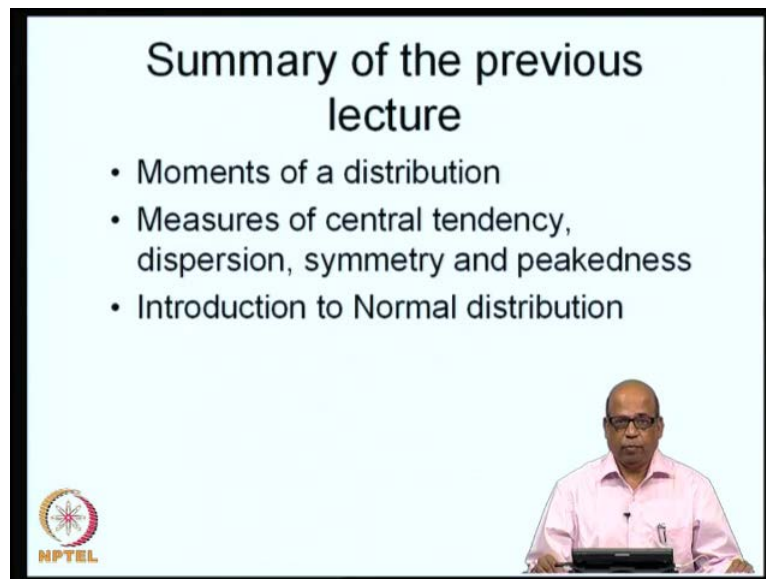**Stochastic Hydrology**

Prof. P. P. Mujumdar

Department of Civil Engineering

Indian Institute of Science, Bangalore

Lecture No. # 05

Normal Distribution

(Refer Slide Time: 00:25)

Good morning and welcome to this the fifth lecture in the course stochastic hydrology. If you recall in the last lecture, we introduced the moments of a distribution. First we took the moments about the origin, and then we started taking moments about the expected value or the mean value. Then we introduce measures of central tendency, specifically the mean mode and the median, and then measure - measures of dispersion the standard deviation, and the variance, and the coefficient of variation c v, measures of symmetry where we introduced the coefficient of skewness, and the measures of peakedness, the kurtosis. Then we also went on to discuss the normal distribution, I just introduced the normal distribution.

So, today we will progress further on normal distribution. What we will do is to make sure that you understand one distribution correctly, we will solve several numerical

examples related with normal distribution before going on to other distributions such as log normal, gamma distribution, etcetera.
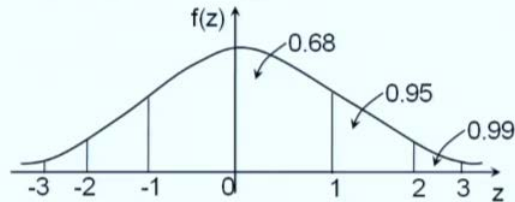
(Refer Slide Time: 01:41)

So, in the last class we mentioned that if x has normal distribution, a linear function y is equal to a plus b X also has a normal distribution, and its parameters are given by a plus b mu that is Y will have the mean of a plus b mu, and it will have a variance of b square sigma square. We use this result to look at so called standard normal distribution, Z is equal to x minus mu over sigma as you can see this is a linear transformation of on X, and then Z will therefore have the normal distribution, and we can see that Z has a normal distribution with mean 0 and unit variance.

This result becomes extremely handy in dealing with normal distribution and we use this pdf of z defined as F of z is equal to 1 over root 2 pi e to the power minus z square by 2, because it has a variance of 1 and mean of 0. The f of z turns out to be this expression and for cdf of z which is probability of z being less than equal to a given value of z we get 1 over root 2 pi integral between minus infinity to z e to the power minus z square by 2 d z as mentioned last time this integral is not analytically this expression cannot be analytically integrated by normal means by the usual methods that we have and therefore, we adopt the numerical integration for this.

 (Refer Slide Time: 03:38)

So, the f (z) is referred to as the standard normal density function this has a 0 mean and the distribution is symmetrically distributed around z is equal to 0 and we also saw that the plus 1 standard deviation plus 1 to minus 1 standard deviation contains about 68 percent of this area then 2 standard deviations that is minus 2 to plus 2 contained about 95 percent of the area and plus 3 to minus 3 contain about 99 percent of the area, which means that about 99 percent of the area is contained within a deviation of plus or minus 3 sigma because sigma is one here the values of capital f of z which is a cdf of z are tabulated. So, we use a numerical integration and then tabulate these f of z, so that we can use these tabulated values of our talking about the probabilities associated with any value of z.
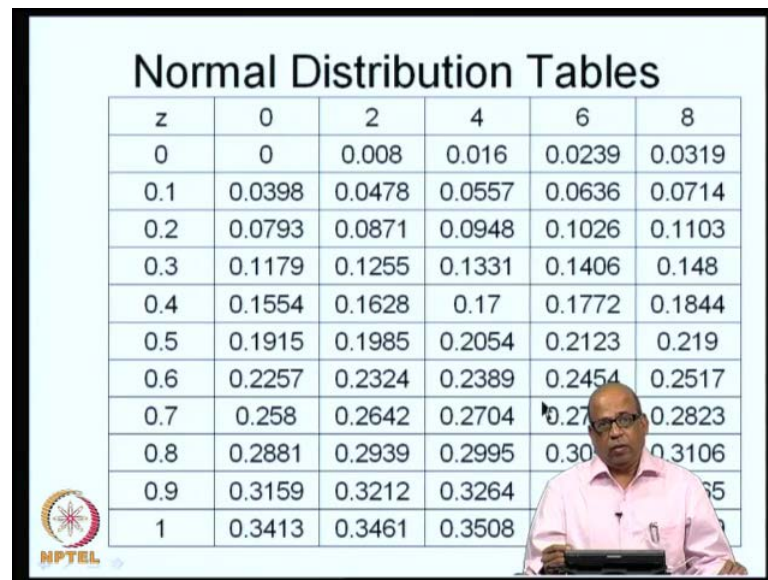
(Refer Slide Time: 04:47)
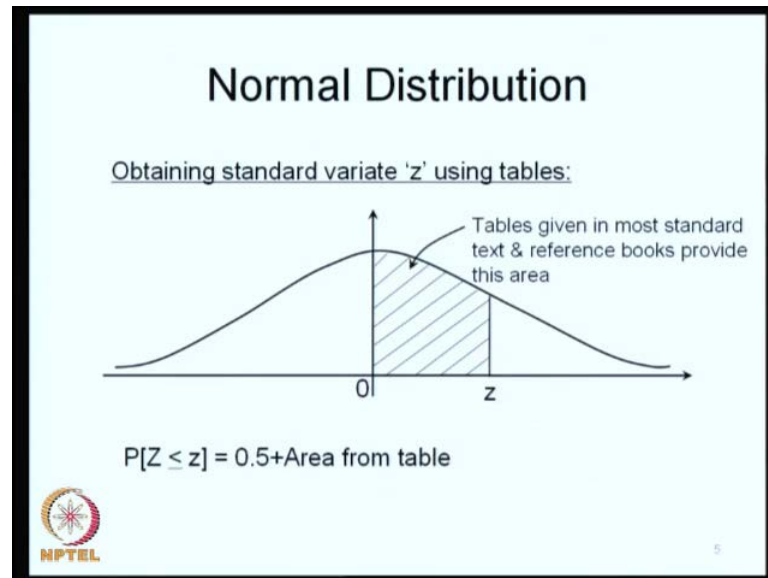
So, most of the text books and the tables that are available in the reference books provide this area that is to the right of 0 so, for all positive values of z they provide this area. So, what does this mean let us say we get this area from the table for a given value of z you need to add 0.5 to that so, that you get the area between minus infinity to z. So, whenever you are using normal distribution tables please ensure that you are looking at this particular area before following the problems that we are discussing in this lecture, if some of the books provide the total area z is equal to that is the area between minus infinity to plus z they include this 0.5 area also in which case you do not have to add 0.5.

What indicates here is probability of z being less than equal to z. So, which means you are looking at this total area and this total area includes 0.5 which is area up to z is equal to 0 and the area as obtained from your standard tables. So, in this course we will adopt this particular method where the areas are provided for positive values of z.

(Refer Slide Time: 06:17)

Normal Distribution Tables

z	0	2	4	6	8
0	0	0.008	0.016	0.0239	0.0319
0.1	0.0398	0.0478	0.0557	0.0636	0.0714
0.2	0.0793	0.0871	0.0948	0.1026	0.1103
0.3	0.1179	0.1255	0.1331	0.1406	0.148
0.4	0.1554	0.1628	0.17	0.1772	0.1844
0.5	0.1915	0.1985	0.2054	0.2123	0.219
0.6	0.2257	0.2324	0.2389	0.2454	0.2517
0.7	0.258	0.2642	0.2704	0.27	0.2823
0.8	0.2881	0.2939	0.2995	0.30	0.3106
0.9	0.3159	0.3212	0.3264		5
1	0.3413	0.3461	0.3508		

So, we will see an example of how the normal distribution tables look in fact, you know these normal distribution tables can be readily obtained from standard software like micro soft excel and so on. So, you can encourage you to just experiment to these and generate these tables yourselves. So, essentially what it does is for a given value of z it integrates the pdf of z between 0 and infinity 0 and z, and then provides the associated values of capital f of z. So, for example, if we are looking at let us say z is equal to 0.24 let us say you are looking at z is equal to 0.24 what is it that we are looking at? We are looking at an area under the standard normal distribution up to 0.24. So, from the table you get area up to 0.24 as 0.0948. So, to that you add 0.5. So, 0.5 plus this is not visible here. So, 0.5 plus 0.0948 that is what you will get 0.5 plus 0.0948.

Now, that gives you the area up to z is equal to 0.24. So, essentially what you are doing is for positive values of z you will read the area up to that point and then add 0.5 to that to get probability of z being less than equal to z. So, this is it goes on like this for various values of z you are enumerating the integral values here and you get the associated tables. So, as you can see from here in about area of about 0.5 or of up to about 99 percent if you add 0.5 to the left of z is equal to 0 you about 99 percent of area is contained in about z is equal to plus 3 to minus 3 and up to 4 when you go it is almost equal to 1.

(Refer slide Time: 08:46)
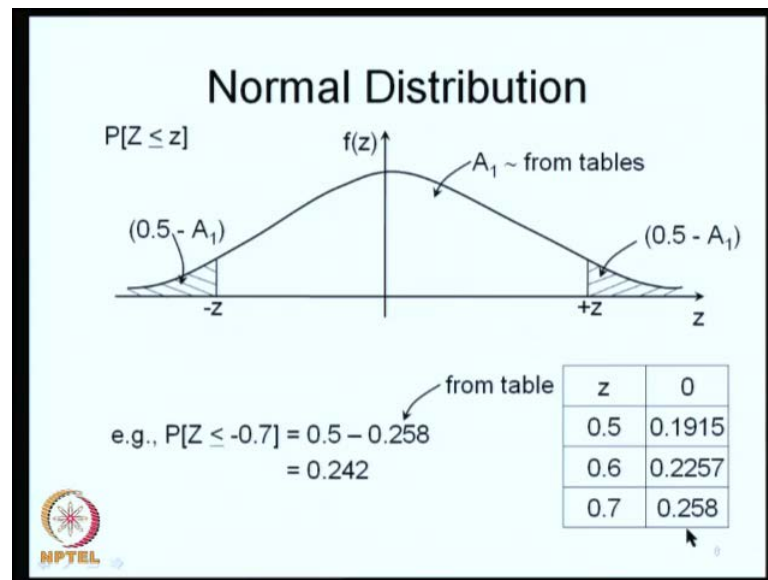
We will see how we obtain for z taking on negative value, let us say you are interested in getting probability of Z being less than equal to minus 0.7 in all the problems dealing with normal distribution we must always remember that the standard normal distribution is symmetrically distributed about z is equal to 0. So, we use that fact and then obtain the areas from the tables and convert that into the probabilities, that we are interested in let us say we are interested in probabilities that z is equal to Z is less than equal to a certain value on the negative side minus z, which means you are looking at this area and the total area up to this point is 0.5. What we do is associated with this you read the area from the table for positive value of z.
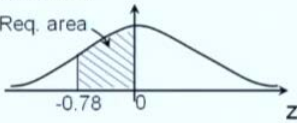
Let say in this case minus 0.7 you read the area corresponding to plus 0.7 what is the area that tables give? The table gives this particular area up to this point then, because the total area right of a curve is 0.5 the area to the right of plus z is 0.5 minus A 1 where A 1 is the area that you just read from the tables and by symmetry this area will be equal to this area. So, 0.5 minus A 1 which is the area to the right of this plus z is the same as 0.5 minus A 1 which is the area to the left of minus z and this defines the probability that z is less than equal to a negative value of z. So, for example, probability of z being less than or equal to minus 0.7. So, what we do is from the table we read for plus 0.7 z is equal to plus 0.7 which is from the table 0.70 it comes out to be 0.258.

And then we take 0.5 minus a one which is 0.5 minus 0.258 which will be 0.242. So, probability of z being less than equal to minus 0.7 is equal to 2, 0.242. So, in all these problems we must first understand what is the area under the standard normal curve? That we are looking at and then use the fact that the standard normal distribution is symmetrical about z is equal to 0. Pick up associated areas from the tables and then convert them into the associated probabilities.

(Refer Slide Time: 11:27)

We will do several examples on normal distribution today. So, that you are well versed with usage of the tables for standard normal distribution, let us say we will first start with getting the area between z is equal to minus 0.78 and z is equal to 0. So, we are interested in probability of z lying between minus 0.78 and 0. So, this is the area that we are interested in if you use numerical integration what we would have done we will integrate the cdf of z which is one over root 2 pi e to the power minus z square by 2 we will integrate between minus 0.78 to 0.

We would have integrated the pdf of z that the probability density function which is e to the power of minus z square by 2 divided by root 2 pi that is a pdf we will integrate between the area between z is equal to minus 0.78 and 0 to obtain this area. So, if we do this numerically we get an area of 0.2823. So, we write this integration as minus 0.78 to 0 which by symmetry we write as 0 to 0.78 and we integrate. It numerically using some
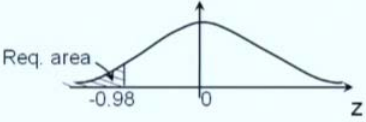
standard software like mat lab or something and then get this probability. So, the area of this turns out to be 0.2823.

Now, from the tables, because z is negative now what we do is corresponding to z is equal to plus 0.78 which comes somewhere here plus 0.78 you read the area which comes out to be 0.2823. So, this is the area between 0 and plus 0.78 which is 0.2823 which by symmetry is also the area between minus 0.78 and 0. So, that is how we obtain the area between area for z is equal to area between z is equal to minus 0.78 and 0.

(Refer Slide Time: 13:37)

Now let us look at the area under the standard normal curve for z being less than equal to minus 0.98 that is we are looking at the area to the left of this as said this is also same as area to the right of this where this particular, is 0.98 that is equal to plus 0.98. So, look up for 0.98 you get 0.3365. So, you are getting the area up to this point. So, let us say this is plus 0.98 by symmetry this is plus 0.98 and the area that we are getting here is 0.3365 which is this area this is 0.3365.

And we are interested in this area which is also equal to this particular area. So, this area will be equal to this area this area is what 0.5 minus 0.3365 which is 0.3365 that is how we obtain the area that is required here in this particular case z is less than equal to minus 0.98.
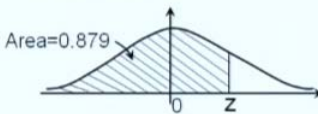
(Refer Slide Time: 14:44)

Example-3

Obtain 'z' such that $P[Z \leq z] = 0.879$

Since the probability of $P[Z \leq z]$ is greater than 0.5, 'z' must be +ve

Area=0.879

z	6	7	8
1.1	0.377	0.379	0.381
1.2	0.3962	0.398	0.3997

area between 0 to z = 0.879 − 0.5
= 0.379

From the table, for the area of 0.379, corresponding z = 1.17

Again now we will do a different type of example where we have specified the probability of z being less than equal to z has been specified as 0.879 we are interested in getting the associated value of z. So, we are asking the question what is that value of the z for which probability of z being less than or equal to z is equal to 0.879 that is the question that we are asking now, you must remember because this probability is greater than 0.5 we are looking at the positive side of z is equal; that means, to the right of z is equal to 0 is what we are looking at. So, this total area has been specified to be 0.879. So, from the tables what is it that we have to look at we look at this particular area which is 0.379, which is 0.879 minus 0.5.

So, you go to the table and look at the value that is closest to 0.379 you can also do numerical interpolation and if you want exact values. So, in this particular case it turns out to be 1.17. So, for the area of 0.379 we get z value of 1.17 we will now do an example where we are looking at a given sample value let us say you have stream flows at a particular location.
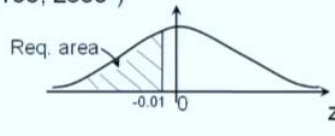
We denote that as a random variable X, and that has a mean of 100 estimated mean of hundred and the variants of 275100 square and we are interested in getting probability of x being less than equal to 75. So, we convert x into z by using z is equal to x minus mu over sigma and therefore, the right hand side which are the specific specified values of x are also converted into the associated specified values of z as z is equal to x minus mu over sigma. So, 75 we take it as 75 minus 100 which is the mean by 20, 5, 100, which is sigma this is sigma square. So, this is sigma. So, that z is equal to minus0.01. So, from the table you get you are looking at minus 0.01. So, you are looking at this area what you do is you get this particular area associated with z is equal to plus 0.01, so 0.01 which is 0.04 and because you are looking at this particular area.

This area becomes point 5 minus 0.004 which is 0.496. So, that is how you get probability of x being less than equal to 75 is also same as z being less than equal to minus 0.01 which is equal to 0.496. We will do another example where specifying mu x this is similar to the earlier example except that we are looking for probability of x being greater than equal to x and we are looking at what is that particular value of x for which probability of x being greater than equal to x is 0.73.

(Refer Slide Time: 17:46)
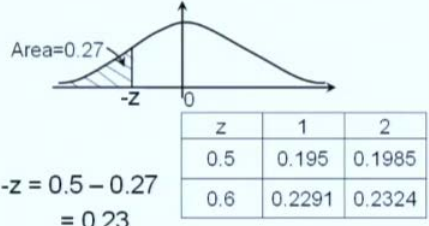
In fact, these kinds of problems come quite often in water resources where we are saying that the flow value that is exceeded with 70 percent of probability 75 percent probability and so, on. So, we are interested in that particular value of x which is exceeded in this particular case 73 percent of time are probability that that particular value is exceeded is equal to 0.73 and from the samples we have estimates for mean and standard deviation that is why in this particular case you have mu x is equal to 650 and sigma x estimated is 200.

So, as you can see probability of x being greater than equal to x converts itself into probability of x being less than equal to x is equal to 1 minus 0.73 which is 0.27. So, you are looking at an area of 0.27 because it is less than 0.5 it has to lie on the negative side of 0 on the left side of 0. So, we are looking at minus z here. So, area between 0 to minus z is equal to area between 0 to plus z here. So, we are looking at an area of 0.23. So, you look up the tables and go to the area where you get closest to 0.23 you can do the numerical interpolation and you get z is equal to minus 0.613.

Once you get z you are actually looking at the value of x here you are not interested in z that value. So, once you get the value of z which is in this case minus 0.613 we use the fact that z is equal to x minus mu over sigma and get the associated value of x in this case it turns out to be 5, 207, now another similar type of example where we are dealing with x which is normally distributed and we have two probabilities given.
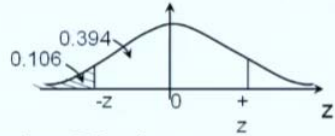
(Refer Slide Time: 19:56)

Example-6

A rv 'X' is normally distributed with following probabilities:

$P[X \leq 50] = 0.106$ and $P[X \leq 250] = 0.894$

Obtain μ and σ of 'X'

$P[X \leq 50] = 0.106$
$P[Z \leq z] = 0.106$
Since the probability is less than 0.5, z is –ve.
From tables, for area of 0.394, -z=1.25

$z = \dfrac{x - \mu}{\sigma}$; $\dfrac{50 - \mu}{\sigma} = -1.25$

$\mu = 50 + 1.25\sigma$

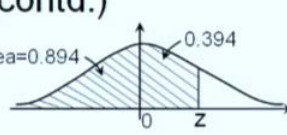| z	4	5	6
1.1	0.3729	0.3749	0.377
1.2	0.3925	0.3944	0.3962
1.3	0.4099	0.4115	0.4131

Probability of X being less than equal to x is given and probability of X being less than equal to 250 is given as 0.894 we are interested in getting the standard deviation and the mean of x. So, what are we given we are given X being less than equal to 50 equal to 0.106 which is this. So, we first convert this into probability of Z being less than equal to z is equal to 0.106. Since the probability is less than 0.5 the z has to be the z that we are talking about has to be negative value of z. So, from the tables we get for an area of 0.394 that is 0.5 minus 0.106 this area is 0.394. So, this is 0.394 here corresponding to 0.394 here you get a z value of minus 125.

So, 1.25 corresponds to this area and we are interested in this particular area being 0.394. So, z value corresponds to 1.25 and that we convert it as minus 1.25 here, because we are interested in this particular area. So, z is equal to minus 1.25. So, we write one equation mu is equal to 50 plus 1.25 sigma from this expression.

(Refer Slide Time: 21:44)

Similarly, we use the other one other condition that is probability of x being less than equal to 250 is equal to 0.894 this is given. So, what are we given we are given that the area up to this point including area to the left of that is equal to 0 is given as 0.894. So, area to the right of 0 will be 0.394 and therefore, we go back to the tables and look at varies at area which is 0.394. So, again we get 1.25. So, z becomes 1.25 remember because we are talking about area being greater than 0.5 we are looking at the right side of z is equal to 0. So, z becomes equal to 1.25. So, from the earlier expression we write we use mu is equal to 50 plus 1.25 sigma and then obtain sigma as a t and mu as 150 one more example we will do where we are considering the annual rainfall in a particular basin this is normally distributed with a mean of 1000 mm and a standard deviation of 400 millimeters.

Now, we have an expression which relates the runoff R with the precipitation P for the rainfall P as given by R is equal to 0.5 P minus 150. Now P is normally distributed and we are interested in getting the mean and standard deviation of the annual runoff, because this is a linear function and P is normally distributed R is also normally distributed and we can obtain the mean and standard deviation of R and from there we can get a the probabilities associated with the annual runoff exceeding any given value for example, in this case 600 millimeters.

So, first we will start with this R is equal to minus 150 plus 0.5 P this is a linear function of P. Since P is following a normal distribution with mean as 1000 and variance as 400 square R follows normal distribution with mean as 350 and standard deviation as 200. So, we obtain mean and standard deviation by using simply the fact that a linear function of a normal random variable linear function depend on a normal random variable also follows normal distribution with mean given by a plus b mu and variance given by b square sigma square. So, we obtain the mean and standard deviation as 350 and 200 millimeters in this particular case next we are interested in probability of R being greater than equal to 600. Once we know that R follows normal distribution for example, in the previous case what did we do we said that R follows normal distribution with the parameters mean as 350 and standard deviation as 200.

(Refer Slide Time: 25:06)
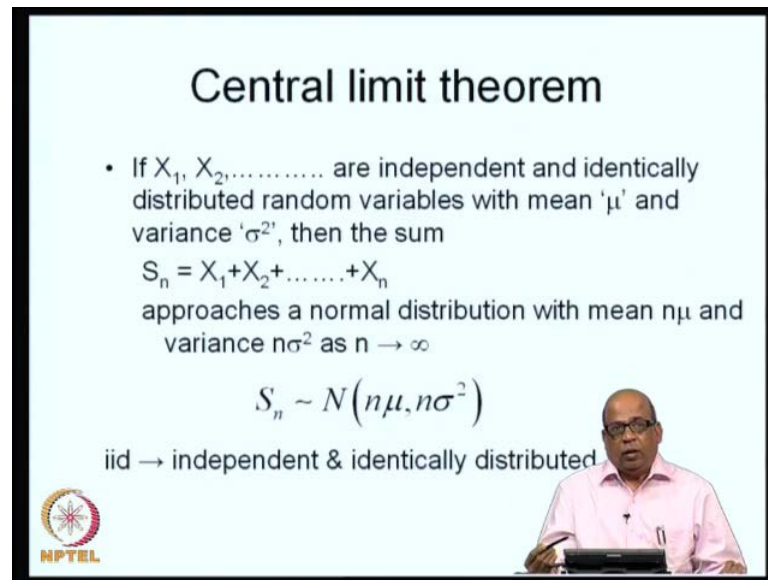
We can start talking about probabilities associated with the variable R which is runoff in this case. So, we obtain probability of R being greater than equal to 0 greater than equal to 600 as 1 minus probability of R being less than equal to 600 and so, on. We use the same procedure and obtain the probability of R the rainfall the runoff at a particular location in the basin being greater than equal to 600 as 0.6056. So, in this example what did we demonstrate we demonstrated the use of linear functions defined on a random variable, which follows normal distribution with known parameters mu and sigma, now

why is normal distribution. So, popular not only in hydrology, but in many other applications many other scientific fields normal distribution is extremely popular in fact, as a first cut analysis you generally use normal distribution when you do not have any other inferences available to you.

(Refer Slide Time: 25:45)

That is mainly because of the central limit theorem now the central limit theorem states if X 1 X 2, etcetera are independent and identically distributed random variables with mean mu and variance sigma square, then the sum defined by S n is equal to X 1 plus X 2 plus X 3 etcetera, up to X n where you have N number of such random variables. Approaches a normal distribution with mean n mu and variance n sigma square as n becomes as N tends to infinity that is we state S n is equal to S n follows a normal distribution which parameters mean as n mu and variance as n sigma square. Now this is an important result and we use this abbreviation iid to indicate that the random variables are independent and identically distributed now look at the implications of this it does not put any restriction on whether X 1, X 2, X 3, etcetera have to follow normal distribution they can follow any distribution as long as they are independent random variables.

And as long as all of them follow same distribution, let us say they are following exponential distribution with the same mean. So, they should all have the same mean and the same standard deviation. So, by iid mean independent and identically distributed as long as you satisfy these conditions the sum of the random variables X 1 plus X 2 plus

etcetera, X n this sum follows normal distribution with mean given by n mu and the variance given by n sigma square. In many situations we can approximate a particular random variable as having being constituted of a sum of several random variables and if we can also make a assumption that they are independent and identically distributed then this result becomes very handy let us say that you have you are looking at a stream flow in a particular month. Let us say stream flow in month June is a we are interested in the distribution of this.

Can we approximate this to be a normal distribution if we consider the stream flow in month June as having been constituted of several random variables, let us say in this particular case 30 random variables, 31 random variables x 1 plus x 2 plus x 3 plus etcetera, where X is are the stream flows in day. So, the monthly stream flow can be looked at as a sum of daily stream flows in that particular month, now if the daily stream flows can be assumed to be independent and identically distributed then we can say that the stream flow in that particular month follows a normal distribution with the mean given by 30 into mu and the variance given by 30 into sigma square where mu and sigma square are the mean and the variance of the individual days of random variables.

(Refer Slide Time: 30:14)

Now, the requirement that these be identically distributed that becomes slightly restrictive, so far most hydrological applications under general some general conditions if X I are all independent with expected value of X I given by mu I and variance of X I

given by sigma square; that means, what we are doing now is we are relaxing the requirement that they are all identically distributed all of them have some distributions X 1 has a distribution X 2 has its own distribution with different parameters X 3 has its own distribution with different parameters and so on. But as long as we can assume X 1, X 2, X 3 etcetera, X n as independent then the sum X 1 plus X 2 plus X 3 etcetera, X n approaches normal distribution with expected value of s N given my sum of mu over all and variance of S n given by sum of all the variances is equal to one to N now, one condition for this generalized central limit theorem is that each of X I that we are considering here has a very limited or very negligible effect on the total distribution of S n itself. So, individually they are not contributing significantly to the distribution of S n, but together they are making sure that it approaches a normal distribution.

Now, this result this generalized condition becomes extremely handy in hydrologic applications where we will be dealing with as just mentioned several random variables which can be taken as sums of individual random variables. Again for examples seasonal rainfall if we are looking at seasonal rainfall or seasonal stream flows at a particular location if we are looking at now, this seasonal stream flow can be taken as having being constituted of daily stream flow, let us say we are talking about stream flow in a monsoon period which has 4 months approximately 120 days. So, X 1 X 2 etcetera, up to X 1 20 there are several individual random variables as long as you can take them as independent random variable then you can approximate the seasonal stream flow as with a normal distribution and once you know the mean and standard deviations of the individual random variables X 1, X 2, X 3, etcetera, up to 120.

(Refer Slide Time: 33:05)

Then you can also obtain the mean and standard deviation of the seasonal flow, now we will see why normal distribution cannot be very generally applied in most of the hydrologic situations although normal distribution is extremely elegant and extremely useful in many applications, you can recall that normal distribution is defined from minus infinity to plus infinity. So, irrespective of how high is you R mean how much to the right of 0 your mean there is always a probability finite probability associate with negative values. So, even if you have the mean to extreme right and that is extremely high value of mean, but there is always a probability associated with a negative value of the particular random variable in most of the hydrologic situations we are dealing with non negative variables.

For example stream flow cannot be negative rainfall cannot be negative cannot be negative and so, on. So, most of the hydrologic variables are non negative unless you are talking about temperature as one of the variables and with the scale you are looking at you may have negative values of temperature or reservoir levels around a particular threshold value that can be negative. So, in only very specific cases specific applications you come across negative variables it is for this reason normal distribution has a limitation in the sense that there is a finite probability associated with negative values when we are dealing with normal distribution and therefore, when you generate values of generate samples of this particular variable using the normal distribution there will be

always negative values that are generated which we will see in subsequent classes of this particular course.

The other property of normal distribution is that it is perfectly symmetrical. So, it is symmetrical about x is equal to mu, but most of the variables that we deal with hydrology for example, rainfall or time between two events two critical events or the flood flows at a particular location etcetera, these are generally skewed distribution these follow generally skewed distributions with gamma as typically being positive in most situations. So, whenever we have a significant skew we cannot use normal distributions. So, both these limitations of normal distribution lead us to use the log normal distribution, which we will introduce now the log normal distribution, if Y is equal to L N X that is Y is equal to log of X log natural of X this follows normal distribution then X is said to follow log normal distribution as simple as that.

That is we take the transformation Y is equal to and if y is equal to l n x follows a normal distribution then Y follows log normal distribution now, the probability density function of the log normal distribution is given by F of X is equal to this is obtained from the fact that Y is equal to l n x follows normal distribution as you can see this is normal distribution for l n x, F of X is equal to 1 over root 2 pi x sigma x e to the power minus l n x minus mu X whole square divided by 2 sigma X square and this is defined for greater than zero and mu x greater than 0 and sigma x greater than 0 . So, all these are positive quantities now this has a property that the skewness coefficient gamma S is given by three C v plus C v cube.

Where C v is the coefficient of variation of x which as you can recall is simply sigma x divided by mu x so, coefficient of variation of x. So, as C v increases the skewness gamma increases for the log normal distribution. So, while the normal distribution was a symmetrical distribution the log normal distribution it has a positive skew as you can see here this has a positive skew.

(Refer Slide Time: 37:48)

Log-Normal Distribution

The parameters of Y= ln X may be estimated from

$$\mu_y = \frac{1}{2}\ln\left[\frac{\bar{x}^2}{1+C_v^2}\right]$$

$$\sigma_y^2 = \ln\left[1+C_v^2\right] \qquad where\, C_v = \frac{S_x}{\bar{x}}$$

Or in most situations because here x bar is in hydrologic applications we are dealing with x bar as positive and S x being standard deviation is always non negative now, the parameters of y is equal to l n x may be estimated this is. So, specific to hydrologic applications where zhou and han have demonstrated that the parameters of y is equal to l n x may be approximated as mu Y is equal to half l n x bar square by 1 plus C v square and sigma y square is equal to log of 1 plus C v square where C v is the coefficient of variation of x, if we are given a sample let us say you have given 50 years of stream flows at a particular location, now this constitutes a sample of the random variable x where the random variable is the stream flow at that particular location.

From the stream flows samples you can estimate x bar and the standard deviation s x and if x follows log normal distribution then y is equal to l n X follows normal distribution and the parameters of y is equal to l n X are given by mu Y is equal to this and sigma y square is equal to this. So, from the sample you would have estimated x bar as well as x and therefore, you know C v and using this you can write mu Y and sigma Y square. Once you specify mu y and sigma Y square the Y is equal to l n X which follows a normal distribution is completely defined and therefore, you can start talking about probabilities on Y, because it follows normal distribution and then its parameters are determined as mu Y and sigma Y square in applications. Another way of doing this is if you are given a sample of on X, let us say stream flows at a particular location as we did just now stream flows at a particular location for the last 50 years monthly values of stream flows are available which means 50 into 12 , 600 values are available.

And you want to approximate this with a log normal distribution then a easy way although slightly irrelevant way of doing this will be that you take log of X corresponding to each of the values for example, you have let us say you have X 1, X 2, X 3, etcetera, observed values X let us say you have X 1, X 2, X 3, and so on; these are the observed values. So, you simply take y is equal to l n X 50 you take log of 50 and then associate y. So, you generate another series y is equal to l n X, if x follows normal distribution log normal distribution then y is equal to l n X follows normal distribution and therefore, you work with the series on and associate the probabilities on X .

So, this is one easy way of doing this. So, if you have a sample on X which follows log normal distribution simply take y is equal to l n of X generate another sample and this sample follows now normal distribution you can its mean and standard deviation by this from this sample then start talking about probabilities on Y which are also related to probabilities on X .

(Refer Slide Time: 41:34)
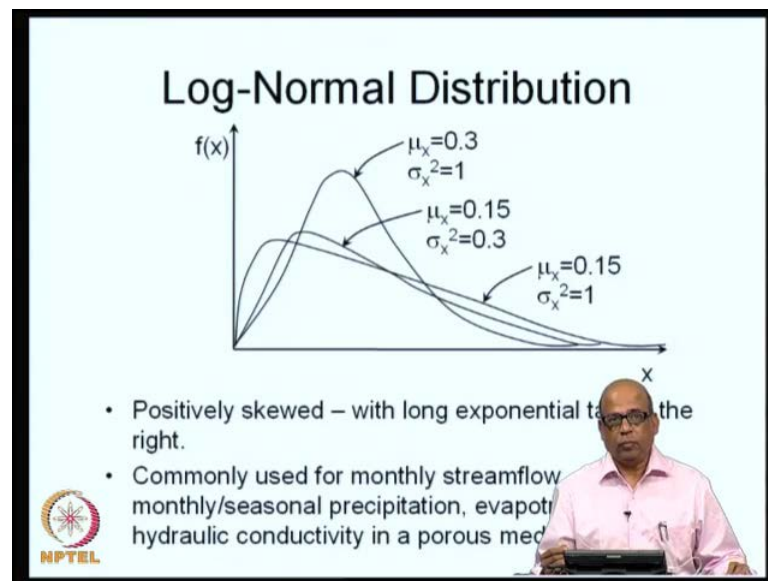
Now, just look at how the log normal distribution appears this is for different mu for example, mu X is equal to point 3 sigma X square is equal to one this is a shape that it takes now as your C v increases; that means, C v is sigma by mu as your C v increases from your previous location previous expression here 3 C v plus C v cube as your C v increases the skewness increases.

So, you just look at these. So, as your C v is increasing for the same mu X of 1 point 5 the C v is increasing between this and this because the S is smaller here and s is larger here. So, the C v is increasing as C v increases the skewness increases; that means, you will have a longer tail to the right. So, log normal distribution is positively skewed with long exponential tail on the right like this and log normal distribution has many applications in hydrology is typically used for monthly stream flows monthly or seasonal precipitation evapotranspiration and hydraulic conductivity in a porous medium and so, on. So, log normal distribution is very popularly used in several hydrologic applications let us consider one example, now let us say you are talking about the annual peak runoff in a river this is modeled by a log normal distribution.

(Refer Slide Time: 43:09)

Example-8

Consider the annual peak runoff in a river - modeled by a lognormal distribution

$\mu = 5.00$ and $\sigma = 0.683$

Obtain the probability that annual runoff exceeds 300m³/s

$P[X > 300] = P[Z > (\ln 300 - 5.00)/0.683]$
$= P[Z > 1.03]$
$= 1 - P[Z \leq 1.03]$
$= 1 - 0.3485$
$= 0.6515$

z	2	3	4
0.9	0.3212	0.3238	0.3264
1	0.3461	0.3485	0.3508
1.1	0.3686	0.3708	0.3729

It has a mean of 500 and a standard deviation of 0.683. So, we want the probability that the annual runoff exceeds 300 meter cube per second. So, we are interested in X being greater than 300. So, we are interested in getting of probability of Z being greater than log 300 minus 5 these values of mean and the standard deviation that we have given are for Y that is y is equal to ln X, and then you talk about because y is equal to l n X follows log normal distribution, you can convert probability of X being greater than 300 as probability of Z being log of 300 minus 5 this is the mean divided by 0.683 which is a standard deviation.

Then we use the normal distribution table as we did in the previous examples on normal distribution table and get the probability of X being greater than 300 as 0.6515, Similarly we will consider now X bar is equal to 135 million cubic meters, we are talking about the stream flow that location then the standard deviation is 23.8 million cubic meters.

(Refer Slide Time: 44:43)

Example-9

Consider
$$\bar{x} = 135 \text{ Mm}^3 , S = 23.8 \text{ Mm}^3 \text{ and } C_v = 0.176$$
If X follows lognormal distribution
Obtain the P[X ≥ 150]

$$\bar{Y} = \frac{1}{2}\ln\left[\frac{\bar{X}^2}{C_v^2 + 1}\right]$$

$$= \frac{1}{2}\ln\left[\frac{135^2}{0.176^2 + 1}\right] = 4.89$$

$$S_y^2 = \ln(C_v^2 + 1) = \ln(0.176^2 + 1) = 0.0305$$
$$S_y = 0.1747$$

And from this we get C v as S by X bar which is 23 by 8 divided by 135 that comes out to be 0.176, now if X follows log normal distribution we are interested in getting probability that x being greater than equal to 150. So, we use the expressions and the fact that y is equal to l n X follows normal distribution we first estimate Y bar and S y square which is the mean of Y and the variance of Y using the expressions that just introduced X bar square plus divided by C v square plus 1 remember both X bar as well as C v deal with relate to the original variable X and we are getting the mean of the transformed variable y is equal to l n X , similarly the variance of the transformed variable Y is equal to l n of X . So, we get the mean as 4.89 and the variance as 0.0305 are the standard deviation as0.1747.

So, once we determine the mean and the standard deviation and given the fact that Y follows normal distribution then we can talk about X being greater than equal to 150. So probability of X being greater than equal to 150 means probability of Y being greater than equal to log of 150, because Y is equal to l n X.

(Refer Slide Time: 46:23)

Example-9 (contd.)

Y = ln X follows log normal distribution

P[X ≥ 150] = P [Y ≥ ln150];

ln150 = 5.011

$$z = \frac{y - \bar{y}}{S_y}$$

$$= \frac{5.011 - 4.89}{0.1747}$$

$$= 0.693$$

P[Y ≥ ln150] = 1− P[Y ≤ ln150]

= 1− P[Z ≤ 0.693]

= 1 − (0.5+0.25583)

= 0.24117

So, we then because Y follows normal distribution we simply use the fact that z is equal to y minus y bar over S y. So, Y is l n 150 which is 5.011 minus 4.89 which is the mean of y as we obtained just now 4.89 and the standard deviation is0.1747. So, using that we get probability of Y being greater and equal to log 150 using the standard normal tables as 0.24117. So, log normal distribution is mostly used for as said monthly stream flows and conductivity evapotranspiration and so on, but there are many situations where we would be looking at time between critical events.
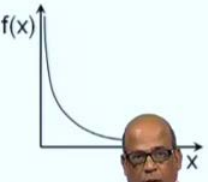
(Refer Slide Time: 47:08)

Exponential Distribution

• The probability density function of the exponential distribution is given by

$$f(x) = \lambda e^{-\lambda x} \qquad x > 0, \lambda > 0$$

• E[X] = 1/λ
• λ = 1/μ
• Var(X) = 1/λ²

$$F(x) = \int_0^x f(x)dx = 1 - \lambda e^{-\lambda x} \qquad x > 0, \lambda > 0$$

Let say that the time that has elapsed between high intensity rainfalls are the time that has elapsed between two critical floods of a given magnitude. So, whenever we are talking about such variables then log normal distribution is not generally suitable we have the exponential distribution which is ideally suited for such purposes. Now, well introduce the exponential distribution the probability density function of the exponential distribution is given by a simply f of x is equal to lambda E to the power minus lambda X defined for X greater than 0 and lambda greater than 0 this is a single parameter distribution, where lambda is the only parameter you can easily verify that the integral between 0 to infinity of lambda E to the power of minus lambda X turns out to be 1 indeed in the first few classes when introduced to the pdf we would have considered a similar example the expected value of this of X is can be shown to be 1 over lambda. So, mu is equal to 1 over lambda or lambda can be estimated as 1 over mu and the variance of X which follows an exponential distribution can be shown to be 1 over lambda square.

The exponential distribution pdf looks like this which is positively skewed there is a long tail to the right and then it approaches asymptotically X is equal to infinity on the X axis asymptotically on the other side y the cdf of F of x which by definition is integral between 0 to 0 and X, f of x d x will turn out to be 1 minus lambda E to the power minus lambda X defined again for x greater than 0 and lambda greater than 0 .So, once you define cdf of x you can talk about associated probabilities probability of x being less than equal to a given value of x and so on remember both the normal distribution as well as log normal distributions had 2 parameters mu and sigma and the exponential distribution has only 1 parameter lambda.

So, if you are given a sample that follows normal distribution from the sample you can estimate the sample mean and from the sample mean you can estimate the parameter lambda because mean is equal to one over lambda. Once you estimate this parameter your pdf is completely defined then you can obtain the cdf and start talking about the probabilities of the random variable taking on certain values also it is easy to integrate the particular pdf here lambda E to the power minus lambda X. So, you can integrate this and obtain 1 minus lambda E to the power minus lambda X typically when we are talking about time to failures we use the exponential distributions, now in many industries they talk about failures of components let us say what is the time to failure of,

let us say a bulb or a machine component that is you start using the component and we are estimating the expected value of the failure of that particular time to failure.
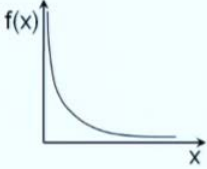
Time to failure of that particular component, but in hydrology in water resources we not do not so much talk about the component failures we talk about functional failures say for example, we may be interested in hydro power generation at a particular location, and then whenever it falls below a threshold hydropower we call it as a failure. Then we will be interested in the distribution of the time between two failures that is let us say in this particular month. We could not generate the power next time when the failure occurs the time elapse between two such events is what we will be interested and such a random variables are generally modeled using the exponential distribution, another example will be that we may be interested in time between two critical events, let us say that the flow the low flows we are interested in flow below a threshold value we are calling it as low flow.

And then we will be interested in the time that elapses between two such low flows or time between two flooding events. So, whenever we are talking about intervals of two critical events and the interval is a random variable we generally use the exponential distribution. So, let us do some example on this as said the exponential distribution is a positively skewed distribution and it is used for expected time between two critical events such as floods of a given magnitude or time to failure of hydrologic water resource systems components and so, on.

(Refer Slide Time: 52:60)

Again by this components do not mean the physical components. So, you may have functional components the ability of the system to provide a certain, let us say demand in terms of hydro power in terms of irrigation and so on. So, whenever it fails to achieve that objective then we count it as a failure and then we are interested in time between two such failures we will take a simple example here.

(Refer Slide Time: 53:42)

The mean time between high intensity rainfall as said a rainfall intensity above a specified threshold events occurring during a rainy season is 4 days that is we are talking

about the mean time, mean time between high intensity rainfall is about 4 days in the rainy season assuming that the mean time follows an exponential distribution obtain the probability of a high intensity rainfall repeating within next 3 to 5 days or within and within next to 2 days this kind of applications come typically when we are dealing with urban flooding. Let us say we are interested in high intensity short duration rainfalls at a particular location and we say that let us say for example, whenever the intensity of rainfall exceeds 90 millimeters per day then we call it as a high intensity rainfall or for design purposes we may be interested in 9 centimeters per hour in certain situations where you are talking about very short durations of 15 minutes and so on. So, we are interested in very high intensities of rainfall and the time duration between such events. So, the mean of that in this particular case mean of such an event occurring is given to be 4 days and it follows a normal.
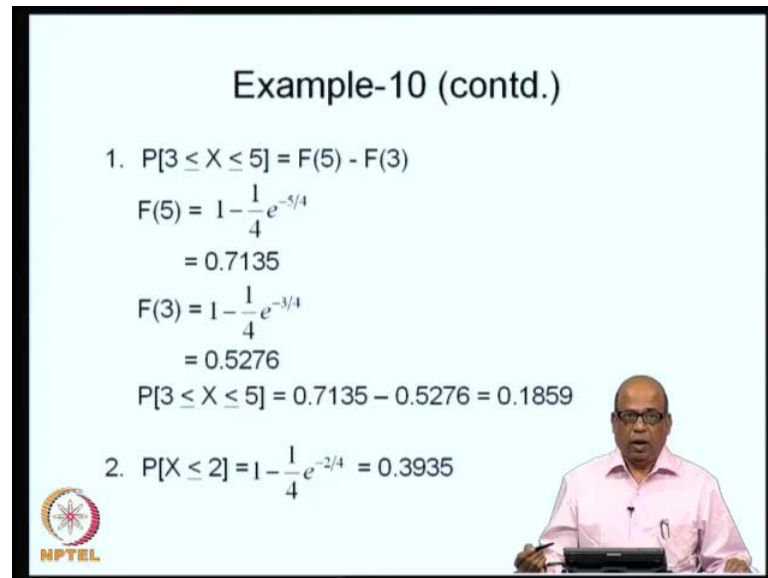
It follows a normal exponential distribution then we are interested in getting once the event has occurred already what is the probability that it will again repeat within the next 3 to 5 days or within the next 2 days. So, we are interested in getting probability that X lies between 3 to 5 days where X is the time between one event and the next and within the next 2 days. So, we estimate the lambda which is the parameter required for the exponential distribution as 1 over mu which is 1 by 4. So, once you get lambda the exponential distribution is completely defined and from that we will be able to talk about F of that is probability that X takes on value between 3 and 5 is given by F of 5 minus F of 3 from your fundamentals.

So, F of 5 is 1 minus e to the power minus 5 by 4 divided by 4 that is we are talking about F of X is equal to 1 minus lambda e to the power minus lambda minus X. So, this you get as 0.7135 and similar then we get probability of x lying between 3 and 5 is equal to 0.1859 there is a correction here. So, this is F of x is equal to 1 minus e to the power minus lambda x. So, there is a correction here well just go back here for a while and see that your F of X is 1 minus lambda e to power minus lambda X. So, we get probability of using this expression we get the probability associated probabilities. So, for today we will close at this point what we started off today is with the normal distribution we defined the standard normal density function and then solved several numerical examples sealing with the standard normal distribution.

And then went on to the log normal distribution as mentioned the normal distribution is a very commonly used distribution; however, the two limitations at the normal distribution has for hydrologic applications. Namely that there is a finite probability associated with negative values and that the normal distribution is a perfectly symmetrical distribution for most hydrologic applications these become real limitations and therefore, we generally use the log normal distribution. Now log normal distribution is a positively skewed distribution and if X follows a log normal distribution then Y is equal to l n of X follows normal distribution and we solve the methods of estimating parameters on Y given the parameters on X and then start talking about the probabilities on X because we know that y is equal to l n X follows normal distribution. Then we also introduce the

exponential distribution and solved a numerical example. So, thank you for your attention we will continue the discussion in the next class thank you.