Stochastic Hydrology

Prof. P. P. Mujumdar

Department of Civil Engineering

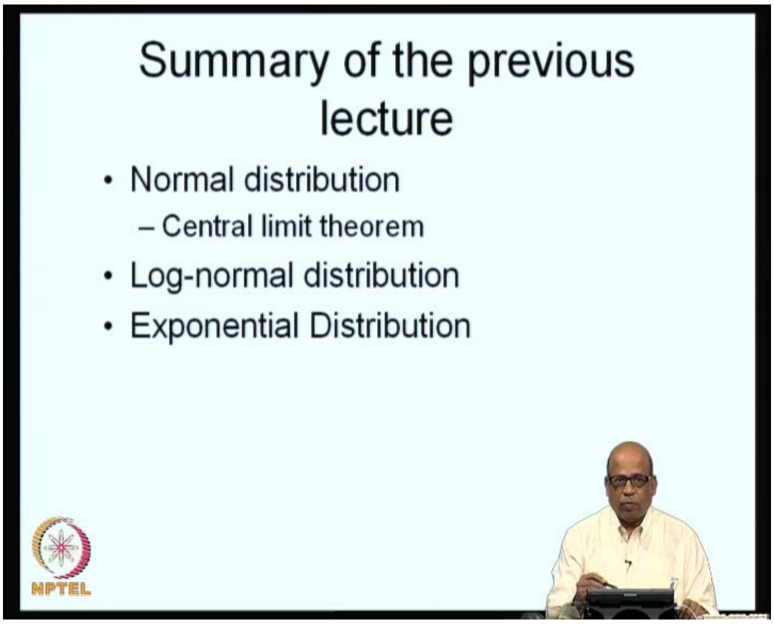
Indian Institute of Science, Bangalore

Lecture No. # 06

Other Continuous Distribution

Good morning and welcome to this the sixth lecture of the course stochastic hydrology. If you recall in the last class we covered the normal distribution in some detail.

(Refer Slide Time: 00:30)



In fact, we also solved several examples related to normal distribution. And we discussed the central limit theorem, because of which the normal distribution application becomes, so elegant and popular. Then we went on to discuss the log normal distribution, if you recall the two limitations of the normal distribution for hydrologic applications or that the normal distribution has a range the random variable following the normal distribution has a range from minus infinity to plus infinity. And therefore, there is a finite probability associated with negative values, when we use the normal distribution.

But most hydrologic variables are non negative. For example, stream flow cannot be negative, rainfall cannot be negative, and evaprotranspiration cannot be negative, etcetera.

So, most of the hydrologic variables that we come across are non negative. Unless we are talking about let say temperature in certain scales or water levels with respect to a certain thresh hold level and so on in which case the variables can be negative. The second limitation of normal distribution for application to hydrologic processes is that due to it is symmetry itself. Normal distribution, if you recall is a perfectly symmetrical distribution. It is symmetrical about x is equal to mu, but most of the hydrologic processes are they follow skewed distributions mostly they may be positively skewed with a long tail towards right and in some situations they are negatively skewed with a long tail towards the left and so on.

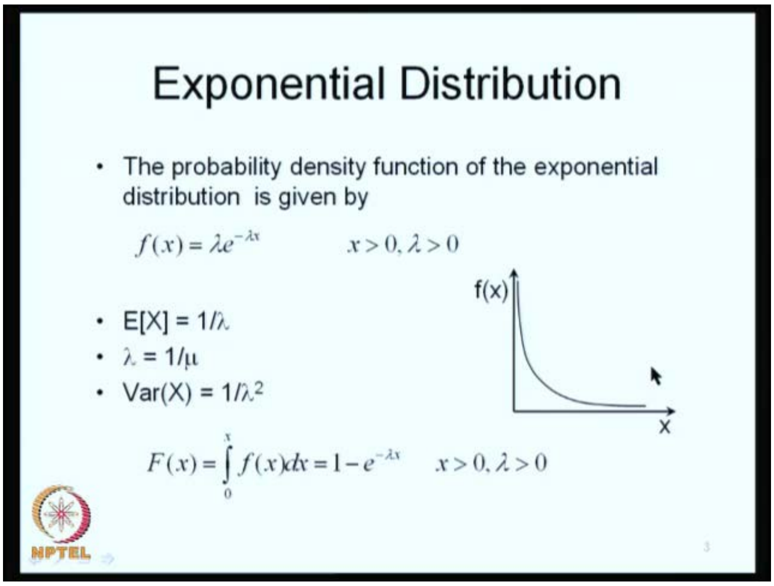
So, the symmetry of the normal distribution in fact, becomes a limitation for hydrologic applications. In such situations, we have used the log normal distribution, and in the last class we covered some details about the log normal distribution, and also we saw an application through a numerical example. Now, the log normal distribution, if you recall is a positively skewed distribution, and we say x follows log normal distribution, if y is equal to ln x follows normal distribution.

So, one easy way of handling the log normal distribution is to convert the sample values that we have x 1, x 2, x 3, etcetera; transform them as ln x 1, ln x 2, etcetera, and then work with the log transform series and obtain the mean and standard deviation of the log transform series and with respect to these we then use the normal distribution. Because log of x follows normal distribution then we went on to cover the exponential distribution exponential distribution is also a positively skewed distribution, and if you recall the exponential distribution is generally used when we are talking about time to failure of certain events.

Let us say the time to not necessarily time to failure, but time between two critical events, let us say we are talking about the time elapse between rainfall intensities beyond that certain thresh hold value or time elapse between flood events of certain magnitude. When we are talking about such intervals and these intervals themselves are random variables then we generally use the exponential distribution. Today what we will do is that, we will continue the discussion with exponential distribution, discussion on exponential distribution cover an example of the exponential distribution. I will quickly go through the exponential distribution pdf again, and how we obtain the cdf, and then how we work with the numerical examples and then we introduce the gamma distribution and move on to the extreme value distributions.

So, if you recall the exponential distribution has the p d of lambda e to the power of minus lambda x and this is defined for x greater than 0 and lambda greater than 0.

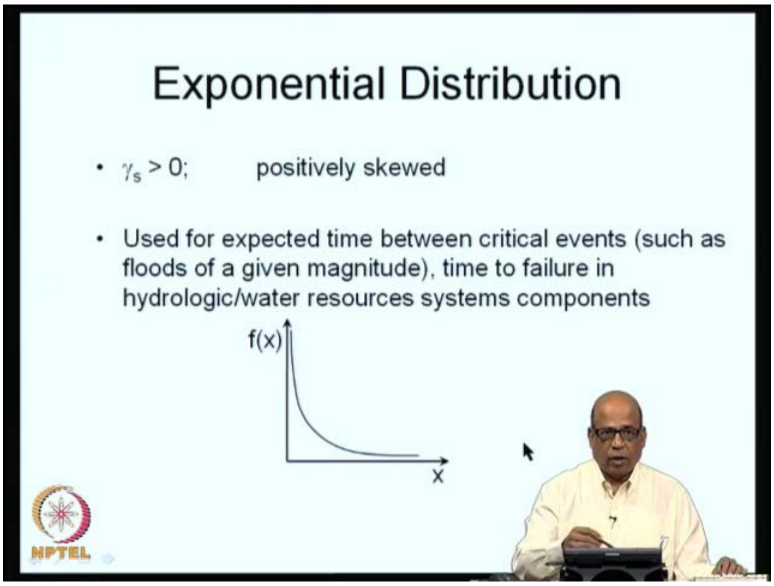
(Refer Slide Time: 05:13)



The expected value of x when exponential when x follows exponential distribution is equal to one by lambda and therefore, we can estimate the parameter lambda with 1 by lambda and variance of x is given by 1 by lambda square. The shape of the exponential distribution is like this it is in fact, the exponential curve classical exponential curve now the cdf of x when x follows exponential distribution is given by integral zero to x f of x dx which is equal to 1 minus e to the power of minus lambda x defined for x greater than 0, and lambda greater than 0.

It is a positively skewed distribution, because the gamma s turns out to be positive in this case and therefore, it has the distribution has a long exponential tail to the right. As I said it is used mostly for the expected time between critical events or time to failure in hydrologic water resource system components.

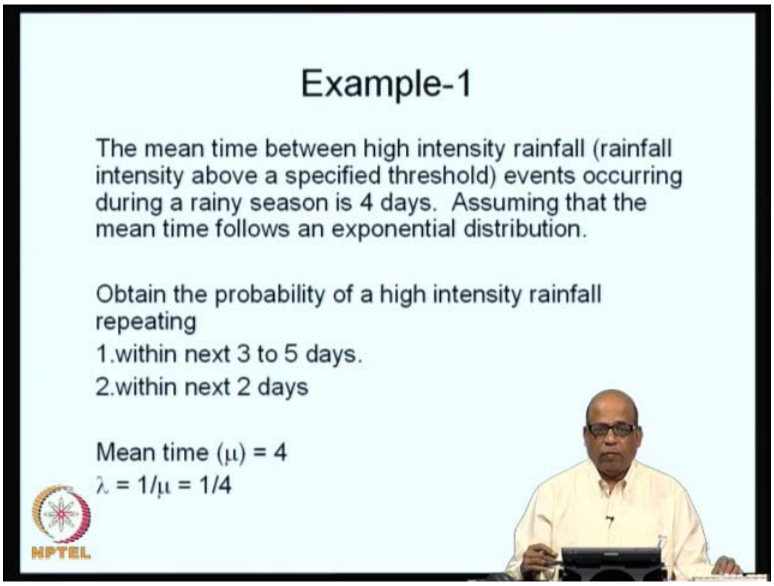
(Refer Slide Time: 06:13)



Classically it is also used in industry for industrial applications for time to failure of critical components that is physical failure we will be talking about there but in hydrology and water resources we are not so much concerned about the physical failure of components, but of the functional failures. For example, if your water resource system is suppose to produce say certain amount of hydro power, and it fails produce that amount of hydro power we call it we recon as a failure.

Obviously this failure we are talking about the failure with respect to the hydrologic inadequacy in the sense that either the head was inadequate or the supply of flow was inadequate. And therefore, the hydro power could not be produced in that particular period. Now whenever such failure occurs, we would be interested in times to such failures whenever such failures are defined, we would be interested in average time to failure again once a failure occurs when will the next failure likely to occur. In such situations we use the exponential distribution, let us take the take a simple example here where we are interested in the time between high intensity rainfall events.

(Refer Slide Time: 07:44)



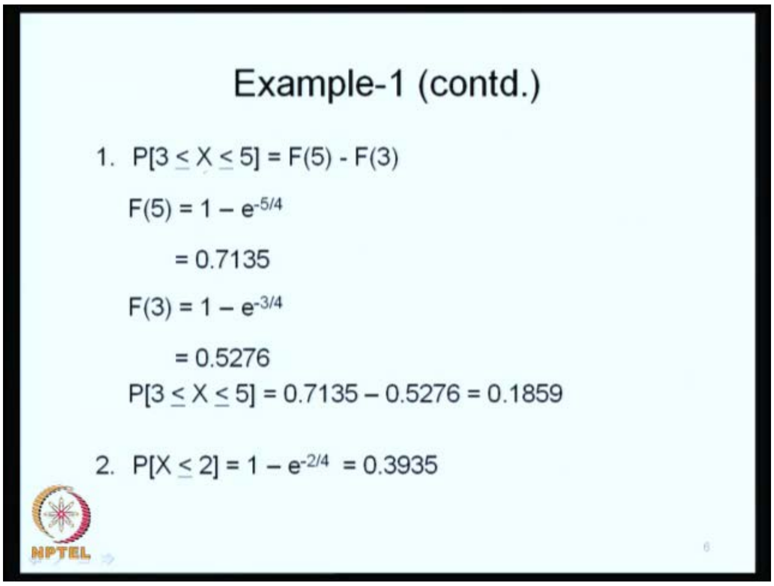
So, we may define rainfall intensity above a specified threshold as a high intensity rainfall event now, these kinds of situations come when we are talking about urban flooding for example, where rainfall intensities of a given magnitude let say 90 centimeters, 90 millimeters per day or 100 millimeters per day etcetera. These kinds of intensities may become critical; in fact, those of you who have gone through urban hydrology course will know that the time of concentration in the urban hydrology situations will be quite small.

For sub catchments it may be as small as 12 minutes, 15 minutes and so on. And then the intensities the design intensities that we will be talking about would be of the order of 90 millimeters per hour. Obviously, 90 millimeters per hour occurring over 12 minutes and not over the whole 24 hour period and so on. So, in urban situations, as well as in fluvial flooding situations where river flooding is concerned, we would be interested in obtaining the recurrence time of high intensity rainfall. So, this is one such example where we are saying the mean time between high intensity rainfalls which is defined as rainfall intensity above a specified threshold.

High intensity rainfall events occurring during a rainy season are 4 days which means on an average this kind of intensity occurs once in about 4 days assuming that the mean time flow follows an exponential distribution, we will now obtain the probability of the high intensity rainfall occurring within next 3 to 5 days which means, let say we denote the high intensity rainfall as a random variable not the high intensity rainfall, but the mean time between the high intensity rainfall events occurring that if we denote it as a random variable x then we are asking the question what is the probability? That x lies between 3 to 5, and then we are also interested in what is the probability that x is less than or equal to 2; that means, the critical event occurs within the next 2 days.

Now, we will use the exponential distribution here as is given in the problem the mean time is 4 days, and the lambda is estimated remember exponential distribution has only one parameter lambda and lambda is estimated by 1 by mu which is equal to 1 by mu in this situation.

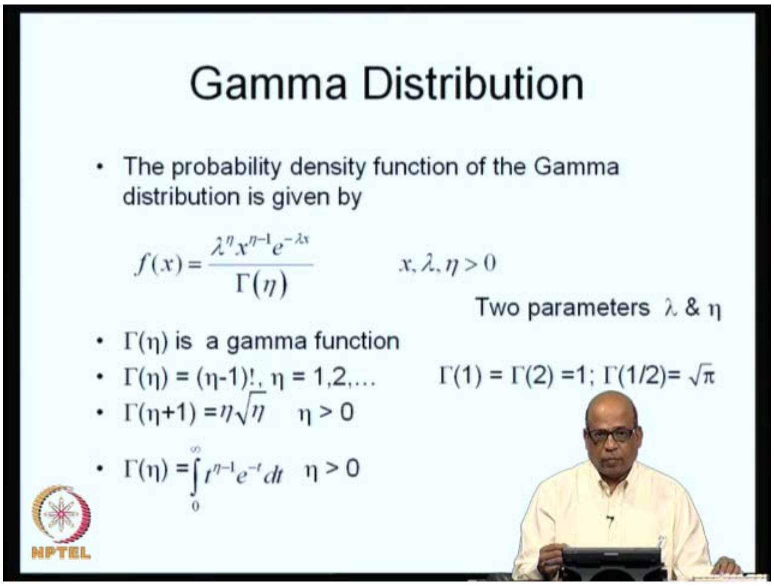
(Refer Slide Time: 11:04)



So, for this we get probability of X lying between e and 5 as F of 5 minus F of 3 from our fundamental principles. So, we obtain this as remember F of x is equal to you have for the exponential distribution F of x is equal to 1 minus e to the power minus lambda x. So, in this case lambda is equal to 1 by 4, and x is that specific value for which you are looking for the F of x. So, F of 5 will be you will have 1 minus e to the power minus 5 by 4. So, similarly you get F of 3, and therefore probability that x lies between 3 and 5 is given by F of 5 minus F of 3 net finite.

Similarly, probability that x is less than or equal to 2 is simply f of 2 F of 2 and that turns out to be 0.3935. So, this is how we obtain the probability associated with the exponential distribution much the same way as we do for any other probability distribution we simply compute the cdf the cumulative distribution function and then talk about probabilities associated with various events on x. Now, we move on to a more general distribution which is gamma distribution. The probability density function of gamma distribution is given by f of x is equal to eta to the power n. I will repeat that f of x is equal to lambda to the power eta x to the power eta minus 1 e to the power minus lambda x the whole divided by gamma of eta this is defined for x lambda and eta all greater than 0.

(Refer Slide Time: 12:41)

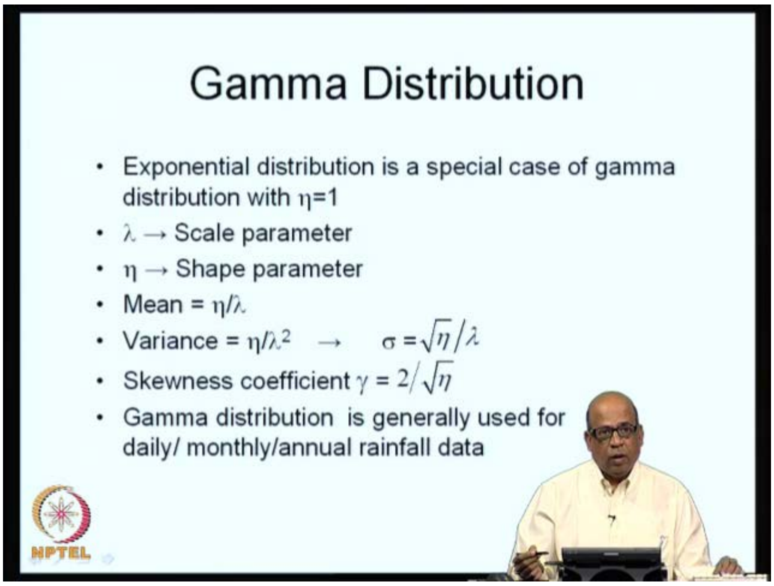


This has two parameters lambda and eta and the gamma of eta that I have written here is in fact, the gamma function gamma eta is the gamma function or is obtained as gamma of eta is equal to eta minus one factorial for eta is equal to 1, 2, etcetera. When it is a integer variable and eta gamma function of eta plus 1, this is the recursive relationship is equal to eta root eta or eta greater than 0, there is no restriction that eta should be integer here and the definition of gamma function is 0 to infinity integral t to the power eta minus 1 e to the power minus t dt eta greater than 0.

I would encourage you to go to engineering mathematics books, and then just revise your knowledge about the gamma function. Some interesting results on gamma function that is gamma 1 is equal to gamma 2 is equal to 1 which follows from this relationship here, and gamma half is equal to root pi look at this expression for f of x, this is a pdf. If we put eta is equal to 1 what happens, this becomes lambda to the power 1 x to the power 0 e to the power minus lambda x which means and divided by gamma of 1 and gamma of 1 is 1. So, this turns out to be for eta is equal to 1. When we said eta is equal to 1 this turns out to be f x is equal to lambda e to the power minus lambda x, defined for x greater than 0 and lambda greater than 0 this will recall that it is in fact, the exponential distribution.

So, exponential distribution is a specific case, a special case of the gamma distribution. So, the gamma distribution is a more general form of the exponential distribution in fact, gamma distribution is a family of distribution as we can presently see now, gamma distribution has two letters, lambda and eta, lambda is called as the scale parameter and eta is called as the shape parameter. Lambda governs the scale of the distribution and eta governs the shape of the distribution which we will see in the next line the variance for the gamma distribution is given by eta by lambda square and from which we write sigma is equal to root eta by lambda as the coefficients of skewness gamma is given by 2 by root eta.

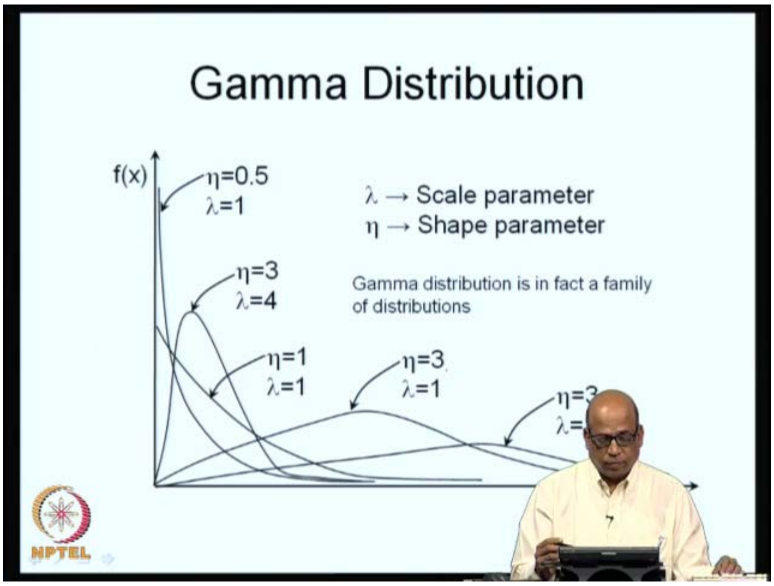
(Refer Slide Time: 16:03)



Now, gamma distribution is quite popular in hydrologic applications, mainly because it is a family of distributions. It is unlike all normal or log normal venues specify the parameters when you change the parameters, you essentially get the same scales and same shapes in the case of normal log normal exponential distributions and so on. Where as in the case of gamma distribution, you can generate a family of distributions by perturbing the parameters and therefore, you have a wide range of distributions, wide spectrum of distributions possible with the same gamma distribution pdf by varying the parameters lambda and eta. So, it is generally used for daily, monthly or annual rainfall data in applications of hydrology.

So, if you see the plots of the gamma distribution, let say eta is equal to 0.5 and lambda is equal to 1 that will plot itself as this particular curve here as eta approaches 1 the curve should approach a exponential distribution then eta is 1 to 3 lambda is equal to 4 you have this particular curve eta is equal to 1, lambda is equal to 1 you have this kind of a curve eta is equal to 3, lambda is equal to 1 you have this kind of a curve and so on.

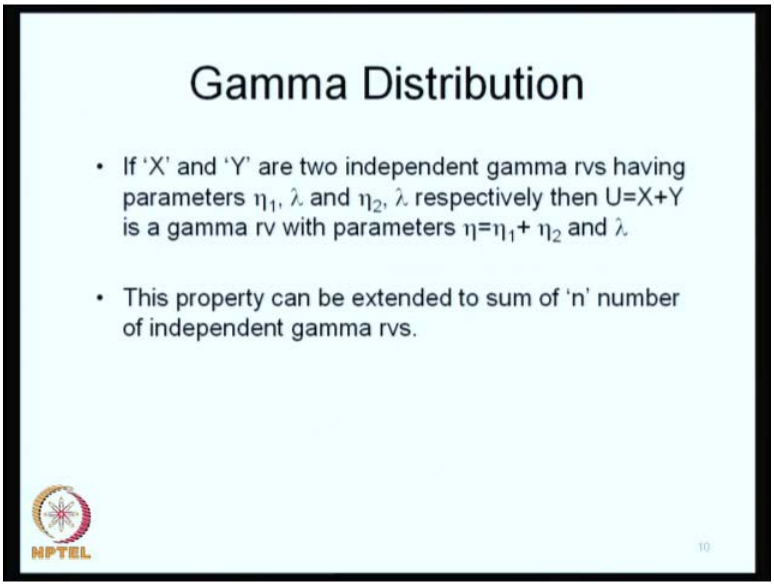
(Refer Slide Time: 17:37)



Let say you look at eta is equal to 3 you have three different curves here eta is equal to 3 here and eta is equal to 3 here and eta is equal to 3 here as you can see the shape remains the same, but the scale is different. So, as your eta is equal to 3 is constant the lambda is varying. So, the scale of the distribution is changing from this distribution to this distribution to this distribution, where as if you look at lambda is equal to 1 this is one curve where you have lambda is equal to 1, another lambda is equal to 1 here and lambda is equal to 1 here.

So, the scale by and large remains the same whereas, because eta is changing eta is equal to 0.5 here eta is equal to 0.1 here and eta is equal to I think one here one here and one more we are considering this is lambda is equal to 1. So, eta is equal to 3 here. So, as eta changes the shape of the distribution changes, for the same lambda and that is why lambda governs the scale of the distribution, it is called as a scale parameter and eta governs the shape of the distribution, and it is called as a shape parameter, and that is why gamma distribution is in fact a family of distributions you can generate several such distributions by varying the parameters.

(Refer Slide Time: 20:13)

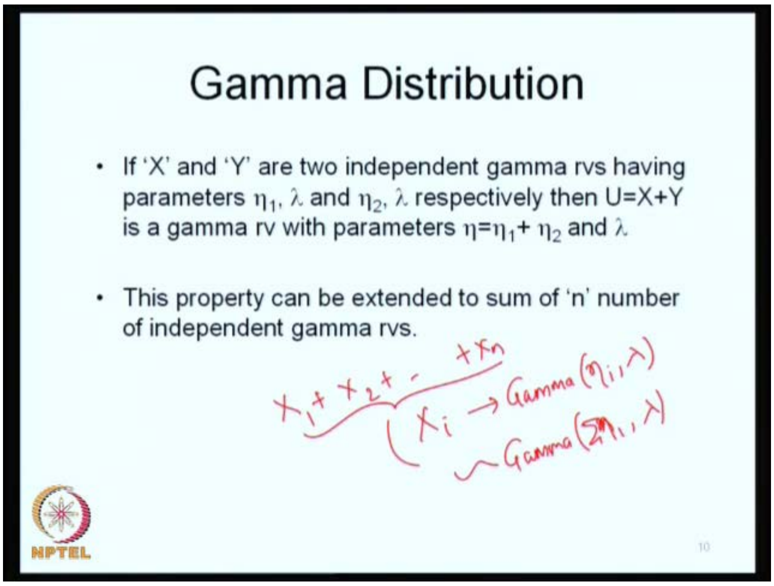


Now, there is another interesting property of the gamma distribution which makes it very elegant and popular to be used in several hydrologic applications. If x and y are two gamma random variables; that means, two random variables follow gamma distribution, they have parameters eta 1 and lambda and eta 1 and lambda respectively; that means, x 1 has a gamma distribution with parameters eta 1 and lambda x 2 has the gamma distribution with eta 2 lambda, which means the scale parameter is the same, lambda is the same for both the distribution; however, the shape parameters are different.

Then we are interested in the distribution of some of the 2 random variables. So, u is equal to x plus y is the new random variable value we are defining as long as x and y follows gamma distribution with the same scale parameter lambda. Then the sum of the random variables u is equal to x plus y also follows the gamma distribution with parameters given by eta is equal to eta 1 plus eta 2, and lambda remember lambda is same for both x 1 as well as x 2. Now, this becomes a very interesting and elegant property of the gamma distribution in fact, this has been extended to n number of independent gamma r v s what we mean by that is that if x 1, x 2, x 3, etcetera x n of all follow gamma distributions with the same scale parameter lambda, but with varying shape parameters eta 1, eta 2, eta 3, etcetera up to eta n.

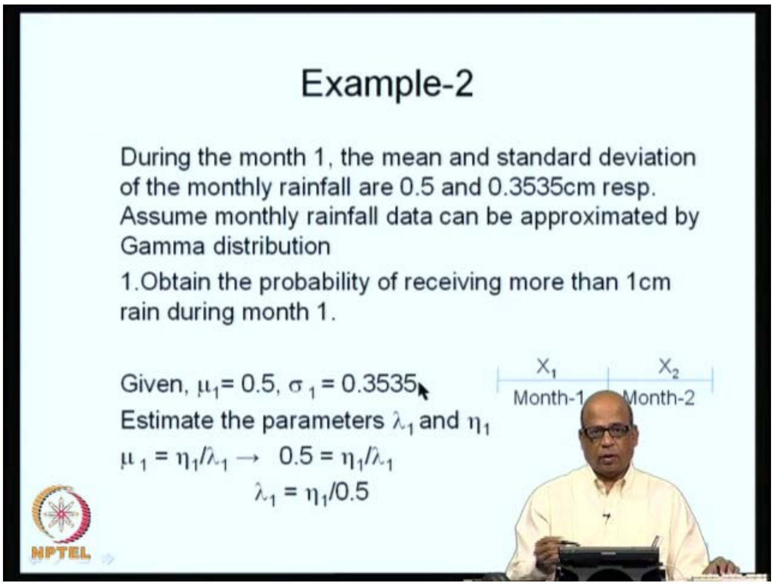
Then the sum x 1, plus x 2, plus x 3, etcetera plus x n follows the gamma distribution with the parameters that is what we are referring to now is the sum of random variables? x 1, x 2, x 3, etcetera x n and x i follows a gamma distribution with parameters eta i and lambda then this sum follows a gamma distribution with parameters sigma n eta i and lambda. So, what we stated here has been generalized can be generalized with n random variables and this is a very useful result which we use in several hydrologic applications.

(Refer Slide Time: 22:38)



Let say, we take an example where we are talking about rainfall during 2 months month 1 and month 2 these are adjacent months. So, x 1 is the rainfall during month 1 and x 2 is the rainfall during month 2, x 1 follow the gamma distribution and it has its moments given here mu 1 is 0.5, the mean is 0.5 and the standard deviation is 0.3535, x 2 also follow the gamma distribution and the properties are given in the next few slides.

(Refer Slide Time: 23:11)

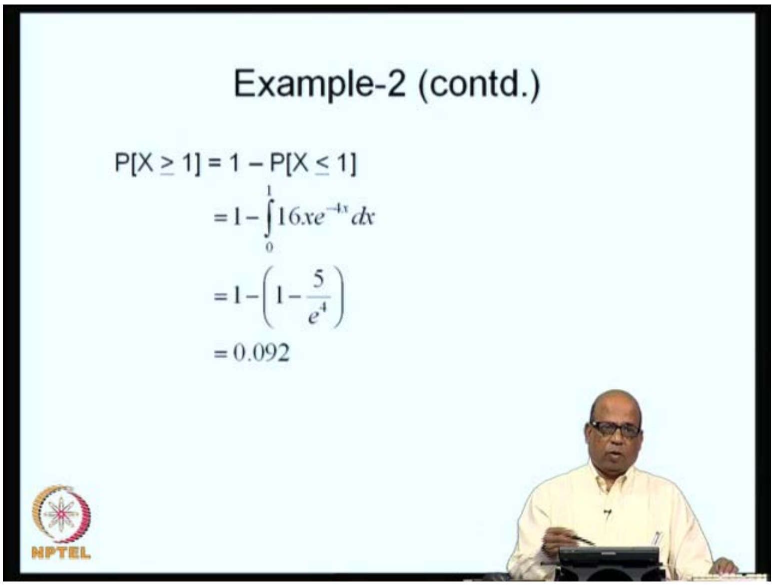


Now, we are interested first in getting the probability that, the rainfall during the month 1 is greater than 1 centimeter and mu 1 is 0.5 centimeter and sigma is 0.3535 centimeters. So, first we estimate the parameters. So, mu 1 is equal to eta 1 by lambda 1 recall that the gamma distribution has the properties, mean is eta by lambda and variance is eta by lambda square now, we use these two and estimate the parameters eta and lambda. So, when we do that we get lambda 1 as eta 1 for example, eta 1, I mean the eta corresponding to the gamma distribution for the rainfall in month 1 which is represented by the random variable x 1.

Now, that parameter comes out to be 2, eta 1 is equal to 2 from this relationship here and from that we get lambda 1 is equal to 4. So, once we get eta 1 and lambda 1, the gamma distribution is completely defined and therefore we write f of x 1 of x this denotes that we are talking about the pdf of the random variable x 1. So, that pdf can be written as lambda 1, eta to the power eta 1, x to the power eta n minus 1, e to the power minus lambda 1, x divided by the gamma function with argument eta 1. So, when we simplify that we get f of x for the random variable x 1 is equal to 16 x e to the power minus 4 x.

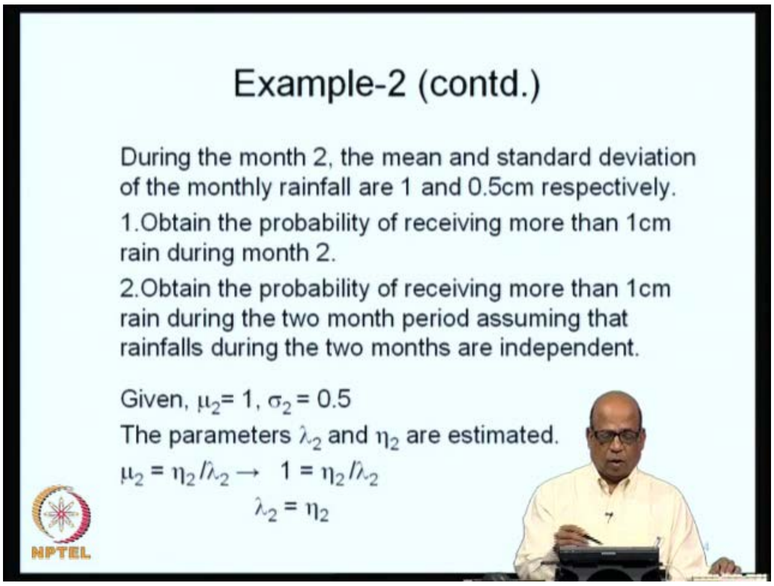
Now, once we defined the pdf you can talk about the probabilities. So, we are interested in probability of x being greater than or equal to 1 that is actually, we are interested in probability of x 1 being greater than or equal to 1. So, this will be given by 1 minus probability of x 1 being less than or equal to 1 and that turns out to be 0.092. So, we are integrating between 0 and 1 with respect to x.

(Refer Slide Time: 25:25)



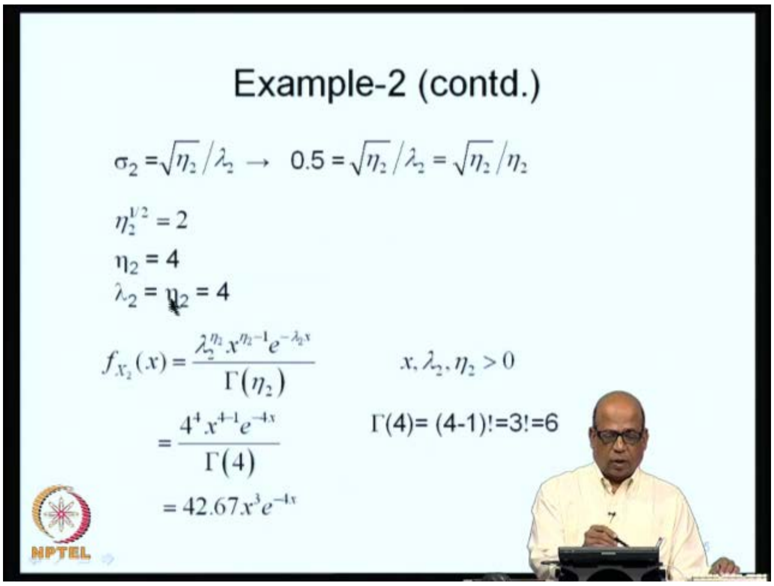
Now, for the month 2 we do the same thing that is obtain the probability of receiving more than 1 centimeter rain during month 2 and month 2 has the moments; for example, the mean is mu 2 is equal to 1 and the variance is 0.5 square.

(Refer Slide Time: 25:52)



So, sigma 2 is 0.5, we again get the parameters lambda 2 and eta 2. So, in this case lambda two comes out to be 4 and eta 2 comes out to be 4, then we obtain the f of x, the pdf for the random variable x 2 which turns out to be 42.67 x cube e to the power minus 4 x the details are given here, and once you get small f of x which is a pdf you integrate the pdf to get the associated probability.

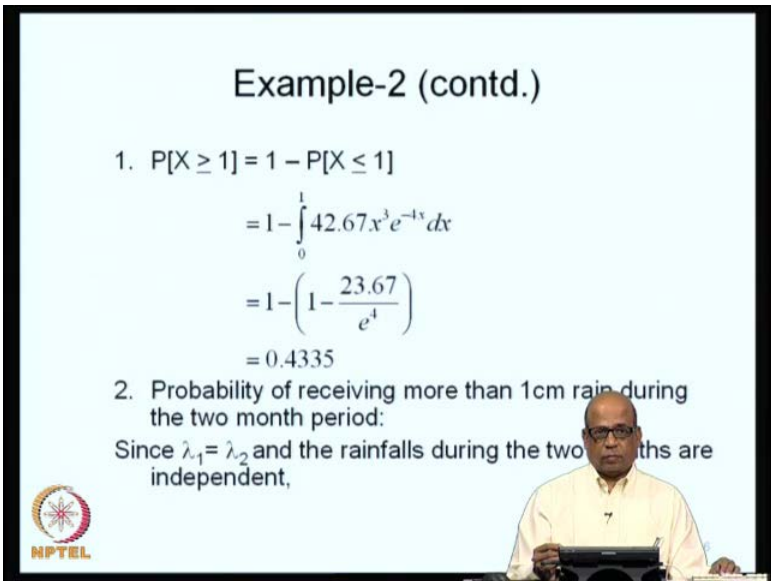
(Refer Slide Time: 26:15)



So, your interest in probability of x being greater than equal to 1 which will be 1 minus probability of x being less than or equal to 1. So, that is how you calculate this and you get the probability associated probability is 0.4335. So, you get a probability of 0.4335 for month 2 and you get a probability of 0.092 for the month 1 that means there is a probability of 0.092 associated with the event the rainfall in month 1 is greater than or equal to 130 meter. There is a probability of 0.4335 associated with the event the rainfall during month 2 is greater than or equal to 1 that is 1 centimeter.

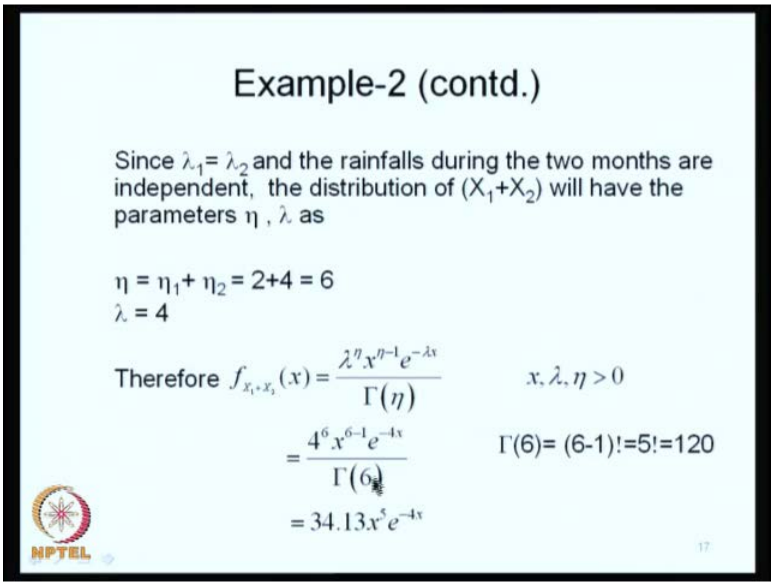
Now, we will look at probability of receiving more than 1 centimeter rain during the 2 month period; that means, x 1 plus x 2 together should be more than 1 centimeter now, because lambda 1 is equal to lambda 2, in this case as you can see lambda 2 is equal to 4 here and lambda 1 is also 4. So, the scale parameter is the same and we are also given the fact that the rainfalls during the 2 months are independent. So, we have been given rainfalls during the 2 months are independent.

(Refer Slide Time: 27:03)



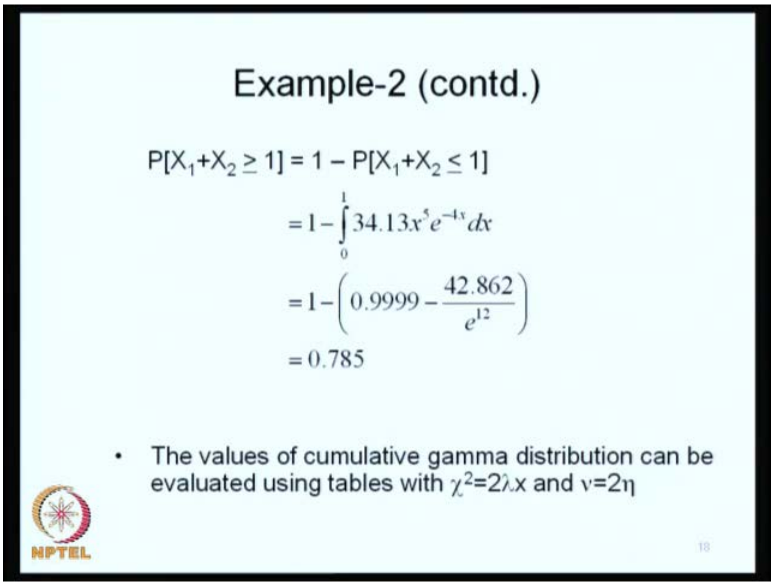
Both x 1 and x 2 follows log gamma distribution both x 1and x 2 follows gamma distributions, they both have the same scale parameter lambda 1 is equal to lambda 2 both equal to 4, but with different shape parameters eta 1 and eta 2. And therefore, we should be able to write the distribution of x 1 plus x 2. So, recall that for this some of the random variables x 1 plus x 2, the eta is given by eta 1 plus eta 2, that is equal to 6 in this case and lambda remains the same which is equal to 4. So, the gamma distribution pdf of the random variable x 1 plus x 2 with the new parameters eta and lambda, that turns out to be 34.13 x to the power 5 e to the power minus 4 x.

(Refer Slide Time: 29:00)



So, once we get the pdf again we can use the pdf to get the associated probabilities. So, we are now talking about the probability that x 1 plus x 2 greater than equal to 1 which is equal to 1 minus probability of x 1 plus x 2 less than or equal to 1, we use the pdf associated with the random variable x 1 plus x 2, which is this pdf now and then integrate between 0 and 1 and get the probability as 0.785, which means the probability of the total rainfall in that 2 month period which consists of rainfall during the month 1, x 1 and the rainfall during the month 2, x 2. The probability of that total rainfall exceeding 1 centimeter is equal to 0.785.

(Refer Slide Time: 30:25)



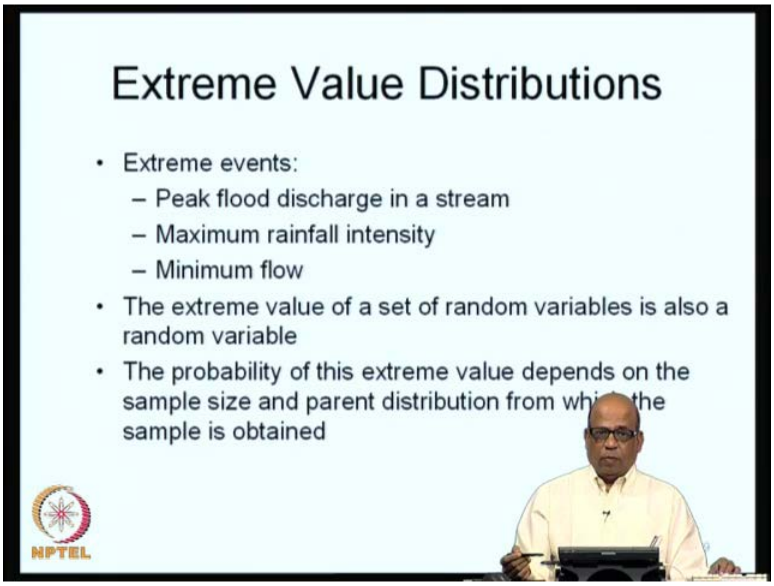
Now, this is how we use the pdf of the gamma distribution there are also tables available then in which case you do not have to go through this integration over and over again. So, the tables for gamma distribution typically use this parameters chi square and nu the parameter chi square is given by 2 lambda x and nu is given by 2 into eta. Eta is the parameter of the gamma distribution, which would be estimated from the sample values. So, once eta is known you can get the nu and lambda is again a parameter of the gamma distribution which you would estimate once lambda is known, the chi square is estimated for that particular value of x for which you are interested in getting the probability.

For example, in this case you are interested in getting the probability associated with the random variable taking on a value greater than or equal to 1 centimeter. So, x you enter with 1 centimeter there. So, this x is associated with the event that you are talking about.

That is x greater than equal to x, x less than or equal to x, x lying between x and x 2 etcetera. So, once you know, once you specify this particular value of x, you get the chi square and you would have known the nu. So, with these 2 values you enter the tables and in general the tables give 1 minus f of x. You can refer to the classic text book by C T Haan, statistical methods in hydrology which I have given the reference for in the first lecture, now such standard text books provide tables for gamma distribution.

However, in the absence of gamma distribution you can use actual integration, because these are all integrals numerically and you can use numerical or direct analytical methods to integrate these depending on the type of functions that you get here, and obtain the associated probabilities. Now, we will move on to extreme value distributions, which are extremely important in hydrology. Often we will be interested not in the normal or smoothen processes; for example, monthly stream flows, annual rainfall, seasonal or monthly, seasonal ground water levels, etcetera which are all smoothened over a period of time and you are talking about normal processes in some normal sense.

(Refer Slide Time: 32:27)



But we would be interested in the maximum rainfall during a certain period, the peak flow in the stream occurring in a month or occurring in a day and so on or the minimum flow that occurs in a particular stream in a year, what is the lowest flow that is expected from this particular steam in a particular year. The lowest flows will have implications directly on the drought situation hydrologic droughts, but more importantly on the water quality situations as you can expect the dilution capacity of the river comes down with the decrease in the flow and therefore, we would be interested in what is the expected minimum flow in the stream.

Similarly the maximum rainfall intensity, as I just mentioned for exponential distribution interested in rainfall intensities in an urban what kind rainfall intensity we can expect let say city of Bangalore, city of Mumbai and so on, and we would be interested in the maximum rainfall intensity because we would be using these high intensity rainfalls for our hydrologic designs. Similarly, peak flood discharge in a stream will have implications on designs of bridges, bridge piers and then on the reservoir capacities how much sluice openings how to be given, sluice gates how to be given, what is the spillway capacity that has to be provided and so on. So, these things will dictate most of the hydrologic design that we provide.

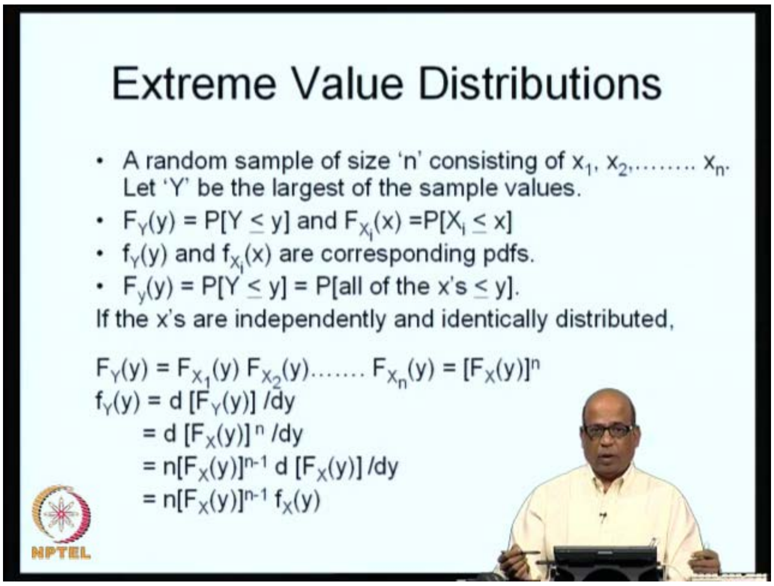
So, in hydrology, we have interest in extreme events. The extreme value of a set of random variables is also a random variable, let say you are talking about the maximum flow that has occurs in a stream over a period of one year. What we would have done, we would have collected the stream flows over the last 50 years occurring during that particular month, let say June period. June period we have 50 values. From these 50 values, we would pick up the maximum value and say that this maximum that has occurred over the last 50 years in the month of June. So, the flow during the month of June is a random variable and we may have obtained a sample of 50 values x 1, x 2, x 3, etcetera x 50. Now all of these become random variables and we are picking up the maximum value out of this sample x 1, x 2, x 3, etcetera x 50.

If you change the sample, then the maximum value will also be changing. So, the maximum value that we are talking about depends on the sample. In fact, both on the sample size as well as on the distribution from which the sample is obtained. So, the probability of extreme value depends on the sample size and parent distribution from which the sample itself is obtained. So, these two are important now, that is first of all the extreme value itself is a random variable; and therefore, we need to start talking about the probabilities of extreme values taking on certain values and so on. Certain specified values, and also that the probability of the extreme value depends on the sample itself.

So, the sample size as well as the parent distribution from which the sample is obtained for example, the sample may have been obtained from a normal distribution or an exponential distribution and so on.

So, the sample size as well as the distribution of the sample from which the sample is drawn both will decide the probability of the extreme values. Often, however, we may not know the parent distribution from which the sample has been obtained and those posses the difficulty of talking about the probability distributions of the extreme values themselves. Let us look at this example now where we are talking about a random sample consisting of x 1, x 2, etcetera x n. If we denote y as the largest of the sample value, that means, you have some n values in the sample and y is the largest value

(Refer Slide Time: 37:49)



Because y is a random variable we start talking about the cdf of y. So, we denote by F Y of y the cdf of y which by definition is probability of y being less than or equal to y, which means what we were saying, the maximum value is less than or equal to a certain value here y and similarly the X i s themselves are random variables therefore, we define the distributions for the X I s and we denote f of X i, which is a cdf which is equal to probability of X i give a being less than or equal to X. And the associated pdf s are given by F Y of y and F X i of X because y is the largest value y is the maximum value in this sample and we are saying Y is less than or equal to this small value y here what does that mean because y is the highest value each of these values X 1, X 2, X 3 etcetera X n.

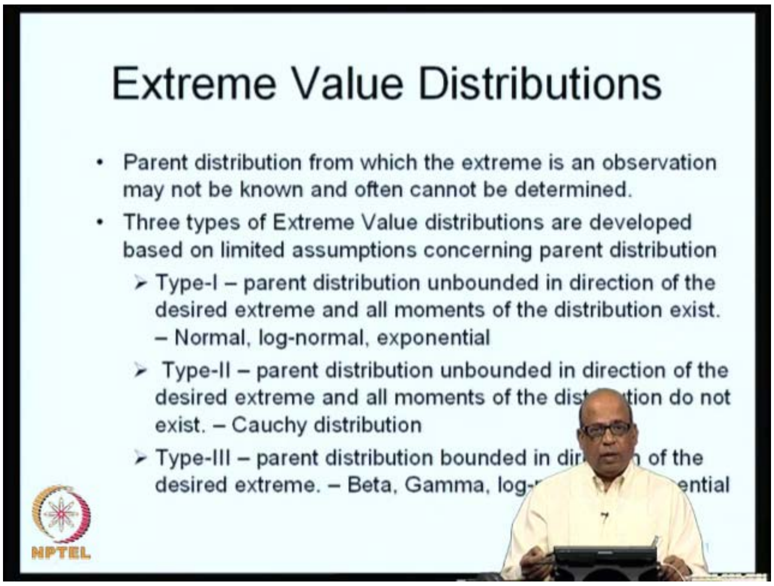
They should be less than the maximum value and therefore, we write this as probability of Y being less than or equal to y. It implies that a probability of all of the X i s being less than or equal to y. So, X 1 is less than or equal to y, X 2 is less than or equal to y, etcetera.

Now, if all the X i s are independent and identically distributed, if they are independent then what will happen, we are talking about probability of X 1 being less than or equal to y and probability of X 2 being less than or equal to y and probability of X 3 being less than or equal to y and so on, because they are independent we can write this as probability of X 1 being less than or equal to y into probability of X 2 being less than or equal to y and so on. So, we can write f y of y as from this f X 1 of y we are using this result now into f X 2 of y etcetera f X n of y this is because we are saying that X 1, X 2, X 3, etcetera are all independent. Further, if they are also identically distributed then what happens, f X 1 of y is the same as f X 2 of y is the same as etcetera f X n of y .

Therefore, we write this as equal to f X of y to the power n from this cdf, we define the cdf to be f X of y to the power n that means, given the cdf of the original population we are able to get the cdf of the maximum value y, from the cdf, you can get the pdf by simply differentiating the cdf. So, f Y of y you differentiate with respect to y you get the pdf. So, d of f Y of y given by dy. So, with respect to y you are differentiating that is d of f x of y to the power n from here with respect to y that becomes n f x y to the power n minus one into d f x y by d y. What is this now? d f X y by d y is nothing, but f X y that is a pdf of X. So, this can be written as n f X y to the power n minus one into f x y this again shows that the pdf of y which is the pdf of the maximum values, is dependent on the sample size n as well as the original distribution f X from which the sample has been drawn.

This is an important conclusion that the extreme values. What we did for the maximum values can also done for the minimum values and it can be seen that the extreme values both maximum as well as minimum the distributions of the extreme values will depend on the sample size as well as on the parent distribution from which the sample has been drawn. Now, as I said the parent distributions from which the extreme is an observation may not be known and often cannot be determined. For example, you may have extremely high intensities of rainfall. So, you may have defined the threshold of the high intensity rainfall to be quite high and then you are looking at extreme high intensities of rainfall. So, you would have collected such rainfalls over let say last 50 years or some such thing.

(Refer Slide Time: 42:39)



But you do not know from which population or you do not know the distribution from which these extreme values have been drawn, but you have the extreme values all right. In such situations how do we obtain the distributions of the extreme values? What we then do is, we place certain limited assumptions concerning the parent distribution and then define several types of extreme value distributions. For example, we define three different types of extreme value distributions. The type-I distribution, we assume that a parent distribution is unbounded in direction of the desired extreme and all moments of the distribution exists which are these type of distributions. Remember here we are talking about the parent distribution that means, the distributions from which the sample has been drawn.

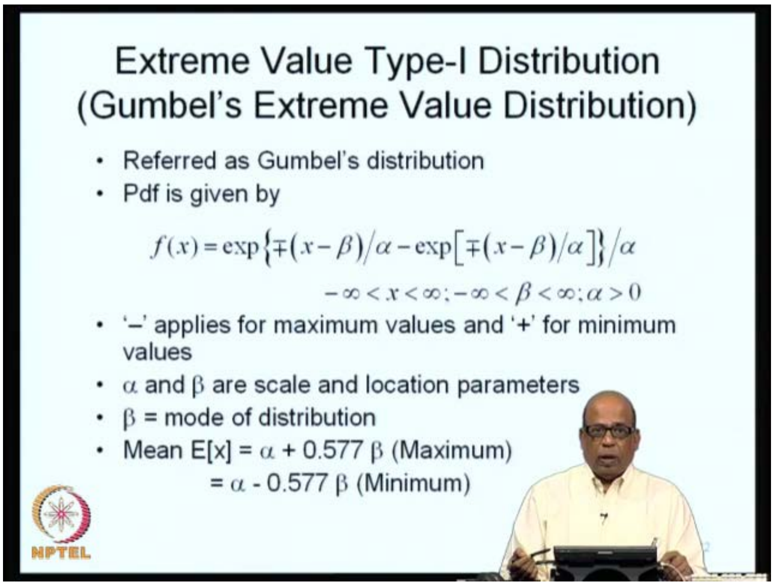
The parent distribution is unbounded in the direction of the desired extreme, let us say we are talking about the maximum value to the right of the distribution. So, it should be unbounded in the right direction. As you can see normal distribution is unbounded in both the directions, log normal is unbounded in the right direction and exponential distribution is unbounded on the right direction. So, these are some of the examples now the normal distribution is also unbounded on the left side, where as log normal as well as exponential are bounded on the left side. So, normal distribution can also be used for getting the minimum values using the type-I distribution.

The type-II distribution is derived based on the parent distribution being unbounded in the direction of the desired extreme all the moments of the distribution do not exist. For example, the Cauchy distribution where it may be unbounded in the direction of the desired extreme, but all the moments do not exist. In fact, type-II distributions have not been very popular in hydrologic applications. So, we do not focus much on the type–II distributions. The type-I and type-II, III are extensively used for the peak flows, the high intensity rainfalls, as well as the low flows and the minimum rainfall and so on. For both floods and as well as the droughts the type–I and type-III distribution have been commonly used in hydrology. So, we focus mostly on type-I and type-III distribution.

What does the type-III distribution indicate? The type-III distribution makes assumption on the parent distribution bounded in the direction of the desired extreme. So, type-I was unbounded in the direction of the desired extreme where as type-III is bounded in the direction of the desired extreme. For example, gamma distribution log normal distribution they are both bounded on the lower side because both are defined for x greater than 0.

So, they are both bounded by x is equal to 0 on the lower side similarly, exponential distribution and the beta distribution which we did not cover in this course beta distributions is bounded on both sides both on the minimum side as well as on the maximum side. So, type-III distribution is generally used for minimum values, but can also be used for maximum values in certain situations where the population can be assumed to be down from a beta distribution. So, we will be dealing with as I said mostly type-I and type -III in hydrologic applications. So, in this course we will discuss only the type-I and the type-III extreme value distributions.

(Refer Slide Time: 47:04)



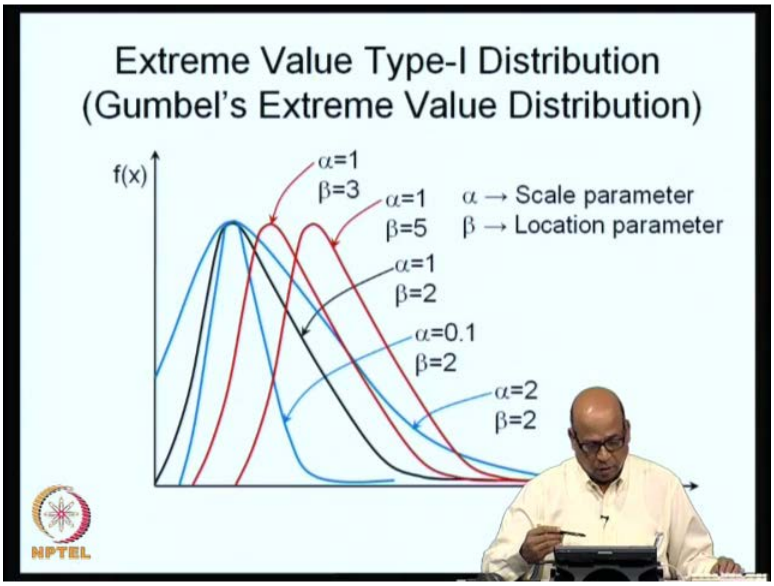
So, the extreme value type-I distribution, this is generally refer to as the Gumbel's extreme value distribution. This is extremely popular in the flood studies. So, the pdf of the Gumbel's extreme value distribution is given as f of x is equal to exponential. This also called incidentally as the double exponential distribution, because there are double exponentials here exponential in large brackets here, minus or plus x minus beta over alpha minus exponential minus or plus x minus beta or alpha this bracket is closed, the whole divided by alpha. As you can see you have an exponential and rise to the power another exponential of the same term this has two parameters alpha and beta.

So, this alpha and beta define this particular pdf and it is defined for x between minus infinity to plus infinity beta between minus infinity to plus infinity and alpha positive values. Alpha should take only positive values. When we are dealing with maximum values we use a negative sign here in both the exponential and when are dealing with the minimum values we use the positive values in the exponentials.

So, alpha and beta are scale and location parameter that means, beta locates where the distribution lies and alpha locates the scale of the distribution itself and the mean of the expected value of X there should be capital X here, the expected value is given by alpha plus X. So, the mean expected value of X is given by alpha plus 0.577 beta. This is for the maximum values and is equal to alpha minus 0.577 beta for the minimum values.

From this you can estimate the parameters alpha and beta this is the expected value and we also have beta given by the mode of the distribution itself.

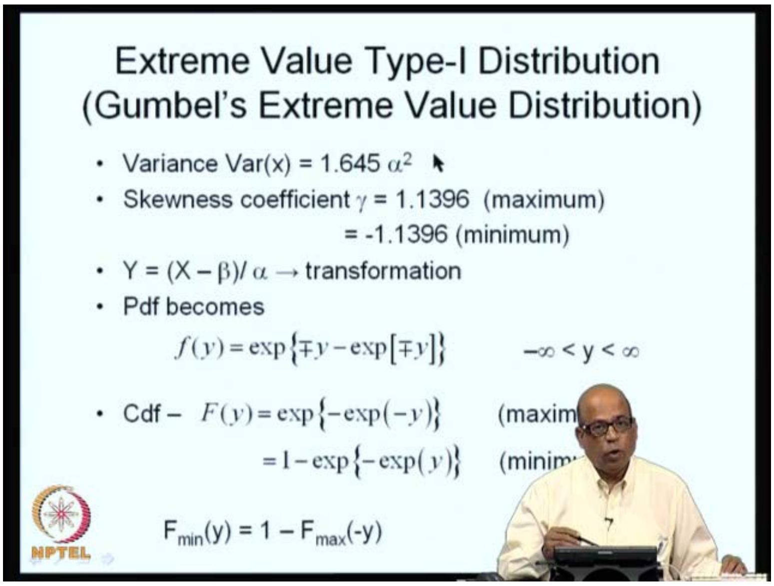
(Refer Slide Time: 49:55)



Now, as I mentioned your alpha is the scale parameter and beta is the location parameter. Let us see how with respect to different betas, the shape, the location changes. You look at this beta is equal to 3 and beta is equal to 5 for the same alpha the distribution has shifted. So, the location has shifted from one point to another point that beta is equal to 2 is at this location for the same alpha is equal to 1. For the same alpha as you change beta the location of the distribution changes and therefore, it is called as a location parameter. For the same beta, as you change alpha for example, alpha is equal to 0.1 here and alpha is equal to 1 here, alpha is equal to 2 here the scale of the distribution changes. As you can see from the blue curves here the scale of the distribution changes and therefore, alpha is the scale parameter and the beta is the location parameter of the Gumbel's extreme value distribution.

As I was mentioning the expected value of x is given by alpha plus 0.577 beta for the maximum values. The variance is given by 1.645 alpha squared for both the maximum as well as the minimum values. By that I mean when you want to use the distribution for obtaining the maximum values or the distribution of the maximum values and the distribution of the minimum values. The skewness coefficient is given by 1.1396 when we are talking about the maximum values and the same with a negative sign for the minimum values.

(Refer Slide Time: 51:23)

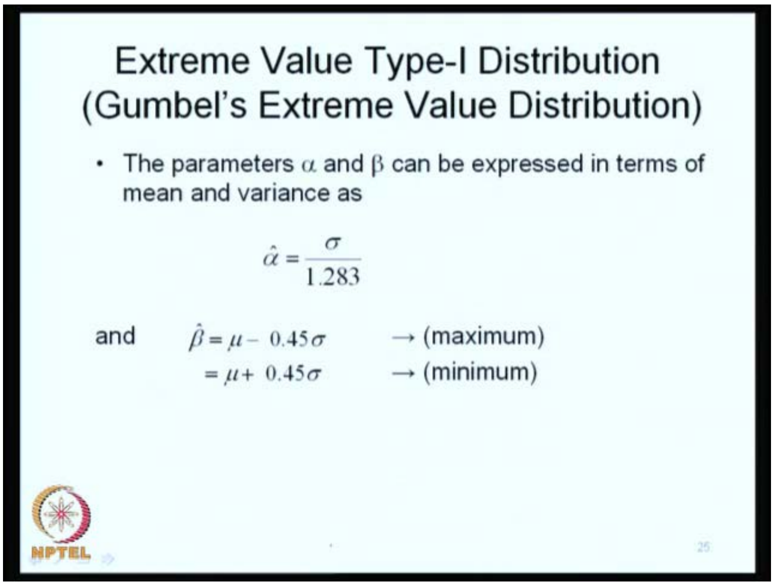


There is transformation that we consider y is equal to x minus beta over alpha now, this transformation, when we use in the original pdf, let us say your pdf is like this. So, you are looking at the transformation x minus beta over alpha. So, x minus beta over alpha this is a transformation we are using when you use that transformation the pdf becomes f of y is equal to now we are talking about the distribution of y f of y equal to exponential minus or plus y minus exponential minus plus y.

This is a useful result and we use this kind of transformation talking about probabilities of extreme events. So, the cdf can be recognized f of y is equal to exponential minus exponential minus y in fact, this is the easier way to renumber the Gumbel's distribution we write here f of y here for example, this one in a easier form we write this as e to the power minus e to power minus y. So, that is why it is called as double exponential distribution. So, this is you really known as double exponential distribution.

For the maximum we use this form, e to the power minus e to power minus y and for the minimum values we use one minus of that e to the power minus e to power minus y which comes from the pdf, f of y defined for maximum or minimum; and therefore, f of y the cdf of y when we are talking about minimum values is equal to 1 minus the cdf of y corresponding to the maximum values.

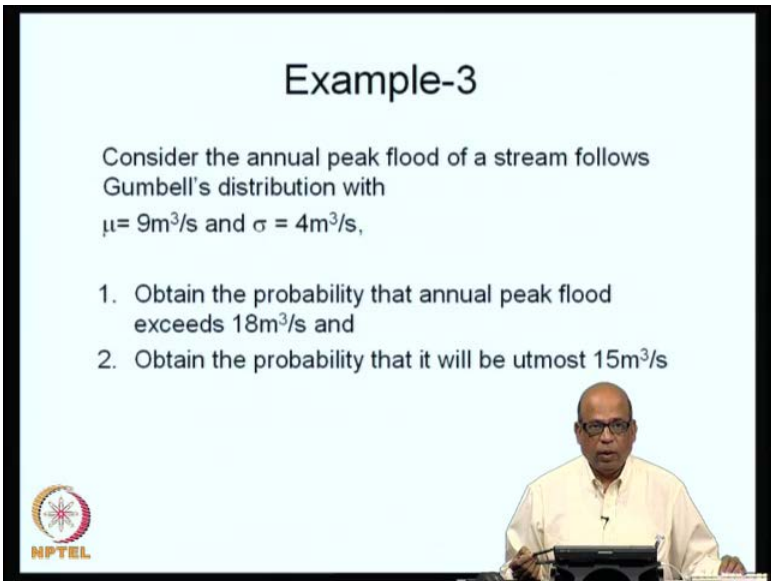
(Refer Slide Time: 54:02)



The parameters alpha and beta can also be expressed in terms of the mean and variance now this is available in standard hydrology text books. It may not be available in standard mathematics books we estimate alpha cap the alpha is denoted with a cap there. So, alpha is estimated by alpha cap is equal to sigma over 1.83 where sigma is the standard deviation of the random variable.

Beta cap is given by mu minus 0.45 sigma for maximum values and plus 0.45 sigma for minimum value. In fact, these expressions are very popular hydrologic situations and especially when we are dealing with floods, we typically use these two parameters or estimates that alpha cap given by sigma by 1.283 and beat cap given by mu minus 0.45 sigma because we will be dealing with the peak flows or the maximum flows. So, we will consider one example now, annual peak flood of a stream follows Gumbel's distribution with mu is equal to 9 meter cube per second and sigma is equal to 4 meter cube per second.

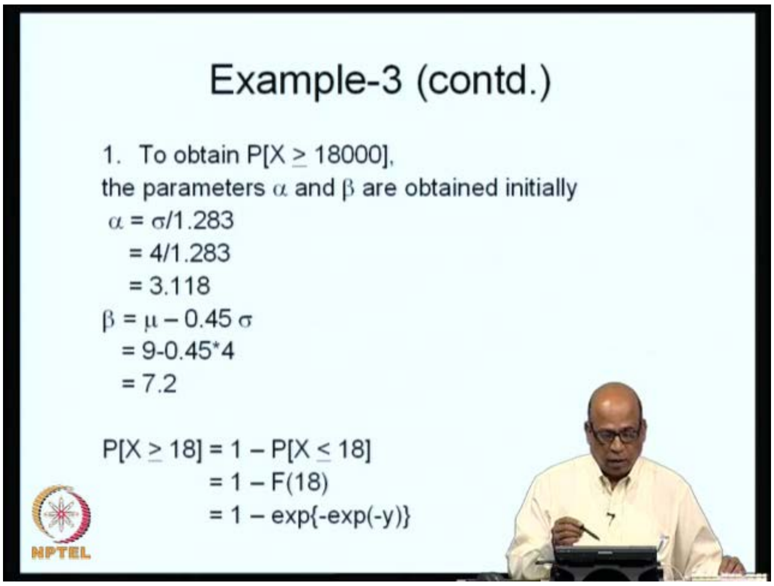
(Refer Slide Time: 55:21)



We want to get the probability that the peak flood exceeds 18 meter cube per second and the probability that it will be restricted to 15 meter cube per second. As you can see these kinds of problems will have implications on hydrologic designs, you want to test the adequacy of an existing design. Then you may want to check the probability that it is restricted to a certain quantity, because your capacity is only limited to so much or you may have designed it for the for a particular discharge 18 meter cube per second and you want to estimate the risk that you are taking what is the probability that it exceeds that particular value for which you have designed. So, these are the kinds of problems that will be interested in while using the extreme value distributions

Now, let us say you want to obtain the probability that it is more than 18 this should be 18 here probability that X is greater than or equal to 18. First you estimate the parameters by using those expressions that I just gave. So, because mean and sigma are given you get alpha is equal to sigma by 1.283 and alpha turns out to be 3.118, similarly beta is equal to because you are talking about maximum values beta is equal to mu minus 0.45 sigma. So, that comes out to be 7.2.

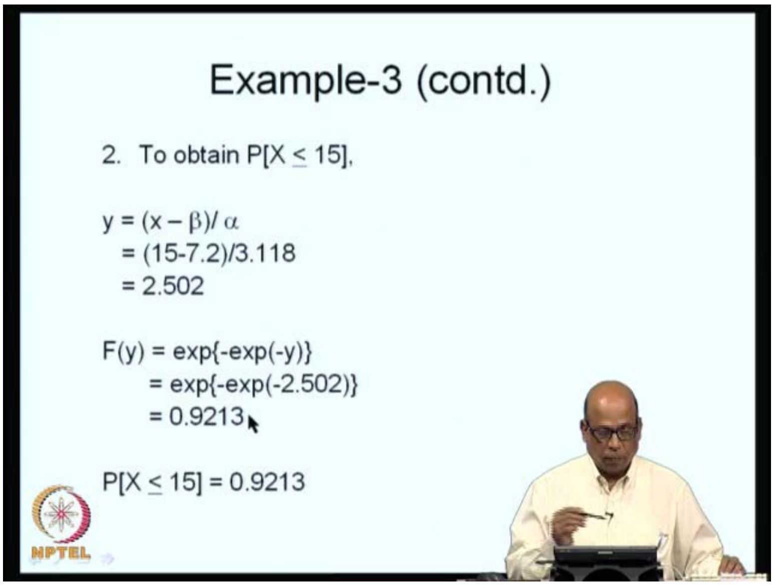
(Refer Slide Time: 56:29)



So, when you are talking about probability of X being greater than or equal to 18 simply get one minus f of 18 which is one minus exponential minus y, where y is the transformation X minus beta or alpha. So, this y here is X minus beta or alpha. So, obtain the y which is 3.646, this is similar to what we did in dealing with the normal distribution. Whenever we are talking about probabilities associated with associated with events defined on a random variable X following normal distribution what did we do we transform that x into z x minus mu or sigma.

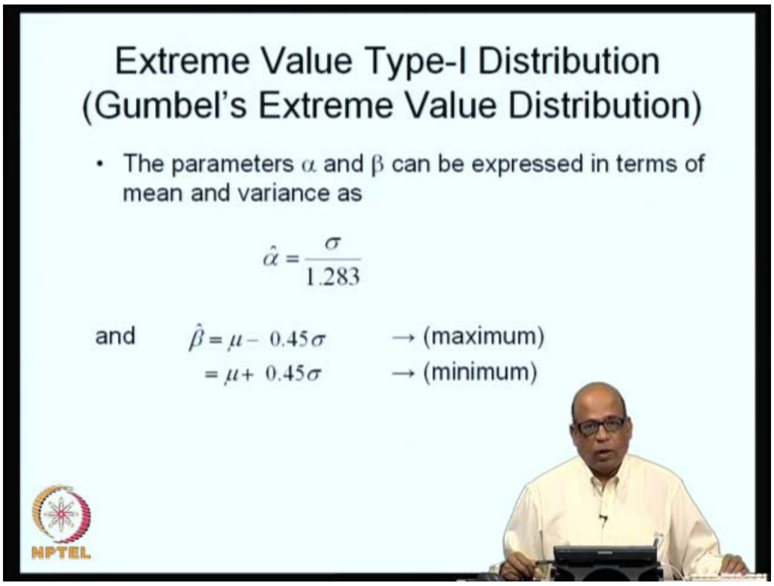
Similar to that when we are talking about probabilities associated with x which follows the extreme value type-I distribution, we use the transformation y is equal to x minus beta or alpha and then use the associated p cdf of y which is a double exponential function and therefore, we get probability of x being greater than or equal to 18 as 0.0308. Similarly, we get the probability of X being less than or equal to 15 we you go through the same steps. First transform the variable into y and then get probability of X being less than or equal to 15 as 0.9213.

(Refer Slide Time: 58:31)



So, this is how we use the extreme value type-I distribution. I again repeat the extreme value type-I distribution is extensively used for flood studies. Typically we want to estimate the probabilities of floods of a given magnitude exceeding particular location.

(Refer Slide Time: 59.22)



So, we fit the extreme value type-I distribution to all the flood discharges, and then estimate the parameters and then start talking about the probabilities associated with various flood magnitudes being exceeded.

So, in today's class what we did is that, we cover the exponential distribution and the gamma distribution recall that the gamma distribution is a family of distributions and then we went on to define the extreme value distributions, in extreme values depending on the parent distribution from which the samples are drawn. You may have type-I distribution, type-II distribution and type-III distributions which are defined based on certain assumptions placed on the parent distribution, and we also examined the type one distribution in detail we define the pdf of the type one distribution and also solved a numerical example associated with the type one distribution. Thank you very much for your attention.