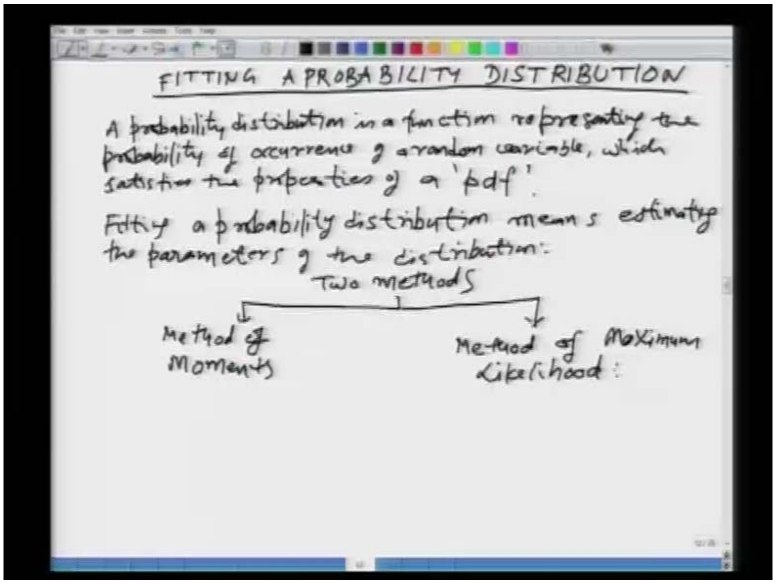
Advanced Hydrology
Prof. Dr. Ashu Jain
Department of Civil Engineering
Indian institute of Technology, Kanpur

Lecture – 31

Good morning and welcome to this post graduate video course on advanced hydrology. We are in chapter 11 of (( )) book on a statistical hydrology. In the last class we looked at the definitions of a pdf probability density function and the cdf which is cumulative distribution function. We also looked at the equation for a very popularly used pdf which is called an normal distribution or the normal pdf. We said that it is used in a standard reduced variate z and then we said that the values of the cdf of this normal distribution are available in standard tables. Then we looked at an approximation of this normal cumulative distribution function which is given by certain researches which is very practical or useful in implementing this normal distribution in a computational in environment or in a computer program.

Then we moved on and we try to look at various characteristics of a pdf in the form of what is called the statistical parameters. We looked at three main parameters or statistical parameters of a pdf which are those well; the first one was we said expected value which is equal to the mean or population mean, and we defined the equation and so on. And the first movement is always taken with respect to the origin. The second movement we defined was what is called the variants and it is the expected value of x minus mu whole square or it is the second movement about the mean. The first statistical parameter is a measure of a central tendency whereas the second one is the measure of the spread in the pdf. The third one is the third movement about the mean which is the expected value of x minus mu to the power 3. We define that with respect to the pdf, and then we said that it is an indication or a measure of or quantification of the Skewness in the data or in the pdf. That is where I think we stopped you know yesterday, and what we are going to do today is we will look at couple of methods of fitting a probability distribution.

(Refer Slide Time: 02:47)



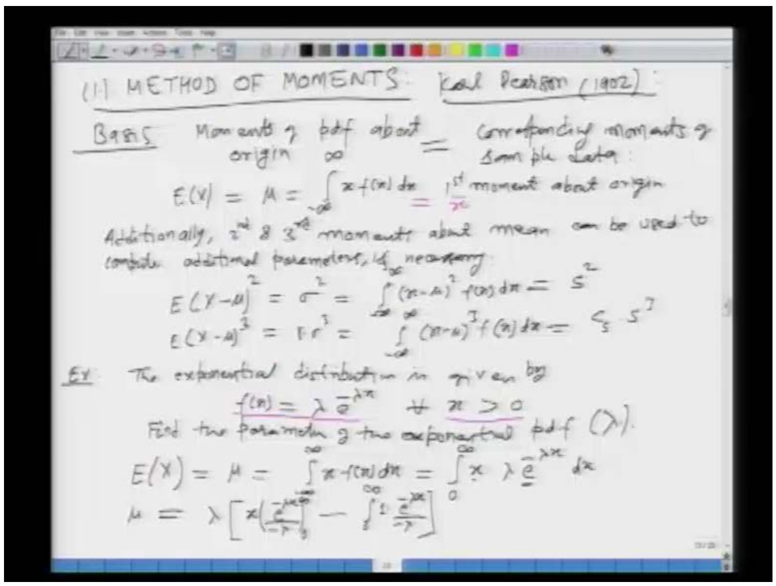
So, we start with fitting a probability distribution. What do we mean when we say that we want to fit a probability distribution to a given set of data? Well this is like a model calibration you have a model which has a pre defined structure and each model would have a certain parameters or some coefficients which need to be calibrated.

So, fitting a probability distribution basically is an equivalent to model calibration in stochastic hydrology. A variant what you are trying to do is, we are trying to find out the estimates or the values of the parameters of that particular distribution. For example, let us say you have the annual rainfall data for last 50 years and you want to fit a normal distribution to that. Normal distribution has a certain parameters it may be a b c d or it may be you know mu and sigma or lambda or whatever. So, what we do is, we try to find out the estimates of those parameters using certain methods.

So, let us look at that first of all we would like to say what is a probability distribution? A probability distribution is function representing the probability of occurrence of a random variable which satisfies the properties of a pdf. So, what is a probably distribution it is any function which satisfies or meats the properties of a pdf and it gives you the probability of occurrence of a random variable within a certain range or a particular value.

Now what do you we mean by fitting a distribution as I just said fitting a probability distribution, fitting a probability distribution means estimating the parameters of the distribution, there are many methods of estimating the parameters of a probability distribution given a set of data, but as I said in this course we will look at only two methods. What are those two methods? First one is the method of movements and the second one is the method of maximum likelihood.

(Refer Slide Time: 07:23)



So, let us get started and first we will look at your method of movements. This was propose by Karl Pearson this particular gentlemen in 1902 long time ago. And this is the very popular actually the most widely use method and it is very easy. So, what we will do is so first let us look at what is the bases of this method and based on that we will just go ahead and look at one numerical example of how to find the estimates of a probability distribution. The parameters that are using this method so we will understand this method based upon a numerical example, but first let us look at the concept on which it is based which is very simple. So, what does the bases say a basically says that the movements of the pdf about the origin is equal to the corresponding movements of the sample data.

So, it is as simple as that so what we are saying is that, the movements of the pdf with respect to origin are equal to or is equal to what the movements with respect to origin taken from the sample or estimated from the sample. Now, what is the movement we have learned which is taken with respect to origin we have learnt only one which is called the population mean or mu or expected value. So, what we have is then expected value of random variable x which is equal to what mu and how is it defined?

You integrate the whole thing over the whole range of which function your x multiplied by f of x and then dx this is your first movement about the origin. That is what we defined yesterday. If we have a pdf which involves only one a parameter to be estimated then we can take this and then we will be fine. So, we need to evaluate only the one movement which we can take normally the first movement about the origin.

However we may have a pdf in which there are more than one parameters to be estimated what do we do in that case, well we take the higher order movements and then we say that the higher order movements that is a second movement about the mean and the third movement about the mean they are equal to what they are the they are equal to the corresponding movements taken out from the or estimated from the sample.

So, we say that additionally that is if required second and third movements about the mean can be used, to do what to compute or estimate additional parameters as I said certain pdf will have two parameters or three parameters if necessary. So, let me actually first say that this is equal to your sample estimate means what x bar and.

If you are doing the second movement which is the expected value of x minus mu whole square which is nothing but the measure of the variants this is sigma squared. How is defined the integral of your x minus mu whole squared f of x dx that you says equal to what the sample variants s square. similarly, you will have x minus mu to the power 3 which is a measure of the a symmetry or the skewness in the data which is let us say I am going to say gamma times sigma cubed that is what it is and this is your minus infinity to infinity of your x minus mu whole cube of your f of x of your dx. So, this is nothing but your coefficient of skewness times your s of q or you know something similar we can perform.

We are saying here essentially is that you take the movements of the pdf and then you equate them with the sample model that is the main base. It is very simple to say, but let us see how it actually works out with the help of one example. It will the understanding or the exact method will be you know more clear once we look at that a numerical example. So, I am going to take up an example today on this. The exponential distribution or pdf is given by this equation that is the f of x is equal to lambda e minus lambda x for all values of x that are greater than 0 with this find the parameter of the exponential pdf. If you look at this equation, what is the parameter? Well the parameter is lambda. So, x is your hydrologic variable and lambda is the parameter exponential constant.

So, let us see how we can find out this lambda using the method of movements. What does the method of movements say? Population movement is equal to sample movement. So, let us first use only the first movement, so we will have e to expected value of x is equal to mu is equal to what is the integral of over the whole range minus infinity to infinity of your x f of x d x is it not? That is the definition.

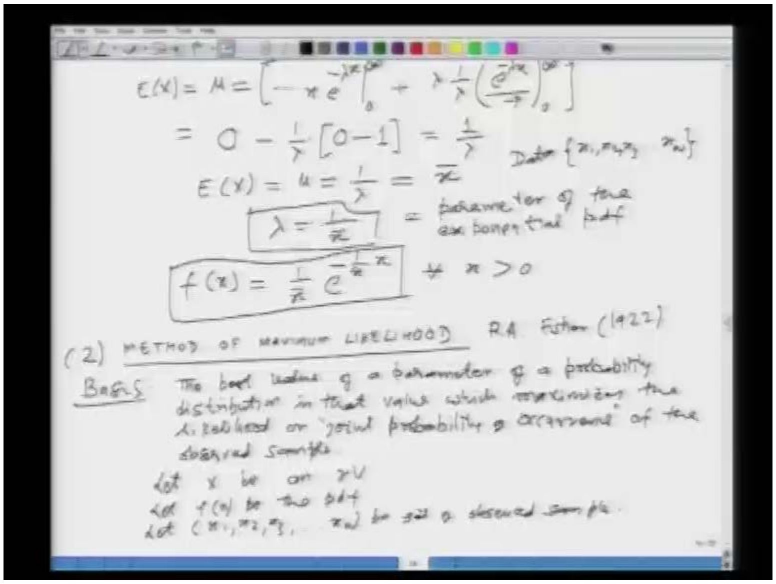
Now what is the range of the hydrologic variable, the exponential distribution is defined in this range means it is defined only for the positive values of x. So, the range if you are integral will be from 0 to infinity. So, the lower limit is 0 it cannot be negative. You have integral 0 to infinity x times f of x, what is f of x? Well you just put the equation for f of x which is the pdf.

So, you will say lambda e minus lambda x d x. Now all of you are very good mathematicians. We are all good at mathematics it is a problem of integral calculus now all we need to do now is integrate this function which is x times lambda e minus lambda x d x within the limits 0 to infinity. And then you will have what the population mean the first movement about the origin, and then we would say that that is equal to what that is equal to the sample mean. So, let us do that. So, from the knowledge of your calculus what you can do is well lambda is a constant here, so it can come out or the integral and then you have an integral of product of two functions which is x times e to the power minus lambda x.

So, you are going to use the chain rule which you have learnt in your earlier classes about the integral of the product of two functions. I am going do that and just try to follow that it should be very easy to see. So, you have first function multiplied by what the integral of the second, what is the integral of the second? Integral of e to the power minus lambda x it is going to be e to the power minus lambda x divided by what minus lambda. So, we have first function multiplied the integral of second between the same limits so evaluated between 0 and infinity.

So, this expression in the brackets evaluated between 0 and infinity that is part one minus then what do you have the integral of for the same limits. The first derivative or the first function what is the first derivative of x 1? First derivative of the first function multiplied by what integral of the second. So, you have e minus lambda x over minus lambda that is what you have I do not think any one of you should have any problem with this, this is a simple integration problem, now what we do is we just simplify this.

(Refer Slide Time: 18:40)



So, what you will have then is the expected value of x is equal to population mean is equal to what you are doing is let me go back in the first expression this lambda will cancel out, and then what you are doing is you are evaluating this whole thing between 0 and infinity. So, if you did that you will have minus x e minus lambda x evaluated between 0 to infinity plus the second one you will have lambda times one by lambda and then integral of that e minus lambda x over lambda.

So, it will be e minus lambda x over minus lambda this thing evaluated between 0 to infinity. Let me go back so that it is clear. So, the second thing what you are doing is this lambda is actually getting cancelled with this one.

So, that is what I have written here lambda times one upon lambda is coming from this lambda sitting outside the brackets and then this lambda down there this minus minus has become plus then all you have is integral of e minus lambda x inside. So, e minus lambda x integral is what this is e minus lambda x or what minus lambda. Now it is a matter of evaluating this between 0 and infinity. So, what will be the first quantity you put infinity e minus lambda x e to the power minus infinity will be 0 then you put 0 x is.

So, the first quantity will be 0 for both the cases. Then the second one you have lambda lambda has cancelled out and one upon lambda let us say you comes out and then in the brackets you will have you will put infinity first and then 0 second. So, if you put infinity you will have 0 and if you put 0 it will be 1 anything raise to be power 0 is 1. So, what are we left with? Your left with minus minus plus one over lambda

So, we have found is then expected value of x is equal to mu is equal to 1 by lambda that is all we have done yet. Now what does the method of movement say as per the method of movement this should be equal to what the sample movement for the population mean it is the sample mean. So, I am going to say that this is equal to what x bar ok.

So, if you have the data about this x let us say x1 x2 x3 up to xn, n number of data points, 50 years of rainfall or 50 years of some hydrologic variable which follows exponential distribution. Can you find x bar? Sure just take the average and equate that to your expected value of x which is one over lambda. So, what will be the value of lambda? Well lambda is equal to one over x bar is equal to the parameter of the exponential pdf.

So, how can you then redefined your exponential distribution it is going to be one over x bar times e to the power of minus lambda x, so it will be minus of one over x bar times what x that is all this is your final fitted exponential distribution were one upon x bar is your lambda and it is to only for x greater than 0.

So, for positive values of random variable this was about the method of movements in which we said that the population movements are equal to the sample movements that is all. Now we are going to look at the second method which is also, very good and most popularly used actually and gives you more accurate estimates of the parameters of a particular pdf this is called method of maximum likelihood. You may avoid this term, but what we will do is we will define what this method of maximum likelihood is? What is the likelihood what is the probability of the maximum likelihood and then.

We will see how we can implement this method. Second one is method of maximum likelihood, again before I actually go it was proposed by this particular gentlemen RA fisher in 1922 a little later than the method of movements and same thing what we will do is, we will look at its bases first, we will look at the basic concept on which this method of maximum likelihood is based on. And then we will define the maximum likelihood function and we will see how we can implement this method to estimate the parameter of a distribution and will take up an example to find out the parameter and the example we will take will be the same.

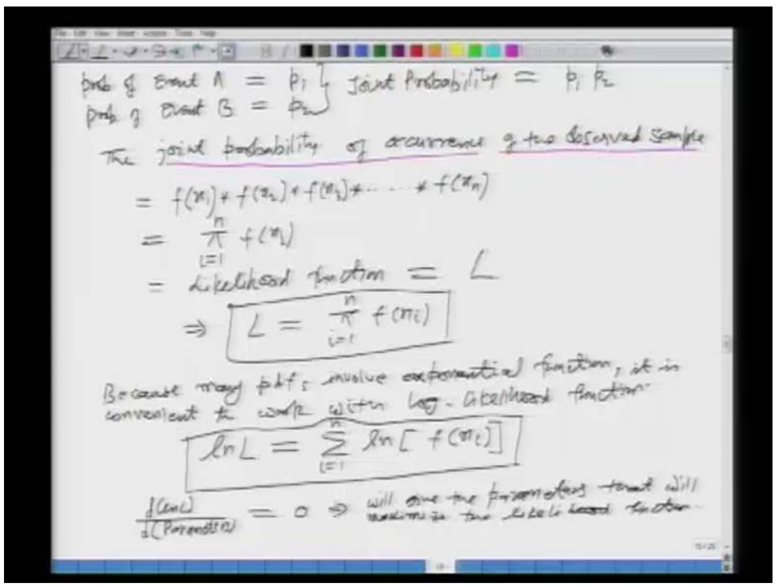
We have just seen which the exponential pdf is so that we can see that you get the same parameter using the method of maximum likelihood also. What is the base of the method of maximum likelihood? It says that the best value of a parameter the best value of a parameter of a probability distribution is that value which maximises the likelihood or what do we mean by likelihood? Joint probability of occurrence observed sample. So, this is the bases. So, what we are saying is that the best value of the parameters of a probability distribution is which ones are the ones which maximise.

So, we are talking about some kind of optimization here which maximises the likelihood function or the joint probability of occurrence of the observed sample. Let us say you have the data points annual rainfall 10 different values for 10 years. Now, how can you fit a particular distribution or normal distribution or any other distribution to this?

Well, what you do is you formulate what is called the likelihood function which will give you what, the joint probability of occurrence of that sample and then you maximise that joint probability of occurrence. So, it says slightly a little more rigorous as for as the mathematics is concerned, but once we formulate that function or the likelihood function things will fall into place. So, let us look at this 10 let capital x be an rv or a random variable let f of small x be the pdf probably density function which we are trying to fit and also, we need a sample.

Let us say x1 x2 x3 and xn are what be a set of observed values or the sample or observed sample. So, now before I actually go to this method what I would like to do is I would like to go back, and revise it some of the laws of probabilities which are going to be used in formulating the likelihood function which is nothing but the joint probability of occurrence.

(Refer Slide Time: 28:43)



so, let us say that if you have, let us say event a the probability of event a let us say is p 1 and the probability of occurrence of some another event is let us say b is equal to let us say p 2 we have 2 different events. And let us assume that they are independent they do not depend on each other.

So, there is no conditional probabilities involved now if I were to ask you what is the probability of joint occurrence of events a and b that is what is the probability that both event a and event b will occur what would that be? We have seen from your laws of unions and intersections earlier what is that joint probability or from the basic knowledge of your probability will be what nothing but the product of the 2 probabilities, so you have one event it is probability is p 1 another event it is probability is p 2 what is the joint probability? That is probability of occurrence of an event in which both the events occur that is equal to the product of the 2 individual probabilities and we know that what is the probability of occurrence? Of another event which says this or that.

So, it is basically you have one corrector is end for end you have the product and if you have either this event or that event then you add those probabilities p 1 plus p 2, but here we are saying event a occurs and event b occurs now with this argument I can generalize this if you have n different events occurring then what will be the joint probability of occurrence of n different events. It will be nothing but the product of the individual probabilities p 1 p 2 p 3 p 4 all the way up to p n that is what we have what so we have here we have an observed sample in which there are n different data.

So, each one of them has certain probability of occurrence which is given by your p d f so what will do is we will formulate the likelihood function or the function which represents the joint probability of occurrence. So, if I write it the joint probability of occurrence extending this above concept of the observed sample if I go back to your bases what is this bases? Telling us we have to maximise what the likelihood? Or we join the probability of occurrence of the observed sample that is I'm trying to formulate here. So, what will be the joint probability of occurrence? Of the observed sample this will be equal to what the probability of occurrence of the? First data point in the sample multiplied by the probability of occurrence of the second data point in the observed sample third fourth and so on.

All the way up to n so what is probability of occurrence of the? First one if the p d f is f x it is nothing but f of x 1 is it clear what does a p d f give? The p d f tells you the probability of occurrence of a particular value or between a range of values. So, using the p d f what will be the probability of occurrence of that particular data point which has been observed it'll be nothing but f of x 1 now you multiply this by what joint probability is what f of x 2 times f of x 3 and so on all the way to f of x ten if you have and different data points in your observed sample is this clear we are or all we are doing is we are applying the basic laws of probability here to write an expression for the joint probability of occurrence.

So, how can I write this is nothing? But I use a symbol phi if it is the sum of some expressions we use the summation and if it is the product if some you know identical things or similar things we use a phi function. So, will have this is equal to the phi will run or the products will run from I is equal to 1 to 10 there are n data points of what f of x I this I have said is equal to what it is the joint probability of occurrence? Of the observed sample and this is also we say is the likelihood function and let us say this is l a. so that means your l is what is nothing but? So, we have formulated what is called the likelihood function? Now our objective is to what is just to maximise this? If you notice this likelihood function what does it involve it involves the p d f.

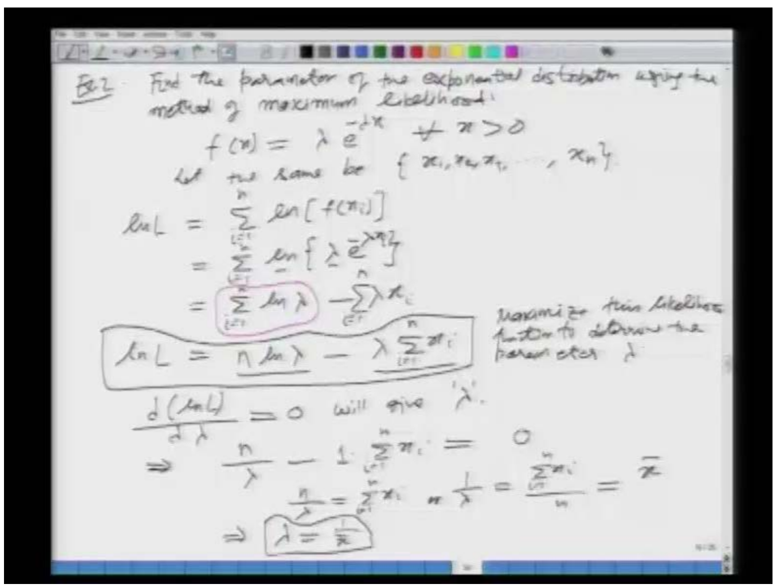
So, you have the equation of the p d f with you let us say you are trying fit an exponential distribution or a normal distribution or some extreme value distributions. So, you have the equation available with you in that in place of x you will put x 1 x 2 x 3 x four x I and then keep on multiplying those individual things you will have a complicated function which will involve the parameters. So, parameters will also be there then what you do I’m sure all of you know how you find the minimum and maximum of a function it is a very simple problem of optimization.

So, what you can do? Is you can differentiate this function with respect to what with respect to the unknowns what are the unknowns? The parameters so you differentiate this with respect to the unknowns equate them to 0 that will give you the values of the parameters. You can take the second derivative to make sure that it is a maximization problem we are not going to do that but I will do is I will formulate this function and then take up an example

So, what I am going to say before the move on to the numerical example is that because many of the p d f s in hydrology and otherwise also involve exponential functions or they involve e to the power something. It is convenient to work with log likelihood function so if you take the log of this equation which I have just written here so you take the log natural log on both side so you have natural log of l is equal to what is going? To be equal to there are phi’s phi means the product once you take the log of 2 quantities which are the product a multiplied by b what is log of m times n log m plus log n. So, what you are going to have is this product or that phi will get converted into summation I is equal to 1 to n of your natural log of what same thing the f of x.

So, this is your new likelihood function and as per as the optimization problem is concerned or the maximisation of the likelihood function is concerned the problem remains the same weather you maximise l or you maximise log of l you are going to get the same result. So, that is what we are going to do next? So, let us see we are maximising this natural log of the likelihood function so how can you find the parameters? If we differentiate this with respect to what the parameters d of your natural log of l of d of parameters if we equate that to 0 will give you what will give the parameters? That will maximise other likelihood function, so this is a very simple method again only concept here was to formulate the likelihood function. So, what we are going to do is? We will look at the same example of the exponential distribution.

(Refer Slide Time: 39:20)



So, let us do that so will have example 2 today in which we will say that find the parameter of the exponential distribution exponential distribution using the method of maximum likelihood. So, what is your equation f of x is what lambda e minus lambda x for all values of x positive and then let the sample be your x 1, x 2, x 3; all the way to x n let us say you have n data points available to you and you want to find out the parameters of the exponential distribution. So, the first thing we have to do is what we have to formulate? The natural log of the likelihood function and we have the equation so what will be the l n of l this will be nothing but what if you go back? It is the summation of natural log of f of xi where this summation is coming from I is equal to 1 to n.

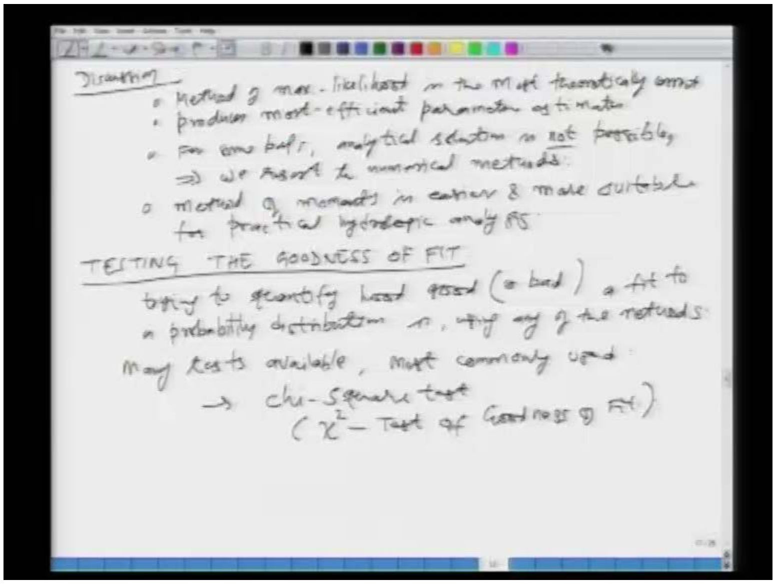
So, you have I going from 1 to n of what of your natural log of f of x I? What will that be equal to? We put the value of your p d f or equation so we have I varying from 1 to n of natural log of what lambda e minus lambda x that is what we have? Now it is a meter of some algebraic manipulations or mathematical simplification you have a natural log you have a summation outside and the natural log is the product of 2 things the product of 2 functions of a log is nothing but the summation. So, what you will have is? You will take out these 2 expressions separately first 1 will be summation of what natural log of lambda? That is the first thing natural log of lambda will come first then you will have what minus because there is an exponential is minus lambda x which will come out as outside of your lambda of your xi; e minus I should say I here lambda of your xi of what natural log of e which is equal to 1 and there is a summation outside as l so you will have this going from 1 to 1 just take a second look at it and convince yourself that what I have written is? What you will get? You have a natural log of product of lambda n e minus lambda x. So, natural log of lambda has come out summation outside similarly, natural log of e to the power minus lambda x I natural log of e is 1 so minus lambda x I and you have a summation outside going from 1 to n what will be? The first quantity this one what will be this is summation of natural log of lambda I going from 1 to 1. So, you have l n lambda plus l n lambda plus l n lambda plus l n lambda up how many times n times.

So what’ll be the result it’ll be nothing but n times natural log of lambda so that is the first quantity minus the second thing is this lambda does not depend upon I so lambda comes out so what do you have? In the summation you will have I is equal to one of 1 to n of x I is this clear this is nothing but your natural log of l this is your what likelihood function? We have not done anything yet all we have done is we have formulated the likelihood function or the log of the likelihood function for the case of exponential distribution now what we will do is? We will maximise this, so maximise this likelihood function to determine the parameters actually in this case there is only 1 parameter which is lambda.

So, what you will do is? You will differentiate this equation with respect to lambda that is d of l n l over d of lambda is equal to 0 will give lambda is it clear let us see how we can do that so if you differentiate the hand side with respect to lambda the first thing is n times natural log of lambda what’ll be the first derivative of that it will be nothing but n is a constant log of lambda what is a first derivative one upon lambda so it is n by lambda minus what is the first derivative? Of lambda times summation of xi lambda is the variable here and what is summation x I summation x i's are it is a constant these xi's are nothing but some numbers or the data point data points are what 1. 2. 3 5. 9 and.

So, on you some all of them what will you get some number let us say 239. 4 so that is a constant what is variable is lambda? So, what is the first derivative of x? It is one so here what you have is 1 times summation of xi's I going from one to n all of that should be equal to what 0 as per your maximization so what we have then is n by lambda is equal to summation of your x I or what will be one by lambda one by lambda will be summation of x i, I am not writing this it is understood I going from 1 to n divided by what n what is this on the hand side? Well summation x I is over n is x bar isn't it that means what is lambda is one over x bar so you see that we get the same result no matter which method use. So, we have looked at the exponential distribution we found out the parameter lambda using the method of movements and it came out to be one over x bar. Similarly, we took the same exponential distribution applied the method of maximum likelihood and we determine that the estimated value of lambda is one over x bar.

(Refer Slide Time: 47:40)



Just a few more points about this fitting the probability distribution few points I would like to make is that the method of maximum likelihood is most theoretically correct or theoretically most correct. Because we are using we are formulating a function we are using the optimization theory and the finding out the things also it produces most efficient parameter estimates the case of the example in the case of the example, we have taken the parameters from both the methods have come out to be same.

So, however in other p d f s you may have slightly different expressions what we are saying is that the method of maximum likelihood is more enable it is based on more correct theory, and it gives you the most efficient values of the estimated parameters of a PDF. Next thing is for some p d f s analytical solution is not possible, you see that many of the PDFs many of the probability density functions will involve a some complicated equations in which you will have multiplication of this f of x i's from 1 to n and then it may result in to an expression, which is very difficult to differentiate there may be more than one parameters let us say some PDF has 4 parameters.

So, you have to differentiate a very complex non-linear function with respect to many parameters equate them to 0 and then find out the solution you have to solve by some matrix algebra you know four by 4 or 5 by 5 matrix which may become either t d s or it may not be possible in certain cases. What do we do? In that case well we resort to what is called? That numerical techniques if that is a case we resort to numerical methods the method of maximum likelihood is so good that even if you are not able to solve it using analytical methods we would like to solve it using numerical methods because it gives us very good estimates of the parameters. However, method of movements is e g f very easy to use and more suitable or practical hydrologic analysis.

So, this method of movements also is quietly widely use because of its simplicity and use of implementation. So, this way we have looked at these methods and the next step or the next thing which we are going to do? Is we will look at or assess how good or how bad this fit is for example, whenever we are developing any model what do we do we calibrate the model. So, we have done actually the calibration part the next step in any model development is what is to assess? The goodness of fit or is called the model variation. So, the next step is going to be the assessing the goodness of fit, so that next topic is going to be the testing the goodness of fit.

So, what do we do in the model? Validation well basically like any mathematical model when you calibrate calculate certain statistical parameters or errors statistics you do that on a data set which is different when what was used for calibration? So, we calculate certain statistical parameters for any mathematical model. In the case of the probability distribution what we will do? Is same thing we will calculate some statistical parameter depending upon what kind of test? We are using the many tests available and we compare the computed statistic with the standard statistic. So, objective in the testing of the goodness of the fit is what it means basically trying to quantify how good or bad if fit to a probably distribution is using any of the methods. So, what we are saying is? That we will be trying to quantify the goodness of this fit how good this fit is that let us say you have a data set you have a particular distribution, if you do not test the goodness then it may turn out that when you start using hat probably distribution your results will not be reliable because the fit was not good.

So, if you are fitting only one distribution or only one fit then we have to asses it how good or bad it is or in certain other situations what we may want to do? Is we may want to fit more than one probably distribution. Let us say you want to try you are not sure you may say that normal distribution is suitable here or Carl Pearson distribution is suitable here. So, you may fit more than one let us say 2 distributions you have fit now you want to find out which one is better how do you do? That so you need certain kind of quantification of the goodness of fit of a model, and here are many tests available for testing this goodness of fit, and in this course we are going to look at only one of them.

So, let me say that there are many tests available for what for testing? The goodness of the fit and the most commonly used one is what for testing? The goodness of fit probability is what is called a chi square test it is pronounce as chi square written like this and the symbol is this chi square test of goodness of fit so I’m afraid I’m running out of time today. So, I would like to stop here I have given you a very brief background about the testing the goodness of a fit of a PDF tomorrow what we will do is? We will look at this up to seizer or the step by step to seizer of how we can carry out this chi square test will look at the procedure and take up and example. So, I will like to stop here today.

Thank you.