at a probability of accidence, this would be the flood in a river and naturally this probability would depend on what time period we are talking about. For example we will know the probability of it occurring in the next 10 years, next 100 years and so on.

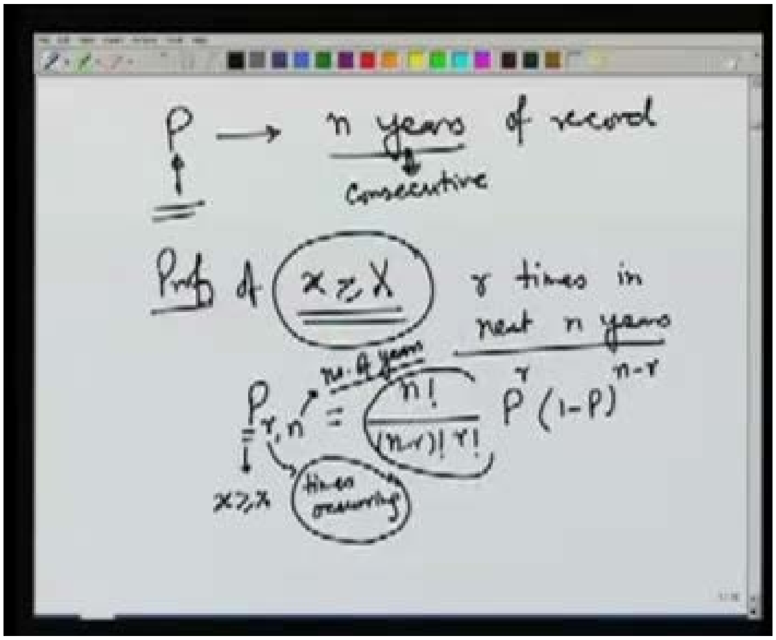
This would be denoted by P and we would write it as probability x greater than or equal to X or in most of our discussions, we will just use the symbol P to denote the probability that the variable x which may be let us say the maximum annual flow in the river is greater than or equal to some specified value X. Now based on the probability, we can define a return period T. It would be simply equal to 1 over P or we can also call it recurrence interval, which indicates the average time lag or time gap between two occurrences of an event. It could be given as 1 over P where P is the probability. If probability of an event occurring in 1 year is let us say 1/2 then the return period will be 2. It means on an average, the event will occur once in two years. We should be careful with the term return period. It does not mean that the event will occur every T years. It does not mean event occurs every T years. This is important because an event may have a probability of .5 years. It does not mean that it will occur once in two years. We can take the example of a coin.

(Refer Slide Time: 05:32)



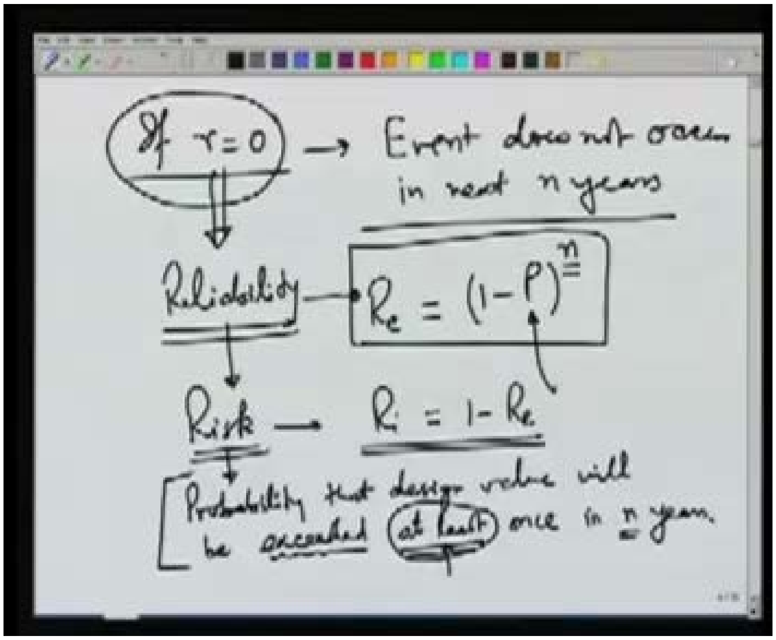
If we toss a coin we know that the probability of heads or tails is 1/2. This means that on an average, out of every 2, the return period in this case would be 2, 1 over the probability. This is the probability of heads or tails and therefore the recurrence interval would be 2. We can say that on an average, in every 2 tosses of the coin 1 head and one tail will occur. It does not mean that if we toss the coin twice or let us say 4 times, there will be 2 heads. If we take large number of trials, (the key word here is that large number), if we toss the coin let us say 1000 times, then we would expect there would be 500 heads and 500 tails. It is not that every two occurrences will have 1 head and 1 tail. But on an average, every 2 occurrences will have a head. That is why we should be careful in using this return period. It is not that if we have a return period of 10 years, the event will occur every 10 years or once in every 10 years, it may occur twice in a period of 10 years or it may not occur even once in the next 10 years.

(Refer Slide Time: 07:13)



The other term which we use or we can base on this value of P is, let us suppose we have n years of record and these will be consecutive years. Suppose P is known to us and we want to predict the number of times the event will occur in the next n years. We can define a probability, the probability of x being greater than or equal to X. This event occurring r times in let us say next n years. In theory of probability, there are equations which give us this probability and we can write it as Prn which is the probability that x will be greater than capital X or equal to capital X r times and n is the number of years. We would deal in terms of years, so in case of annual flood we will say that n is the number of years in which the event is occurring r times and this can be given in the probability theory. There is an equation which relates and this is nothing but ncr, factorial n over factorial n – r, factorial r, then we have probability of occurrence and probability of non occurrence. If we look at this equation, this is the combination parameter ncr, P is the probability of occurrence of the event and 1 – P is the probability of non occurrence. This equation gives us the probability that an event will occur r times in next n years. Based on this, we can decide at what probability the event will not occur at all in the next n years.

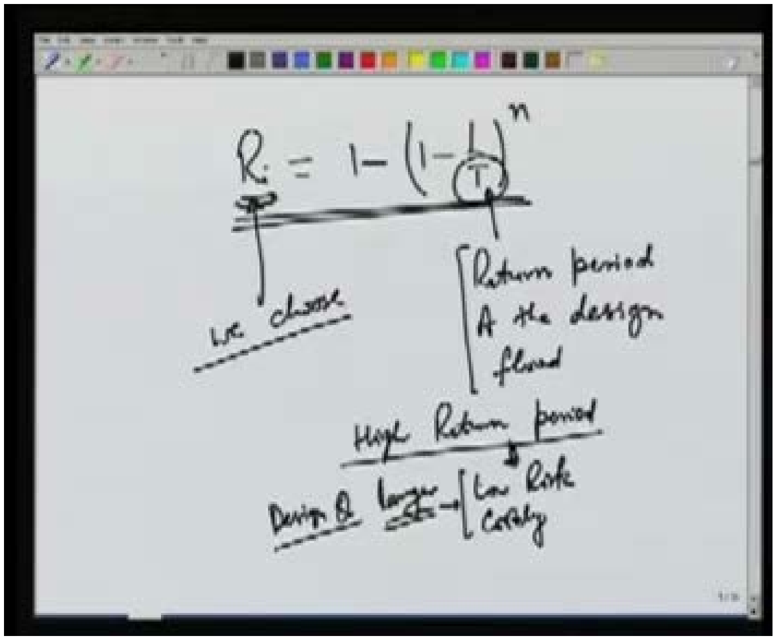
(Refer Slide Time: 10:14)



If r = 0, this indicates that the event does not occur in next n years. This is a turn in which we are interested quite a bit because if we design a project with some value of let us say design flood and its design life is n years, we would like to know what the reliability is. We will call this term reliability because this indicates that if we design some structure for n years of useful life then in those n years, the design flood will not be exceeded because r = 0 which means that the event is not occurring even once.We can write the reliability R_e by putting r = 0 in this equation and therefore the equation which we get is 1 – P to the power n. By putting r = 0, this term will become 1. This term will also become 1 and this term 1 – P to the power n will be left. The reliability is given by this simple equation, if we know the probability of occurrence of the event in a single year then in any years, the probability that the event occurring even once is the reliability of the structure or that project.

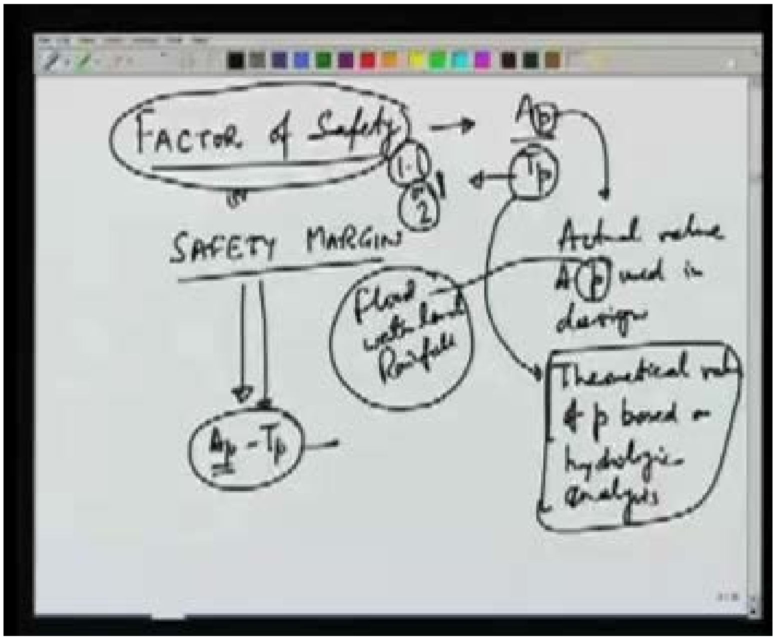
We say that the reliability of the project is the probability that the event will not occur during the life time of the project which is let us say n years, and therefore the structure will be safe because it has been designed for that design flood. Similar to reliability, there is another concept which we associate with it, which is risk. Let us call it R_i. This is nothing but one minus reliability or R_e which is logical and this risk is the probability that design value will be exceeded. In this case we will talk mostly about the floods. We will use the exceeded term here and in drought, it will be lesser than that. The probability that the design value will be exceeded at least once in n years, where n will be the design life for the useful life of this structure. At least once means it may occur more than once but at least once if it occurs, then we will assume that if the flood is more than the designed flood, then the structure is likely to fail. Our risk associated with the project will be this probability and therefore we can also write it as one minus to (Refer Slide Time: 13:58) the power n or since P is 1 over T that is the return period of the particular design flood. We can put that value here or here and get the reliability and the risk in terms of the return period.

(Refer Slide Time: 14:18)



We can write the risk as one minus (Refer Slide Time: 14:26) where T is the return period of the design. Let us call it design flood or design discharge. If we use the design flood for a particular return period, we can get the risk associated with that design flood and risk is something which we have to choose. We choose the amount of risk we can accept for a particular structure. If the structure is very important, for example if there is a dam failure, the dam may cause a lot of loss of life and property. We would try to have minimum risks or very small risks. If the structure is such that its failure will not lead to loss of life or property, it may just be inconvenience to people, for example there may be a culvert on a road and the flooding of that road will not cause loss of life or property. It may be just inconvenience to the travelling people that the road may be blocked for some time but the loss will be minimum. In that case we can go for the higher risk and to reduce the cost. As this is clear, if we use a low risk, then the cost becomes large because low risk means that the return period has to be high. High return period would mean low risk but would mean it is a costly affair because Q will be larger. The design Q is obviously flood once in 2 years or once in 100 years, they will be very different. Flood once in 100 years flood will be very high. If we design for that, the structure will be very expensive and risk will be very small. We have to balance the cost and the risk based on our engineering judgment as to how important the structure is.

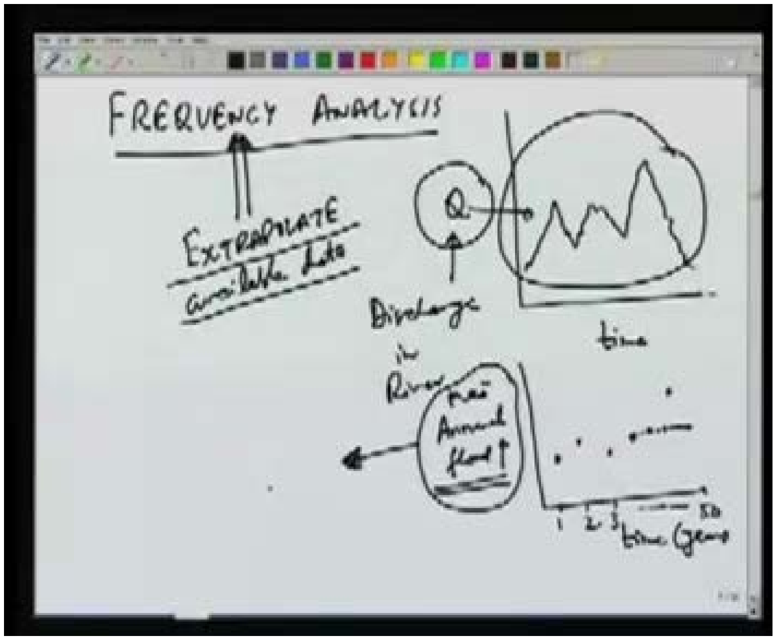
(Refer Slide Time: 17:07)



Similar to the concept of risk and reliability, there is a concept of safety which we can describe in terms of factor of safety or sometimes we use the term safety margin. They are a little different. Factor of safety is a ratio as an event lies and it is defined as the actual value used for some parameter P divided by theoretical value of the same parameter P. This parameter may be rainfall. It may be flood, it may be water level. A_p is the actual value of p used in design and T_p is the theoretical value of the same parameter. This p may be flood, design flood, it may be water level, for example river, it may be rainfall. So this may be any parameter P and T_p is theoretical value of P which is based on the hydrologic analysis. The factor of safety means that once we find out T_p using theoretical considerations, we would further increase it by some factor to take care of other uncertainties. We perform a theoretical analysis but this analysis also is based on some assumptions and therefore to be on the safer side we would increase the value of Tp by some factor.

This may be for example 1.1 or may be 2. It may have any range and it will depend on what is the importance of the structure. On a more important structure, we will probably use a higher factor of safety and a safety margin is nothing but the difference of Ap and Tp. For example if theoretical flood comes out to be 10,000 m cube per second. The actual value which is used may be 11,000 meter cube per second with the factor of safety of 1.1. Our safety margin will be a 1000 meter cube per second. This means that if our design flood is let us say 11,000 meter cube per second. This means we have a margin of safety of about 1000 meter cube per second. The actual maximum which we expect is only 10,000 meter cube per second but in some cases, if there is higher flood due to unforeseen circumstances upto 1000 meter cube per second, higher flood can be accounted for. This gives us a margin which is further safe in our design. Risk reliability and factor of safety are all these are very important concepts whenever we design something which depends on chance. All these hydrological variables for example flood, rainfall, etc would depend on chance and therefore these 3 must be accounted for the risk reliability in the factor of safety. Let us see how we can obtain the probability and how we can analyze a given data this is typically done using frequency analysis.

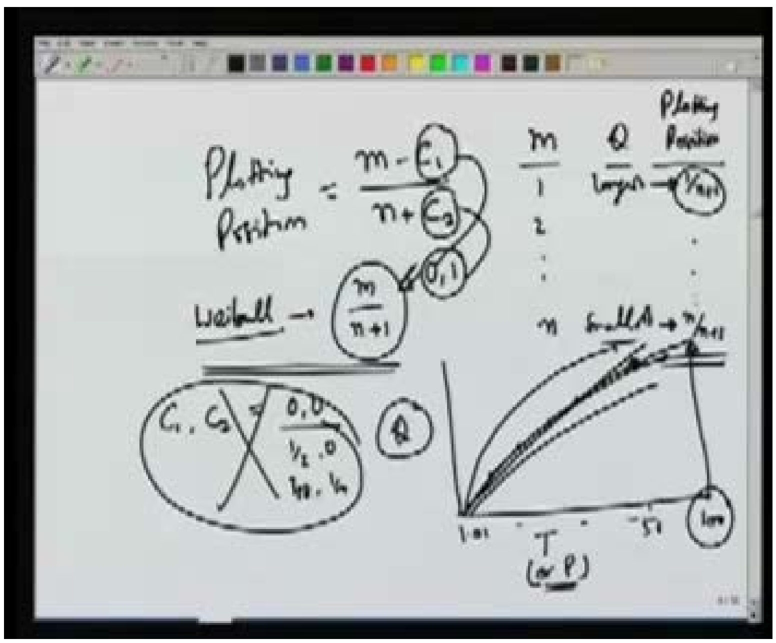
(Refer Slide Time: 21:13)



The data which we have will have values of parameters given like time, maybe Q. Suppose we are dealing with the flood in a river, then this Q may be the discharge in the river. The time may be daily, hourly, it may even be weekly or ten days. That depends on what frequency of measurement we are using but the discharge values in the river may be shown in continuous curve. It may look like this (Refer Slide Time:) will respond to rainfall in the catchment area or slow to ground water river corrosion. All these factors will decide the shape of the discharge curve we will get. Based on this data, this can be thought of the time series. Q is a variable which is the function of time and we can think of this as the time series and then there are lots of methods commonly called used. These are the frequency analysis methods. This can be used to obtain the frequency distribution or the distribution with time. We can say that suppose this Q data is available to us for 10 or 30 years and we want to design our structure for let us say 100 year term period.

One aim of the frequency analysis is to extrapolate the available data. If the data available is only for 10 years or 30 years and we want 100 years, then we would have to do some kind of extrapolation and those frequency analysis techniques can be used to decide. Suppose we are interested in the flood and time is in years, 1, 2, and 3 and suppose we have this 50 year of data here, we can say annual maximum flood or maximum annual flood. For the first year, what is the maximum flood in the river for the second year, what is the maximum value for the third year what is the maximum value and so on. We would have 50 of these values available to us and based on these 50 values, we have to decide on the probability of certain events. We may have extrapolated to get 100 year flood or 200 year flood. When we choose this maximum annual flood, generally what we do is for a single year, we take the maximum value and we ignore the values which are smaller than the maximum value. There is one other option in which we can consider all the floods and then choose. Let us say 50 largest floods. In that case within the first year there may be 2 occurrences of floods. In hydrologic, we use maximum annual floods such that only one which is the largest flood in a year is accounted for. We consider only the largest flood in a particular year. If the second largest flood in a year is larger than the other largest flood of any other year, then we do not consider that. So the maximum annual flood is for a particular year. Suppose there are n records we will arrange them in order. So let us say we have here the rank and there Q. These n records are arranged in order like m will be going from 1, 2 and suppose n is 50, or we can use some number n here. The largest value would be written here and the smallest here. These would be in decreasing order of magnitude. For example if you look at these 50 values, suppose the largest value occurs in year 40, then that value will go as rank 1 and the smallest of these floods may be will occur in let us say year 30. Then that value will be at the end the smallest.

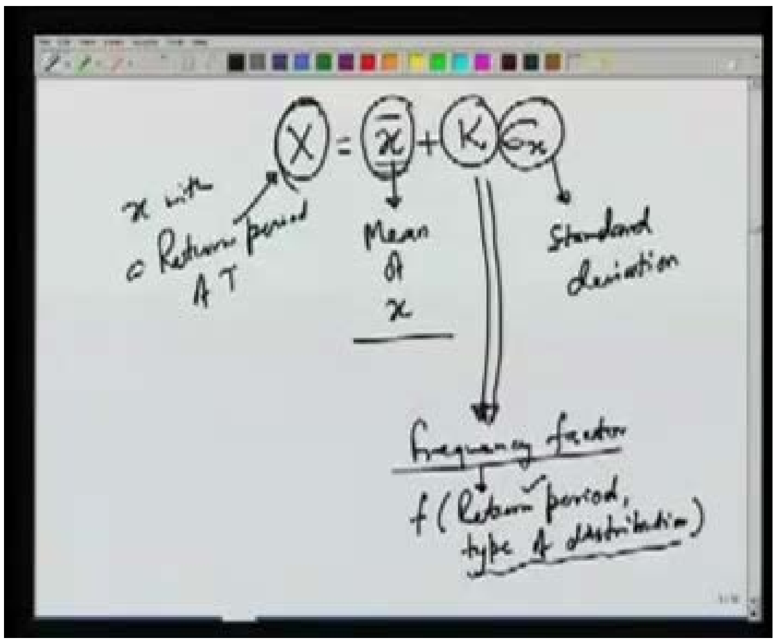
(Refer Slide Time: 27:26)



Once we arrange them in the order, we have a table like this where we have m and Q. We define a new term which is the plotting position. So m goes from 1, 2… upto n. Q, largest and smallest and the plotting position is generally taken as the plotting position m – some constant C_1 over n + some constant C_2. These constants vary for different techniques. It is the most commonly used techniques known as the Weibull plotting techniques and in the Weibull technique we have m over n + 1. The plotting position here which is used is 1 over m over n + 1, so for example in this case, it will be 1 over n + 1 and for the last value it would be (Refer Slide Time: 28:48). So this tells the probability of accidents of the particular event. For this discharge this would be the probability for this discharge. This would be the probability and the recurrence interval is just one over probability. If we do that we can plot either probability verses Q or we can plot recurrence interval T verses Q. T or P since P is 1 over T, we can plot either one. They are essentially the same. The time period or the recurrence interval would let us say go from 1.01

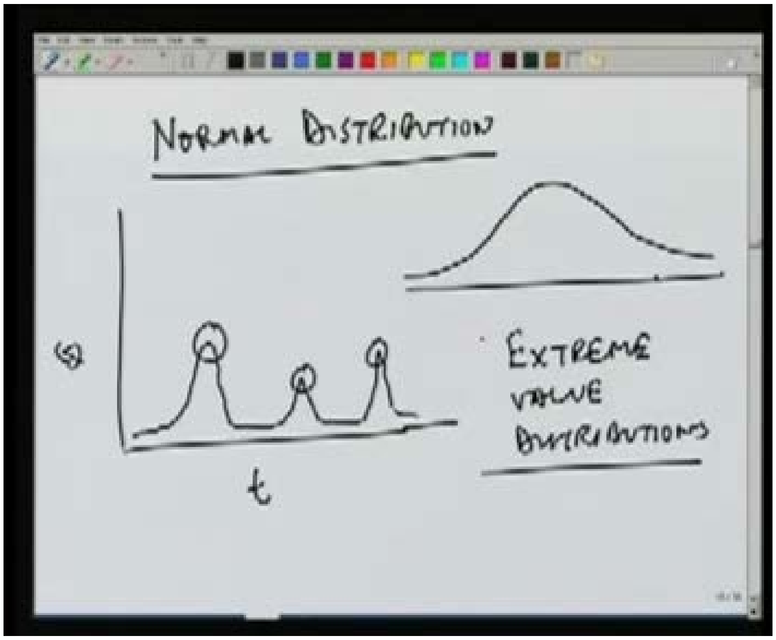
and we may reach recurrence interval of about 50, 100 and so on. From the data plotted from this, we can see that the largest value will have a high recurrence interval than the smallest value will have. There will be some kind of plot here which will tell us about recurrence interval verses discharge relationship and then we can extrapolate it. We can extend following the same trend can extrapolate it based on extension of this curve. What will be the hundred year flood in the river? This probability plotting is an essential component of extrapolation. Weibull is just one method, there are other methods too. For example you can use C_1, C_2 as 0, 0 or you can sometimes use 1/2 and 0, sometimes 3/8 and 1/4. These are various combinations but we would not be using these. We will just stick with $C_1 = 0$ and $C_2 = 1$ so 0 and one which gives us the Weibull plotting position.

(Refer Slide Time: 31:07)



All these frequency distributions are typically written in terms of (Refer Slide Time: 31:10). There is some mean value of the variable x bar and K sigma x. This is the mean of x, so x may be the annual maximum flood and this is the standard deviation. This X is T year return period of T. We want to find out the value which has the return period of T. It will depend on the mean the standard deviation and a frequency factor, which is the function of return period and type of distribution. If we look at this figure, whether our distribution is like this or like this or it is like this or it is a straight line like this. So depending on what kind of distribution we have, for the variable X in this case Q, we will have some equation which will relate the frequency factor with the return period and the type of distribution. Once we know the value of K, we can find out the X for any return period T and use that in our design.

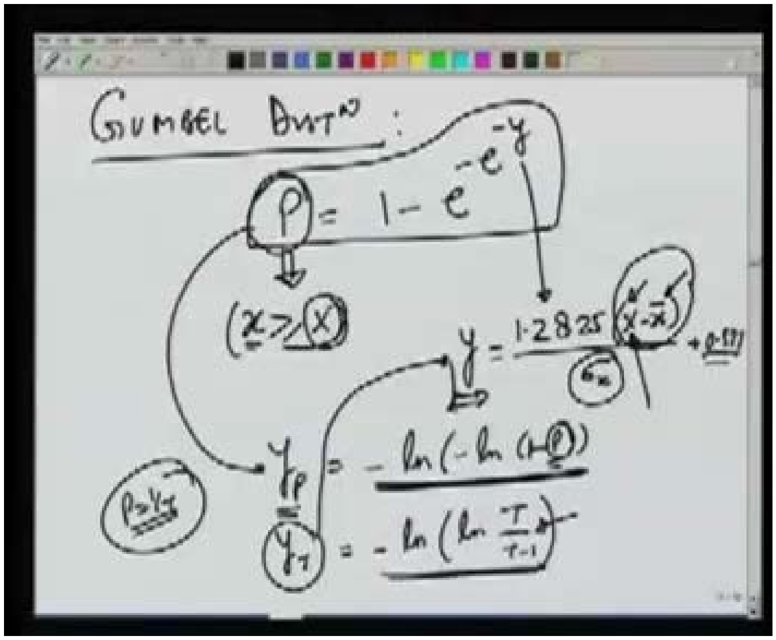
(Refer Slide Time: 33:22)



We shall discuss some of the commonly used distributions. We know that we mostly use the normal distribution for most random variables. Normal distribution or which is also known as Gaussian distribution is quite common for most of the variables but here we should realize that the events which we are talking about are extreme events. Q may be normally distributed but the maximum Q in a year are extreme events. If we look at discharge in a river, it goes (Refer Slide Time: 34: 00) in one year. Let us suppose this is maximum, and in the other year there is some other maximum and so on. So this is a very simplistic approach in which we are saying that most of the times it is almost constant. Once there is a big flood, (sometime in the monsoon) it goes like this (Refer Slide Time: 34:13).

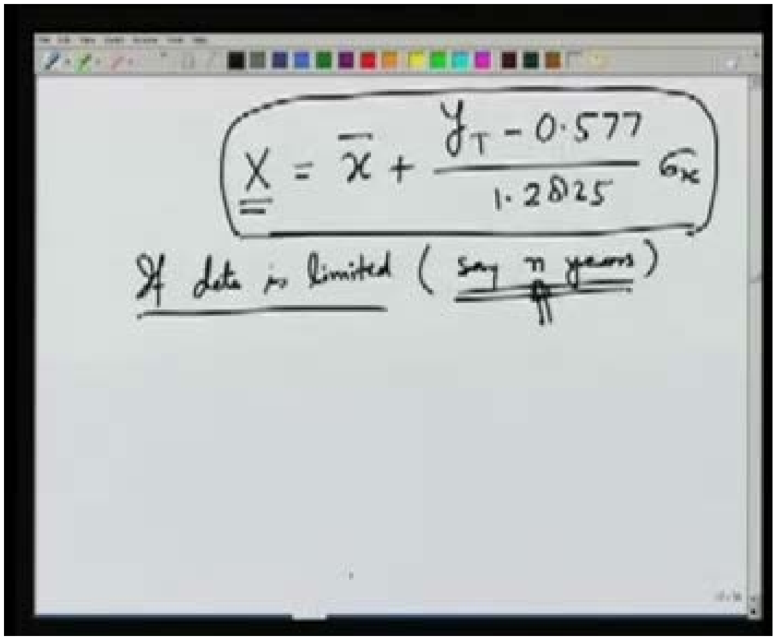
The actual distribution of Q may be closer to normal but distribution of the extreme events which are the floods will not be normal and therefore we do not use the normal distribution for extreme values. There are some distributions which are known as extreme value distributions which are commonly used for analyzing the maximum floods in a river. We will look at 2 or 3 common distributions, for example Gumbel distribution, and then there is a Pearson type three distribution and we normally take the log of variable, so that we called that log of log Pearson type three distribution. Sometimes we take log normal distribution where we say that the values are not normally distributed, but their logarithms are normally distributed.

(Refer Slide Time: 35:26)



Let us start with the Gumbel distribution and the Gumbel distribution is probability is given in terms of a variable and again the probability that x is greater than r equal to some X is given by this equation. y is a new variable, we can call it a reduced variable. It depends on the mean and the variants of x. So 0.577 + 1.2825 X – x mean. This term represents the deviation from the mean and then we divide by the standard deviation sigma x to get a reduced variable y. Based on this y we can obtain the probability of accidents of x from that value X. This probability distribution for any give P, we can also obtain y. If we say that we want the reduced variable y for a given probability P, then we can use this expression which is nothing but transformation of this equation to Y_p. Based on return period too, suppose we know the return period T, then simply using P = 1 over T, we can write this in terms of y_T as – ln, ln of T over T – 1. If we know the return period, we can estimate what is Y_T. We can put that y_T here and get X corresponding to that return period or if we know the probability, the same thing could be found out y_P from this. They are identical equations except that we are using this probability in one case and return period in other case and then we can obtain the X corresponding to that probability or return period using this equation.

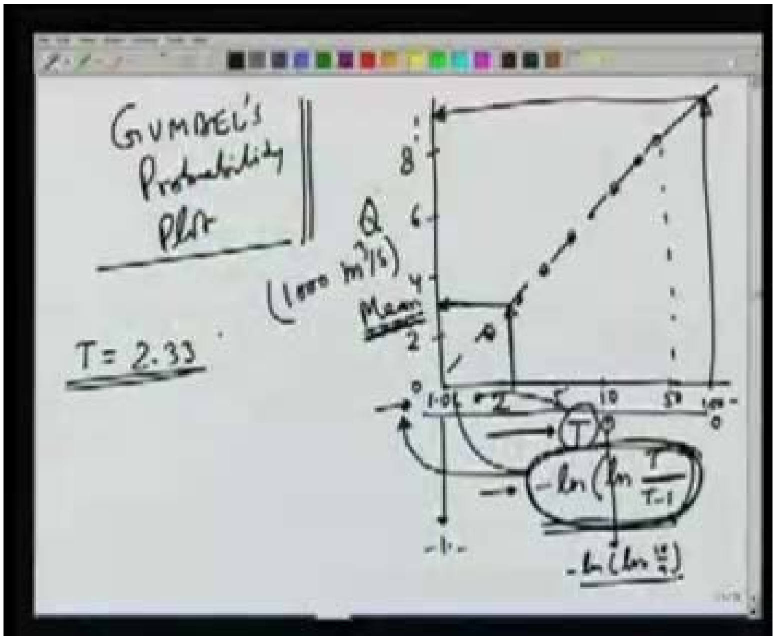
(Refer Slide Time: 38:21)



The equation which we use is the X = x bar. For any given time period, we can find out y_T from this equation. If T is known, we get y_T from this and knowing Y_T, we can get X corresponding to the return period T. This equation is valid when we have very long period of records. If the data is limited we will not be able to use this equation. Let us say n years, so this equation which is based on the Gumbel distribution is valid for a long period of record. If it is limited to n year data, then we will have to modify this equation a little bit. By comparing that Gumbels equation with this, we see that this factor K which is the frequency factor in Gumbels distribution is equal to (Refer Slide Time: 39:46) so this is K. This K is valid for very long period of records. If we say data is limited to n years of record, then this K has to be modified and the equation which we use for K becomes y_T y_n bar over S_n. Now this value and this value which are mean and a standard deviation would depend on n. That is why they subscript n is put there. y_n bar is generally the function of n of course, y_n bar is a reduced mean for n years of data and for example, if n is ten then, y_n bar is about 0.5.

If n is about 50, then y_n bar is about 0.55 and as we know, if n is infinity then we get the old value 0.577. There is a table of values given in various books and other references which provide the value of y_n bar verses n. You can have a table of values in which n will be given and the corresponding y_n bar value is also given. This table can be seen in various books. Similarly this S_n is reduced standard deviation. This also is a function of n and again for different n values, the value of S_n is given in books and other references. Tables of n and sdr S_n will be given for different n values and we have different sd values or S_n values for infinity of course, you get 1.2825. Some other values are for example n = 10, the value is about 0.9,5 for n = 50 it is about 1.16. These values again can be looked up from the tables. We can put the value here and get a value of K from this equation. For a smaller number of data points n, we can modify the K value and then use that K value here to obtain the value of X. Once we have the data, we should check whether the data really follows Gumbels distribution or not, so for that, plotting of the data is required. One of the plots which is commonly used is known as the Gumbels probability paper plot.

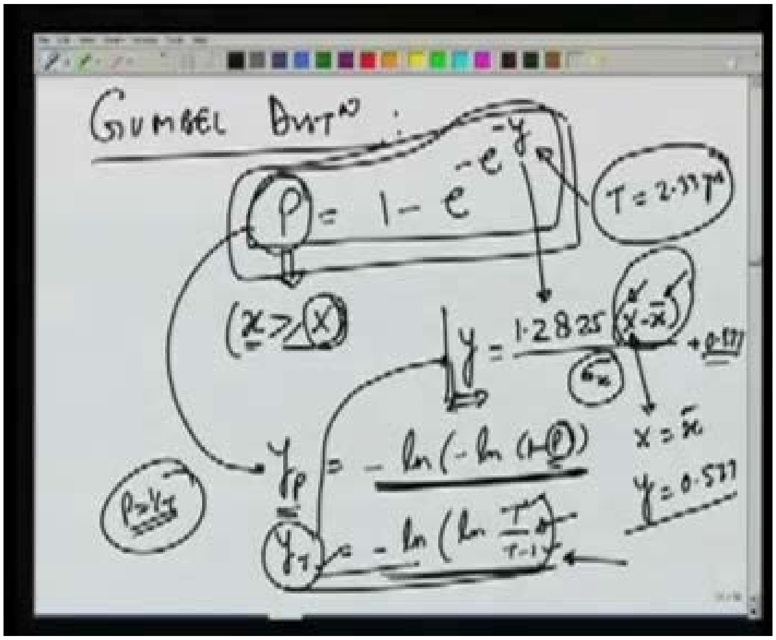
(Refer Slide Time: 43:29)



In this as we know, the time period and corresponding probability which is given by (Refer Slide Time: 43:54), so this is the time period and this is the corresponding probability. The return period T corresponds to a probability of this. What we can do is we can plot this value, but not write the values here. We write the corresponding values of T and that way we prepare a probability paper which let us say has time return period, here 1.01, 2, 5, 10, 50, 100 and so on. On the simple scale these values show the time period, but they really indicate the value of distance. For example if you take 1.01 corresponding to 1.01 from this equation, will give us a value which is around minus one point something. Similarly when you take the value of 10, you will get – ln, so it is natural log of 10 over 9. We can compute this value and that will be the value corresponding to 10, similarly for 100 and so on. On this one, you can plot the time period here and then the values of let us say discharge which will be meter cube per second and sometimes when we are talking about floods, this will be in 1000 meter cube per second because the discharges are generally of the order of a Q thousand meters cube per second.

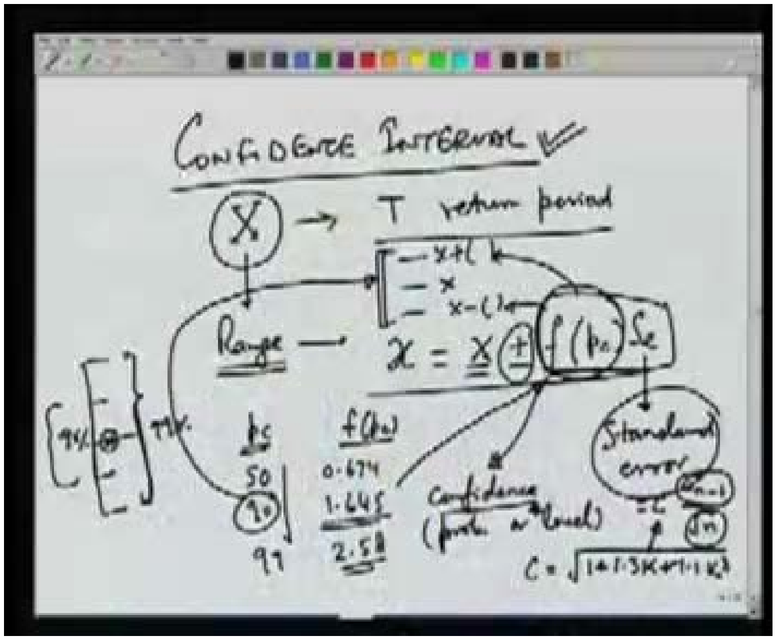
That value may go from let us say 0, 2 4, 6, 8 and on this curve, if we plot the observed data and it shows a straight line that means our data follows the Gumbel's probability closely and from this suppose we have data up to 50 year return period, then we can extend this line and estimate the 100 year return period flow or the 100 year return period flood using this distribution. We can note down in Gumbels distribution, there is a point which is, if you say T = 2.33, that is the flood which has a return period of 2.33, this will correspond to the mean of X or Q in this case.

(Refer Slide Time: 47:18)



We can see from this equation when the X is equal to mean. When X equal to the mean y is equal to 0.577. By putting y = 0.5577 in this equation, we will get a return period of 2.33 years. Using this equation or this equation we can get for y = 0.577, T equal to 2.33. In the Gumbel's plot the value corresponding to 2.33 would be the mean of all the values. This is one distribution which can be used. As we have already discussed, the data has some uncertainty. The 100 year flood which we predict will also have some uncertainty and therefore one other thing which becomes important to know is what is known as the confidence interval.

(Refer Slide Time: 48:33)

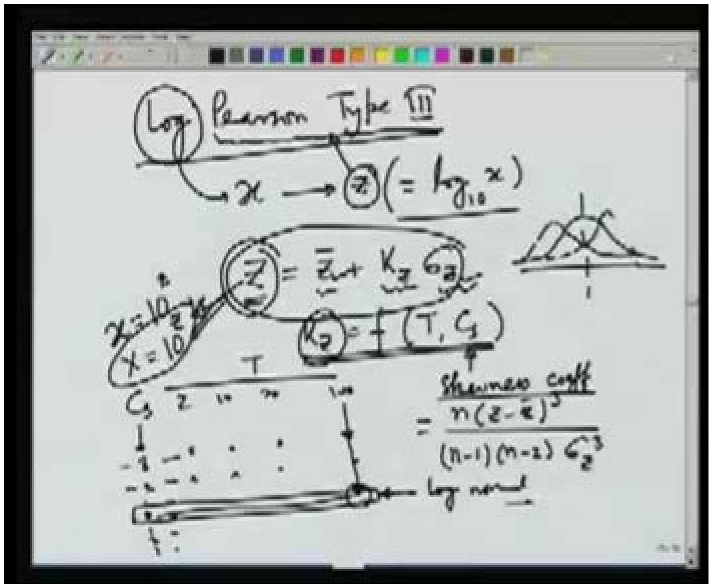


Let us say the X is T return period flood. Then what is the confidence level which we attach to this X or what would be the range so that this X is the single value, what we would now like to have is the range of values which will have 90 percent or 95 percent confidence interval. Since there will be some errors in our estimates, we call this a standard error. The fpc is the factor which is dependent on the confidence probability or confidence level. Our confidence level may be, let us say 90 percent or 95 percent or maybe 99 percent.

Depending on that, this factor will change. Some values, for example if we have a 50 percent confidence probability, the value are about .674. This is the confidence level or confidence probability and this will be a function which we use here. If we have about 90 percent probability then x plus minus means, from the expected value which represents the design flood, we will go plus minus this value and there will be two values. So one would be with x in the middle, then here we have x plus this thing and the third value is x minus the same thing. So within this range, there is 90 percent chance that our actual value will be within this range. If we want 90 percent probability, we use 1.645 factors here. To get the two values of x, their standard error is given by some constant c. A standard deviation of the data points which are n data points. So sigma n – 1 standard deviation, divided by square of n multiplied by some constant c and there are empirical equations available to express the c as the function of k and one of the equation which is commonly used is 1+1.3 k + 1.1k square. This equation can be used to obtain c and then we can obtain the standard depending on the probability. For example if we have 99 percent probability, we have to use a large multiplying factor, 2.58 and therefore this range will become higher. If we have this expected value x then the range will be here. Let us assume 90 percent, for 99 percent, the range will go even higher. We say that with 99 percent confidence, we can say that a 100 year flood will be within this range or with 90 percent confidence; we can say that the 100 year flood will be within this range i.e., the confidence interval is also an important concept because there is really so much uncertainty in that data that we cannot say that there is a single value of

100 year flood. This confidence interval gives us the possible range of values and we can choose the higher or lower values depending on the importance of this structure.

(Refer Slide Time: 53:43)



The other type of distribution which is used commonly for extreme value is known as the log Pearson type three distributions. There is a distribution which is a Pearson type three and when we say log Pearson type three, it means instead of a variable, let us assume x. We will be using log of x and say that the log of x follows a Pearson type three distribution. The equation for this is, first we transform x to z where z is nothing but log of x and then z will follow a Pearson type three distribution. First we do log of x and then z will follow Pearson type three distributions which is given as (Refer Slide Time: 54:38), so now we will be doing in terms of z rather than x, but we use the equation which is similar to the previous one. We have this frequency factor, mean, and a standard deviation and this is return period p value. k_z again is the function of the return period and there is coefficient which is known as Sheuness coefficient which depends on the particular distribution. We can change the values of C_s and we can get different distributions within Pearson type three. This Sheuness coefficient is defined as n number of records, z minus z bar cube. This Sheuness in a sense represent how far away from a symmetric distribution we are. If we have a symmetric distribution, the Sheuness will be 0.

We may have sometimes a positive sheu or a negative sheu. So, depending on that we get a Sheuness coefficient. It may be positive or negative or it may be 0 if it is completely symmetric than this. There are tables available for this K_z is a function of C_s. There is C_s value here. There is a time cleared here, so it may be 2 years, 10 years, 20 years, and 100 years. C_s, let us say – 3, – 2, 0, 1, 2, and 3 so on. For these C_s values, the value of k will be given in the table. K for this combination would be given like this. The values which correspond to 0 are known as log normal distribution because they have 0 Sheuness which means they are symmetric and therefore it is like a normal distribution. Depending on this, we can find out the K_z then we can find out the z from here and once we know z, then x can be easily obtained as 10 to power z and therefore capital X which is our design value would be 10 to power z. These distributions help us in obtaining the value of the variant, for any return period. If you want 100 year flood, we will go here. Suppose we want to use log normal distribution, we will go to C_s is = 0. Find out the value of K_z, from the mean of z and standard deviation of z we can find out capital Z and then find out X. We have looked at some methods of extrapolating available data. It means, suppose we have a record for 30 years or 50 years and then we want to extrapolate, find out what is the 100 year flood, then we can use some of these standard extreme value distributions and extrapolate the values. We have also looked at the concept of risk reliability and safety factors and in the next few lectures we would look at how to obtain the design values for different kind of structures.