Water Resources Systems

Prof. P.P.Mujumdar

Department Of Civil Engineering

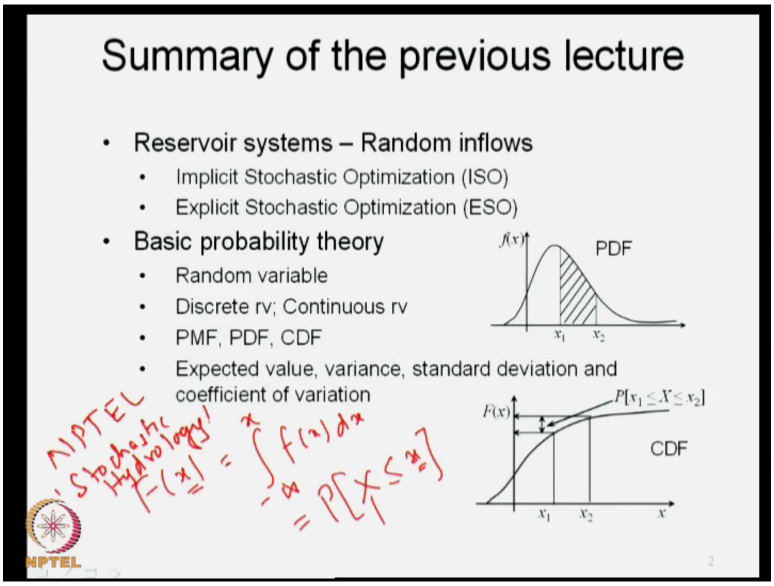
Indian Institute Of Science, Bangalore

Lecture No # 28

Basic probability theory (2)

Good morning, and welcome to this the lecture number 28, of the course, Water Resource Systems - Modeling Techniques and Analysis. In the previous lecture, we introduce the concept of random inflows, and we have just started talking about the uncertainties in the water resource systems; and how to address these uncertainties in the models; that we have developed earlier. So, in the previous lecture, we introduced the basic concepts of probability

(Refer Slide Time: 00:47)



First, I talked about, how we address uncertainties in the optimization models, recall that we can have two different structures of the optimization models, when we are addressing the uncertainties through the randomness of hydrologic variables, and specifically, we are always talking about randomness in the inflows, reservoir inflows.

So, there are two ways of doing it, as I mentioned in the last class, one is through the implicit stochastic optimization, where the optimization problem itself remains deterministic. And you address the uncertainties in the inflows implicitly by running this optimization model, which is the deterministic model, several number of times, each changing the sequences of the inflows, each time changing the sequence of the inflows.

And therefore, you will generate a sequence of outflows, a large number of outputs different corresponding to, each corresponding to the different sequence of the inflows; and then you make the analysis of uncertainty on the outflows output. This is called as the implicit stochastic optimization, because we are addressing the uncertainty implicitly.

whereas, in the explicit stochastic optimization, you will incorporate the probabilistic information, in terms of the probability distributions, in terms of the excepted value, in terms of the variance etcetera of the inflows explicitly in to the optimization model itself. So, the optimization model structure itself changes, and you address the uncertainties in the inflows explicitly through the probability distribution. And that is, what is called as explicit stochastic optimization, as I mentioned, we will cover two different types of explicit stochastic optimization in this course, one is the chance constraint linear programming, and other is the stochastic dynamic programming.

But as a (( )) to those techniques, we need to have some review of the basic probability theory; and that is what we started in the previous lecture, we introduce the concept of the random variable; and specifically, the discrete random variables and the continuous random variables, because that the discrete random variables take on only discrete values such as, 0,1,2,3 etcetera whereas, the continuous random variables can assume infinite number of values on a real line, if you are talking about a single dimensional random variables.

And then, we also introduce the concept of the probability density function for the continuous random variable; and the probability mass function for the discrete random variables; and the cumulative distribution function for the continuous random variable as well as for the discrete random variables.

Now, recall that for the continuous random variables, the capital F of (x) is given by the integral between minus infinity to x f of (x) d x; and this in fact gives probability of X being less than or equal to x; and this is the specific value of x; that were talking about.

For example, you may talk about probability of stream flow being less than or equal to 3000; and if you know the pdf of the stream flow, you integrate between minus infinity to 3000, and typically, if you are using a f of (x), which is defined for non zero values, non negative values, then you will integrate between 0 to x that particular value of x.

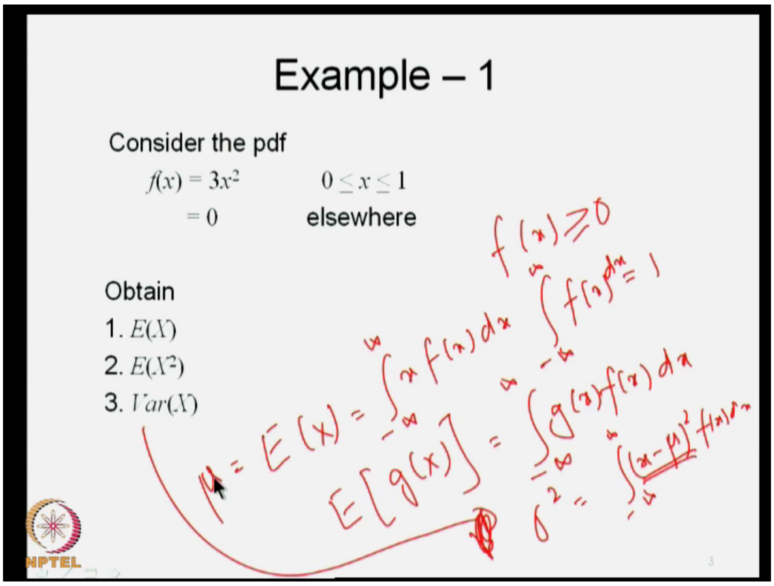
And therefore, if you look at cdf probability of X lying between x 1 and x 2, the specified values of x 1 and x 2 will be given by F of (x 2) minus F of (x 1); that is essentially you are taking the area under the pdf between the values x 2 and x 1; and that is what gives you probability of x 1 being less than or equal to x being less than or equal to x 2, which means X lies between x and x 1 and x 2.

And then, we also introduce the concept of the expected values, the variance, standard deviation, and coefficient of variation, all of which we will be using in the stochastic techniques, optimization techniques, I also mention that the details of these in slightly more rigor has been covered in the NPTEL course on stochastic hydrology, the students may refer to this course, stochastic hydrology.

Now, we will progress further; and then look at some specific distributions. Typically, the normal distribution, the log normal distribution etcetera; and then proceed to the chance constraint linear programming problem, but before that the , because the excepted value variance etcetera are important concept that will be using often in the stochastic optimization problems.

Let us look at some numerical examples dealing with these, these are extremely simple numerical examples, but thus for the sake of completeness shake for the shake of completeness, we will go through some simple numerical examples, through which we determine the excepted value variance and and so on.

(Refer Slide Time: 06:41)



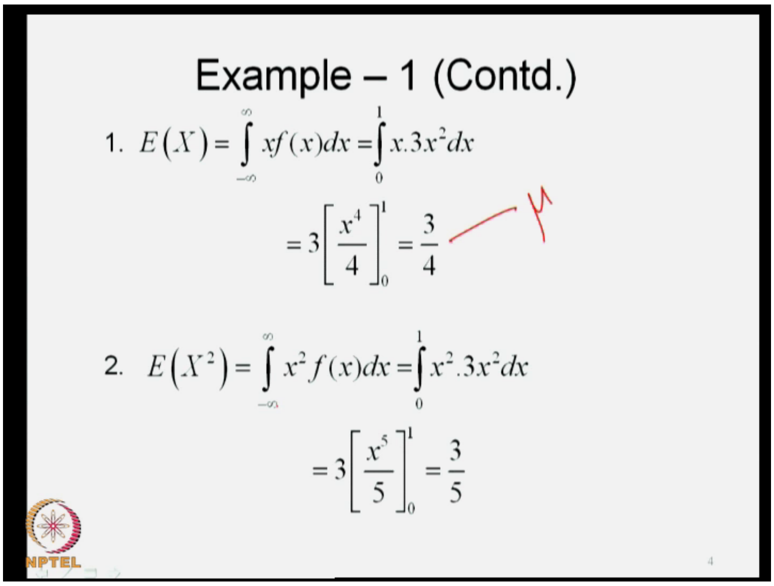
So, let us say, you have a pdf defined as f of (x) is equal to 3 x square; defined over the region 0 less than or equal to x less than or equal to 1; and this is assumed always that it is 0 elsewhere. So, for this f of (x), we will obtain the excepted value of x, excepted value of x square; and then the variance of x.

Recall that your f of (x) in the region that you are defining must be non negative. So, verify that it is nonnegative, and the integral f of (x) for the entire region must be equal to 1 with respect to x. So, this you can verify in the region 0 to 1, where when you to integrate it between 0 to 1; you will in fact get equal to 1.

And also, recall that excepted value of x, we defined as integral minus infinity to plus infinity x f of (x) d x; and excepted value of g of x, which is a function defined on the variable x is integral minus infinity to plus infinity g of (x) f of (x) d x, where f of (x) is the pdf. Now, these are the definitions, we will use those definitions, and then obtain this.

And variance of x, which is denoted as sigma square is in fact, minus infinity to plus infinity x minus mu the whole square f of (x) d x, which is the second movement about the mean. So, mu is excepted value of x and we take the second me second movement of the distribution f of (x) about the mean, and that is what we call it as sigma square. Now, we use this sigma square is also the variance. So, variance of x we denote it as sigma square, we use this; and then use these definitions, we know f of (x) we will obtain these various quantities here.

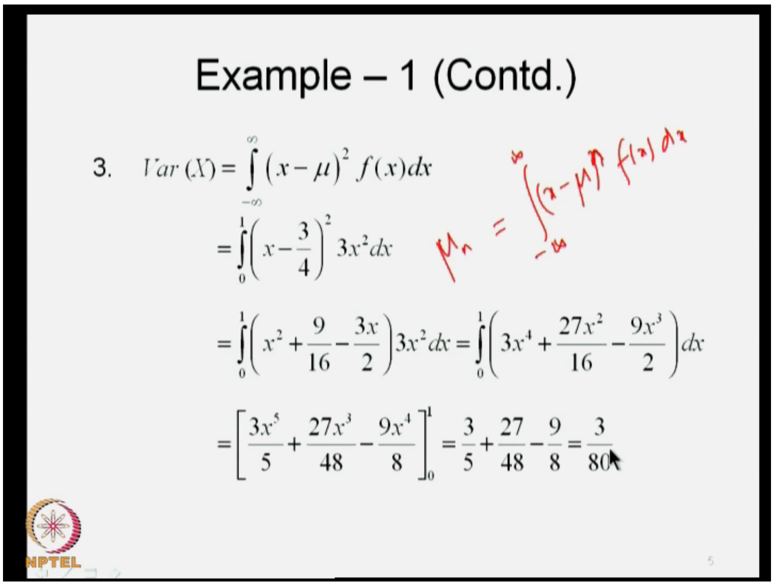
(Refer Slide Time: 09:14)



So, first excepted value of X, as I just mention it is x f of (x) d x, which is now between 0 and 1, because your f of (x) is defined between 0 and 1, it is 0 elsewhere. So, we take between the integral 0 and 1 x 3 x square d x, you get excepted value of x is 3 by 4, this is in fact equal to the mean is in fact, the mean mu.

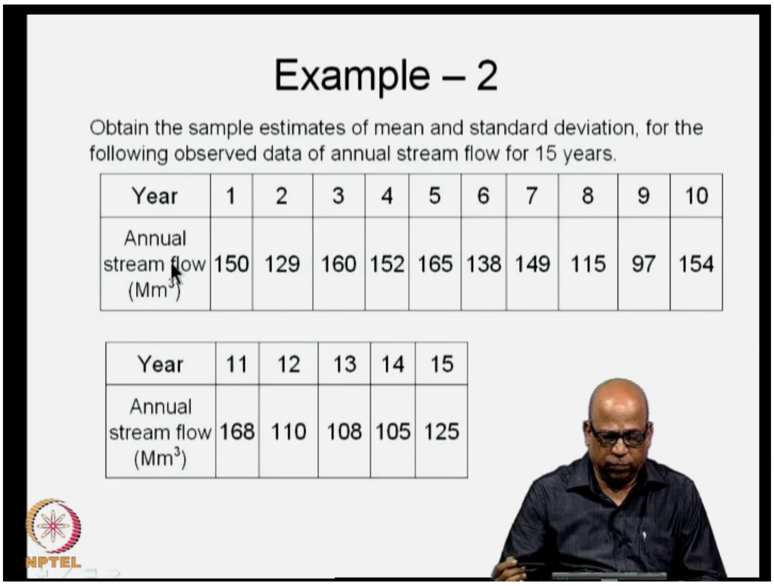
Then you have expected value of x square. Remember, I just said excepted value of any function of x is given by the integral between minus infinity to plus infinity; that function in to the pdf with respect to X. So, X square is the function here, minus infinity to plus infinity x square f of (x)d x, So that you will get it as 3 by 5, so excepted value x square is 3 by 5.

(Refer Slide Time: 10:18)



Then you get the variance of x, which is the second movement about the mean, you define x minus mu to the power n f of (x) d x. As the n th movement, this is to the power n, as the n th movement about the mean. So, and we typically, denote it as mu n minus infinity to plus infinity. So, when n is equal to 2, which defines the second moment about the mean; that is what we call as variance. So, variance is in fact the second movement about the mean, you take second movement about the mean, this integration is fairly straight forward, this is x and 3 mu is 3 by 4 and f of (x) is 3 x square, where integrating with respect to x, all the details are given here. So, you will get 3 by 80. So, the variance of X is 3 by 80

(Refer Slide Time: 11:32)



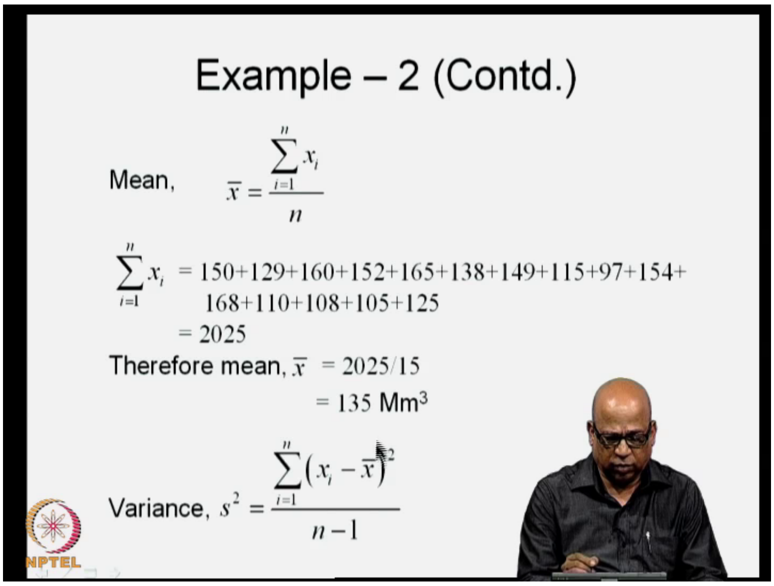
Now often, we will have samples; that means, the observed flows, what we did in thus previous examples is that we had a pdf f of (x) is equal to 3 x square. For a given pdf, we obtained all these values. Now these are population values, as I mention in the previous lecture, these correspond to the population values; for example, the excepted value is the mean for the population and so on.

But in hydrology often, we will have to rely on the samples. So, you may have observations on stream flows for example, and you will have the observations for the last so many years. Using the sample values, we must be able to estimate the population moments; for example, I may want to estimate the excepted value of x, which is the mean for the population, using the sample that we have, which is in fact the observed values of that particular on that particular variable.

So, this we do using the sample estimates of the various movements, as I explain in the previous class; for example, the sample estimate for the population mean is in fact the arithmetic average; and the sample estimate for the variance, we defined in the previous class as S square, which is summation x minus x bar the whole square divided by n minus 1. We use the n minus 1, because we are talking about, what is called as consistent estimate, you can refer to the stochastic hydrologic course of NPTEL for the details.

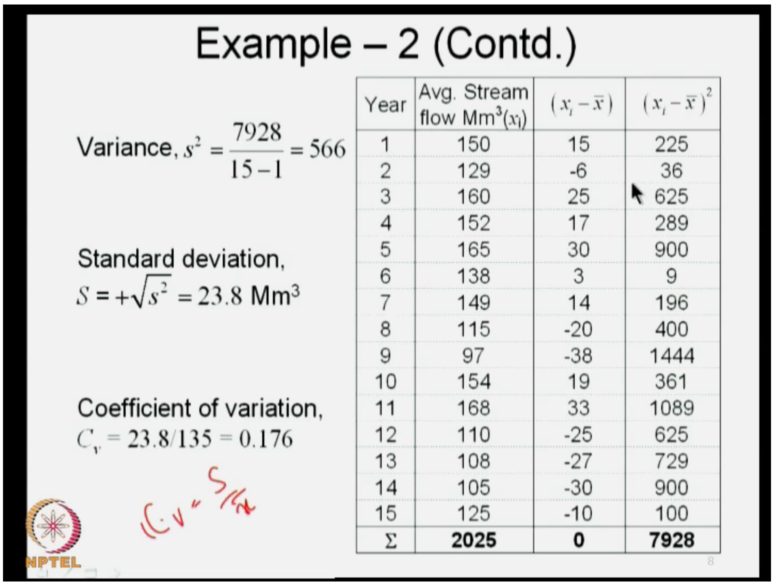
So we will see quickly, how we estimate these various moments for given samples. So, this is the simple example, we have 15 years of annual stream flow measured. And then, we calculate the mean and the standard deviation for this, this is you can do it with the calculator; most of the calculators will have statistic functions.

(Refer Slide Time: 13:45)



So, mean is simply got by x bar got as x bar is equal to sigma x i over n, i is equal to 1 to n; n in case is 15, the number of values number of observations. So, you get the sigma x i; and then you get the means as sigma x i over n, which is 15; and similarly, the variance as sigma x i minus x bar square. Once, you get x i, you get this quantity n is given as 50; and therefore, you will get S square.

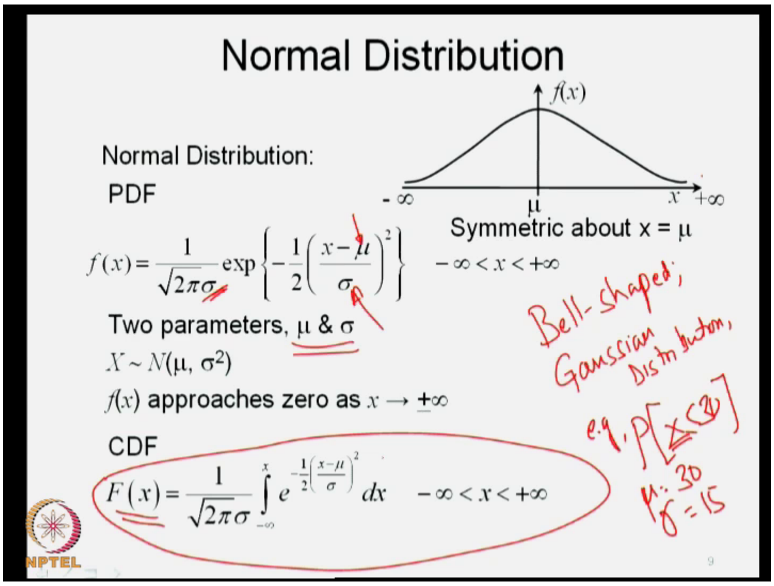
(Refer Slide Time: 14:14)



So, there are very simple calculations. In fact you can use any worksheets, such as Microsoft Excel or you can use the calculator, which has statistic functions; and obtain all of this, remember that the standard deviation is the positive square root of the variance. So, you have the sample estimate of variance, you get the standard deviation.

Once, you get the standard deviation, you get the coefficient of variation as S by x bar. So, C v is equal to S by x bar in the, as a sample estimate. So, you get S is 73 23.8; and your mean is 135 therefore you get a coefficient of variance of 0.176. Now, these are the calculations; however, these are very simple calculations, you can just verify this.

(Refer Slide Time: 15:13)



Now, what we will do is? We will go in to specific distributions. So, we know now, how to estimate the excepted value of x and the sample estimates, how to get the sample estimates of the mean, standard deviation and coefficient of variance etcetera, if you have a data.

There are a few good distributions, if you most commonly used distributions in hydrology and water resources, we will just quickly review those things, those distributions; and then see how we use them in the optimization models that we subsequently develop.

The single most important distribution is the Normal Distribution; it is also called as the Gaussian distribution; it is also called as the Bell Shaped Distribution. It is in fact bell shape and it is symmetric about x is equal to mu, the normal distribution looks like this mu is here, mu is the mean; and it is defined for x is equal x going from minus infinity to plus infinity. So, this is the entire range of x it is defined from minus infinity to plus infinity.

The pdf f of (x) for the normal distribution is given by this expression here, it is also called as the Bell-Shaped Distribution, it is also called as the Gaussian Distribution, it is most commonly used in hydrology in fact as the first approximation, you always use the normal distribution, if you want to address uncertainty inflows and uncertainty in let us seasonal rainfall or monthly rainfall, soil moisture, ground water levels and so on.

As a first approximation, you straight way use the normal distribution for various reasons, the normal distribution is most commonly used; however, you must note that the normal distribution is defined for the entire region minus infinity to plus infinity; and therefore, even if you have a very high mu; let us say, your normal distribution comes from this point to this point and this is 0.

Let us say, the 0 is somewhere here this has syntactically reaches minus infinity. So, even if your 0 somewhere here, there is a non zero probability associated with flows going below 0; 0 and that is why it you may have some difficulty in using the normal distributions for hydrologic variables, which are mostly non negative. However, as I said as the first approximation, as the first trial in most of the processes, we use the normal distribution.

The f of (x), which is the pdf is given by 1 over root 2 pi into sigma. Now, sigma is the standard deviation, exponential minus half x minus mu ,mu is the mean divided by sigma the whole square; and this is defined for minus infinity less than x less than or equal to plus infinity. So, for the entire region, it has two parameters here, yet mu and sigma. And it can be shown that the mu is in fact the mean and the sigma is the standard deviation; that means, you you would have put beta 1 and beta 2 here; and beta 1 turns out to be the mean and beta 2 turns out to be the standard deviation. So, it has two parameters mu and sigma.

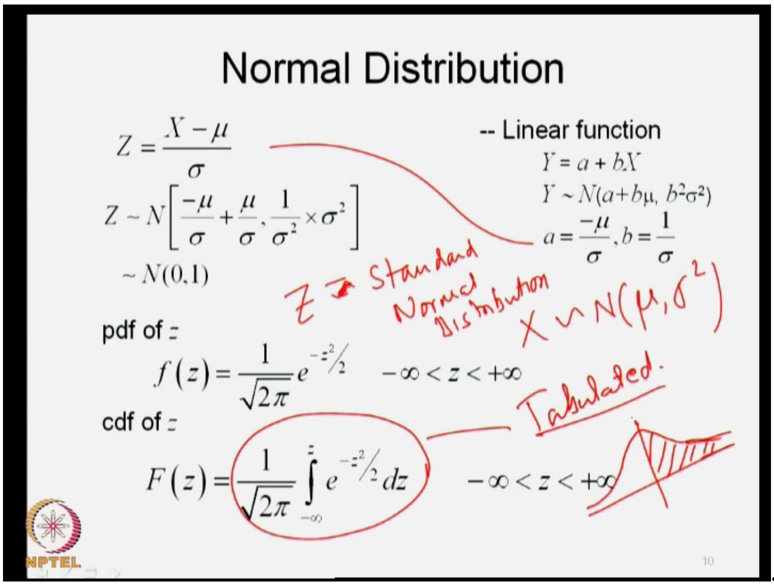
We denote by this notation, X follows normal distribution with mu and sigma square, sigma square is the variation. So, you can use here mu and sigma square. So, these are the two parameters of the normal distribution. So, the normal distribution is indicated by this notation at X, which is the random variable follows the normal distribution with parameters mu and sigma square, which means the movement you give mu and sigma, this distribution is completely defined. Every probability distribution, we will have certain parameters and the normal distributions, has two parameters mu and sigma square.

Now as you can see here, the density function approaches 0 as x tends to infinity as in practically it goes to 0 on either side minus infinity as well as plus infinity. By definition of the cdf f of (x), which is the capital F of (x), you get from small f of (x), you integrate between minus infinity to x the f of (x); you will get capital F of (x). So, this is the expression for the cdf.

Now, to get probabilities associated with the random variable x here; for example, we may be interested in probability of X being less than or equal to 2 sigma or mu plus 2 sigma or proab if you have specify the mu and sigma as some values; let us say, mu is the 30 and sigma is 15 for a particular sample, we may be interested in events such as probability of X being less than or equal to a given value of x. For example, probability X being less than or equal to 30, when mu is 30 and sigma is 15; such probabilities, you can determine based on the cdf f of (x) provided, you can integrate this expression.

Now, the integration of this expression is not straight forward; and analytically it is not possible or it is difficult; and therefore, we adapt numerical integration for this. So, you can use numerical integration to get this f of (x) value, f of (x) denotes probability of X being less than or equal to x; and therefore, you can talk about any related probabilities on the random variable X using the numerically integrated expression for this. However, because this distribution is defined for a given mu and sigma, as your mu and sigma changes, you will have to redo the integration. Fortunately, for normal distribution, this difficulty can be overcome.

(Refer Slide Time: 22:59)



Because of a interesting property of the normal distribution, namely that if you define a linear function on the random variable X; that is X follows normal distribution with mu and sigma square. So, we know that X follows normal distribution with parameters mu and sigma square. And if Y is a linear function of X defined as, a plus b X, then Y follows normal distribution with the parameters a plus b mu and b square sigma square, which means the variance of Y will be b square sigma square, and the mean of Y will be a plus b mu. Now, this is the interesting and important useful property of the normal distribution.

So, if you define a function x minus mu over sigma. And we call it as Z, X follows normal distribution, and Z is a linear function defined as X minus mu over sigma, then you can see that Z follows normal distribution using this in this particular case, a will be minus mu over sigma, b is one over sigma. And therefore, by substituting this, you will see that Z follows normal distribution with mean 0, and variance 1. Now, this is the important and very useful property of the normal distribution, because for a given X, you can obtain Z; and Z you know that it follows normal distribution with 0 0 mean and unit variance. Now, Z is called as the standard normal distribution.

(No audio from 24:49 to 25:05)

And, because Z has a normal distribution with 0 mean and unit variance, you can now define the pdf of z as f of (z) as the pdf of z. In fact, it turns out to be 1 over root 2 pi sigma is 1; and therefore, you do not have sigma here, e to the power minus z square by 2, because mean is 1 and the variance is 1 I m sorry mean is 0 and variance is 1 and this is defined over the entire region minus infinity to plus infinity. And f of (z), which is the capital F of (z), which is the cdf of z is it turns out to be 1 over root to pi minus infinity to z e to the power minus z square by 2 into d z.

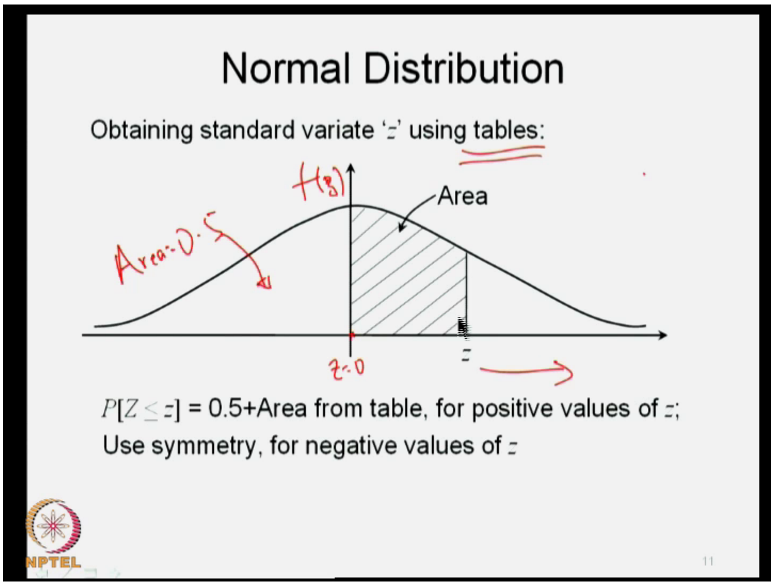
Now, look at this integral now, this integral can be carried out again numerically, and the you can tabulate for various value of z; that means, I specify z. And I have from the tables this integral given, then we can use that integral, which is capital F of (z) to talk about probabilities on X, because we know the transformation Z is equal to X minus mu over sigma.

I can talk about probabilities on Z; and then transfer then to probabilities on X, this is the great advantage of the normal distribution; that means, irrespective of your mu and sigma of the original variable X, you can always convert it into a standard normal distribution, which will have a 0 mean and unit variance; and then using the standard normal tables, as I will just explain, you can talk about probabilities on x.

So, these integrals are tabulated; and there available in most of the standard textbooks on probabilities in statistics, you can also use the tables that are available in MS Excel and other worksheets or any scientific or statistical toolbox; that is generally available in let us say, in math lab or some other program, statistical mathematical program that you have, they will generally have incorporated in them the tabulated values for f of (z).

But when you are using the tables, you be alert that different programs may give different ways, the tables are given in different ways in different programs; for example, typically we may have the tables giving only this area, because of symmetry you can use this tables; and then calculate the probabilities associated with this part of the distribution also, we will do some simple examples now on normal distribution. So, that you can, you will know, how to use the normal distribution of for computing various probabilities, we will use the tables tables as specify as given in some standard text books.

(Refer Slide Time: 28:28)



So, in the particular examples that I will be dealing with we use the tables, which provide this area. So, this is the normal distribution, this is the f of (x) f of (z) here; and this the value of z here in this direction; that tables give in the in the tables that I will be using the tables give the area to right of z is equal to 0. Now, this z is equal to 0 here, for a specify value of z, we get this area and what is this area by symmetry, this total area is 0.5, because the total area under the pdf has to be 1. So, to the left of z is equal to 0, you have the total area has 0.5, and the total area to the right is also 0.5, and the tables give this area, what is this area? This is f of z plus 0.5, because you are talking about this total value, let me rewrite correctly there.

So, up to this point, if you take the total area, you will get associated with this value of z, if you take the entire area from minus infinity to this value of z, you will get capital F of (z). So, we will see how we use these tables and for specified values of mu and sigma; and then talk about the probabilities associated with the random variable x. So, we convert the random variable x into the standard normal variable z, standard normal variant, we call it as z; and then start talking about the probabilities on z.

(Refer slide Time: 30:45)

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	0.2764	0.2823
0.8	0.2881	0.2939	0.2995	0.3051	0.3106
0.9	0.3159	0.3212	0.3264	0.3315	0.3365
1	0.3413	0.3461	0.3508	0.3554	0.3599

What I mean by that is; that you have normal distribution tables, which are given for values of z here. So, z is equal to 0, 0.1; let us say, we are reading z, z is equal to 0.12 or 0.14, 0.16 etcetera. So, these are the areas, and this area corresponds to this area here to the right of z in the tables that I am showing here; and some of the tables may give the complete area up to this point. So, just be alert to the particular area under the pdf that the table is giving, and then use the tables accordingly.

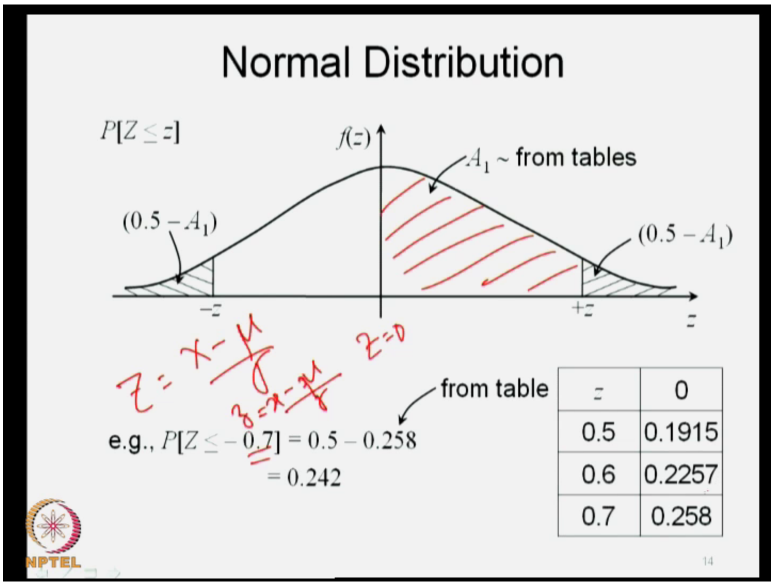
(Refer slide Time: 31:30)

Normal Distribution Tables

z	0	2	4	6	8
3.1	0.499	0.4991	0.4992	0.4992	0.4993
3.2	0.4993	0.4994	0.4994	0.4994	0.4995
3.3	0.4995	0.4995	0.4996	0.4996	0.4996
3.4	0.4997	0.4997	0.4997	0.4997	0.4997
3.5	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5	0.5	0.5	0.5	0.5

So, this is how it looks and typically, it goes up to about 3.99 or something, where the entire area is covered covered. So, the entire area to the right of z will be 0.5 nearly and this will nearly 0.5, it is not exactly 0.5, it will be nearly 0.5, as you can see here. So, we will now do some examples by which you can see, how to apply the normal distribution

(Refer slide Time: 31:57)



We should be also able to use the symmetry. Let us say, this is the normal distribution that is the standard normal distribution f of (z); and from the tables, we get this area for a given value of z, we get this area from the table.

Let us say, I want on the negative side, I want corresponding to minus z, I want the area here, we use the symmetry, we read the area corresponding to plus z, which means essentially, I will be getting this area, and because the area to the right of plus z will be 0.5 minus A 1, if this is the A 1 that you read from the table, this will be 0.5 minus A 1, because the total area to the right of z is equal to 0 will be 0.5.

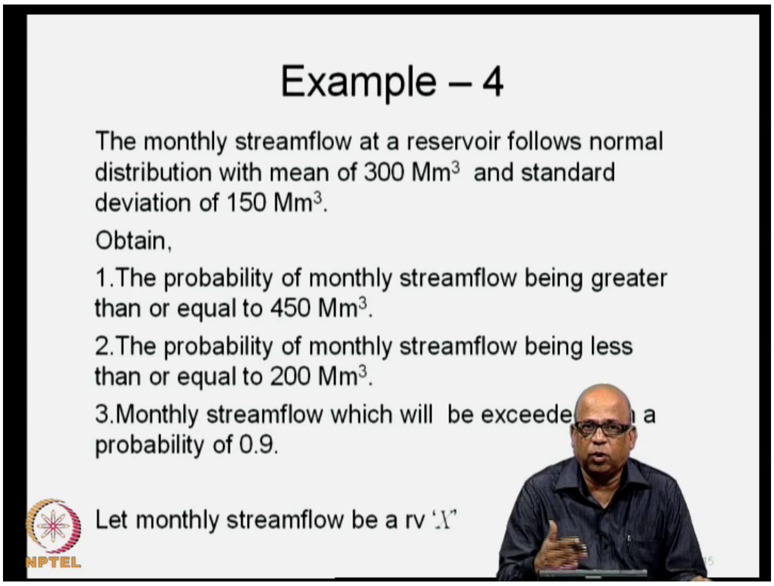
This you transferred to the left side now, takes the mirror image, you will get associated with minus z; you will this area as 0.5 minus A 1. So, this is we are talking about probability of Z being less than or equal to this given value of minus z. So, this is how we use the symmetry, when you are using areas from the tables, areas given by tables.

Let us say, we do one simple example; that is Z being less than or equal to minus 0.7 is what we are interested; that is we want probability Z being less than or equal to minus

0.7, you go to the table first of all, you look at associated with z is equal to plus 0.7, you look at the table area. So, corresponding to z is equal to 0.7, you get 0.258; that is this area is 0.258; and therefore, this area becomes 0.5 minus 0.258 and that will be the same as 0.5 minus 0.258, which corresponds to minus 0.7. So, probability of Z is less than or equal to 0.7 is equal to 0.5 minus the area that you get from the table, which is 0.258, which will be 0.242.

Now, this we are still talking on Z, the variable Z, which is the standard normal deviate, but how do we get this Z, we get this Z from your original random variable X and by using Z X; Z is equal to X minus mu over sigma, and the specific values of this Z here can be got by this specific values of X for which you are talking about talking about the probabilities; and you can use Z is equal to X minus mu over sigma, to obtain these specific values of Z, we will see some the examples. So, this is how you use the standard normal distribution. And the tables that are available with you, which provide typically this area here, the shaded area that I shown in red.

(Refer Slide Time: 35:25)



We will do some examples now, on normal distribution. Let us say, the monthly stream flow at a reservoir follows normal distribution with mean of 300 million cubic meters, and standard deviation of 150 million cubic meters, these parameters you can obtain from the historical data. So, you have the historical data from which you obtain the mean and you obtain the standard deviation.

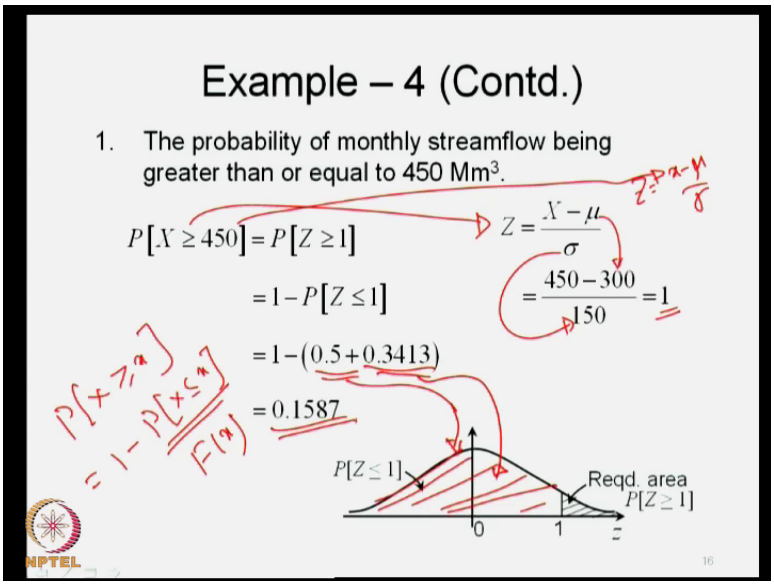
Now, you assume that it follows normal distribution; and then we will calculate various probabilities. So, the probability of monthly stream flow being greater than or equal to 450 million cubic meters. So, this is one of the probabilities, that we determine; and this is the typical problem that we come across in water resource systems, where we are, we need to estimate the probability of the monthly flow being greater than or equal to a certain given value. Now, these kinds of question should be important for planning for let us say, irrigation, hydro power generation, municipal and industrial demands and so on. Where we want to estimate the lightly hold of the flow being greater than or equal to a specified value.

Then the probability of monthly stream flow being less than or equal to another value 200 million cubic meters, in doubt situation, you want to estimate, what is the likely would that the flow will go below 200 million cubic meters, you want to plan for hydro electric power generation, you would like to see, what is the likely hood that it will go down below a certain region; for water quality, you will be interested in low flows, probabilities of low flows, in such situation you will be interested in this.

Then for the planning purposes, let us say, the water sharing between different state or the operation of the reservoir or the sizing of the reservoir, when you want to make decisions and sizing of the reservoir and so on. You typically, use exceedence probabilities; that is, what is the probability that a given flow will be exceeded 75 percent of the times, 95 percent of the times and so on. So, we will also calculate monthly stream flow, which will be exceeded with the probability of 0.9; that means the magnitude of the stream flow, which will be exceeded with the probability of 0.9.

So, we will use the flow as the random variable x, we denote the flow as the random variable x; and then use the concepts that we just develop, just revised, and apply this to determine these various probabilities. So, the first one, which is probability of monthly stream flow being greater than or equal to 450, remember the normal distribution is completely defined the, when you define the mean and the standard deviation. So, you have a mean as well as standard deviation. So, your normal distribution is completely defined, we should be able to use the normal distribution now, for define for determining these probabilities.

(Refer Slide Time: 38:32)



So, first one we write it as probability of X being greater than or equal to 450, now this we convert it into standard normal distribution, we write this as using Z is equal to X minus mu over sigma, now the X that we are talking about, now we may write it in terms of the small x, the particular value of x is this x, 450 minus mu is the mean and sigma is the standard deviation, which is given. So, first you convert it in to Z, which turns out to be in this case Z is equal to 1.

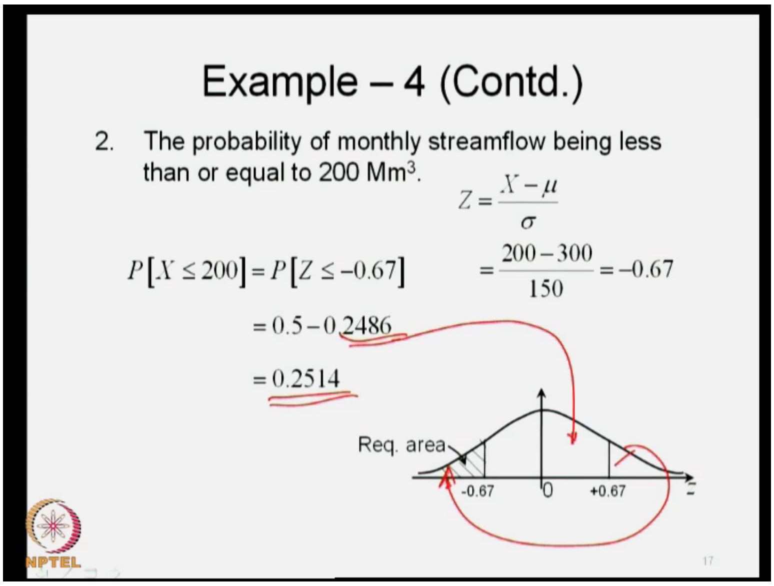
So, probability of X being greater than or equal to 450 is in fact equal to probability Z being greater or equal to 1. So, this is the standard normal distribution now, the moment you converted into in terms of Z, you can go to the tables and then pick up the particular areas. As I mention in the last class probability of X being greater than or equal to x is in fact equal to 1 minus probability of X being less than or equal to x.

And typically, we talk about these kind of probabilities, where we are saying X is less than or equal to x, because you have this as F of (x), which is the cdf. So, I write this probability of Z being greater than or equal to 1, as 1 minus probability of Z being less than or equal to 1; and what is this probability of Z being less than or equal to 1, it is capital F of (z) equal to 1; that is capital F of (1).

So, we go to this table now, look at this figure, we are looking at the area associated with Z is equal to 1, area under the pdf associated with Z is equal to 1. So, our table gives this area now, we want probability of Z being greater than or equal to 1. So, we are looking at this region here Z being greater than or equal to 1. So, we need this area, we are writing that as 1 minus probability of Z being less than or equal to 1.

And therefore, we are looking at this area now, and from the table what we get is, this area. So, this 0.3413 in fact corresponds to this area now, from the table; and when I add 0.5, I come up to this point. So, 0.5 plus 0.3413 brings me up to this area. So, this whole area is 0.5 plus 0.3413 and that gives me probability of Z being less than or equal to 1, but, because I am interested in probability of Z being greater than or equal to 1, I take 1 minus of this, which comes to 0.1587. So, starting with the mean and the standard deviation of the data; we have now, obtain the probability x being greater than or equal to 450.

(Refer Slide Time: 41:59)

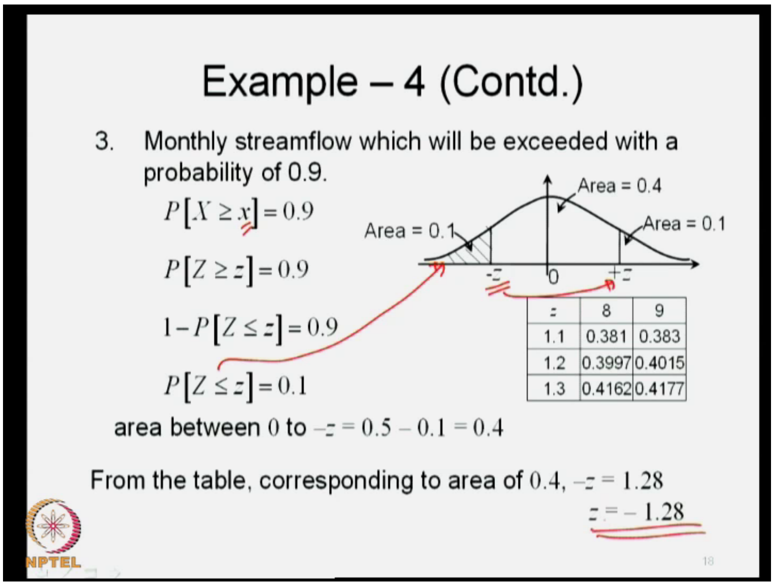


Now, we will go to the second problem, where we are talking about probability of monthly stream flow being less than or equal to 200 million cubic meters, which means we are talking about X being less than are equal to 200. Now, again we like, we did earlier, we convert this in to Z now. So, 200 minus 300, 300 is the mean by sigma which is 150, I get minus 0.67. So, we are talking about probability of Z being less than or equal to minus 0.67.

Now, for minus 0.67, first you look at the area corresponding to plus 0.67 from the table, you get this area as 0.2486. Now, this area will be then 0.5 minus 0.2486, which will be the same as this area; and that is how you get probability of X being less than or equal to

200 as 0.2514. So essentially, we convert this in to Z; and then look at that associated value of Z here, because in the particular tables that I am using, I get the areas to the right of Z is equal to 0, I use the symmetry; and then translate transfer that particular area to this area and this in fact, gives you probability of Z being less than or equal to minus 0.67.

(Refer Slide Time: 43:29)

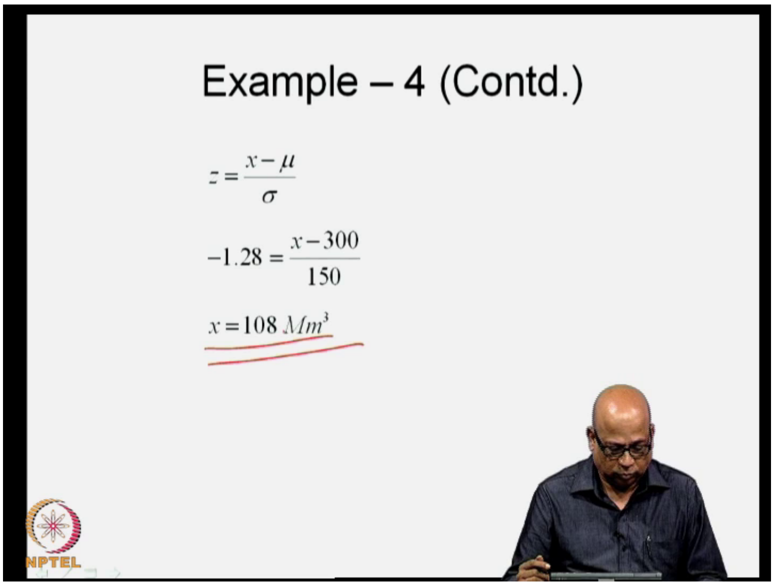


The third problem, where we are interested in probability of X being greater than equal to a particular value of x; and we want to determine the value of x. So we are saying, what is that particular flow, which should be exceeded with 90 percent probability. So, we are saying probability of X being greater than or equal to x, which is unknown will be equal to 0.9; and we write this as state away probability Z being greater than or equal to z is equal to 0.9 and that we write as 1 minus probability of Z being less than or equal to z is equal to 0.9; therefore probability of Z being less than or equal to z will be 1 minus 0.9, which is 0.1.

So, probability of Z being less than or equal to z is 0.1, which is we are saying that this area here is 0.1, because this is what denotes probability of Z being less than or equal to z. up to 0.5, the z values will be negative that is up to the area being equal to 0.5, the associated values will be negative. So, to get this area now, what I will do, I will write this as minus z and take to the right side plus z; and therefore, this area will be 0.1 by symmetry. I will look at this area now, because the tables will give this area. So, this area is 0.4.

So, I will go to the table and look at these values. So, I am looking at that particular value of z, which gives me and area 0.4 between 0 and that z. So, I go to tables and look at the area under the curve to be equal to 0.4. So, this comes out to be 1.28. So, the z transferred to be 1.28 in this case. Now, this is this z. So, I will use minus z here; and therefore, z will be minus minus 1.28.

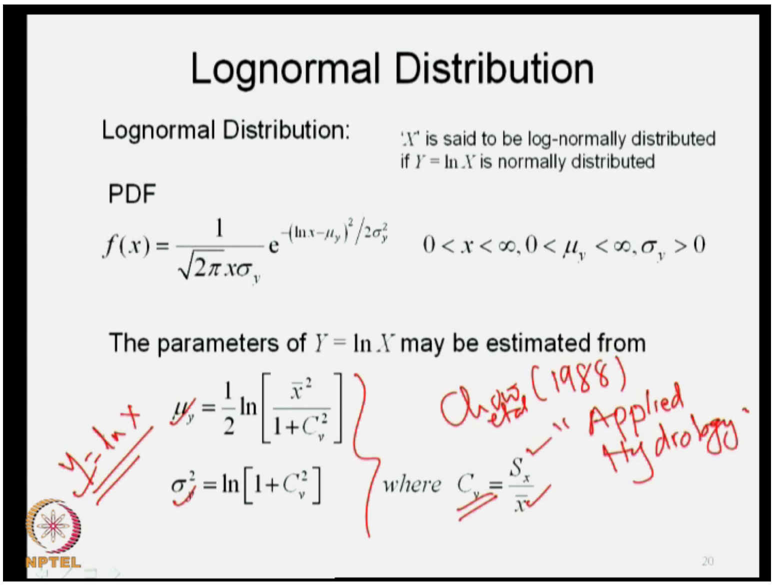
(Refer Slide Time: 45:40)



Once you get this Z, you can get the associated value of x, because z is equal to x minus mu over sigma. So, I will put z is minus 1.28; x is x, mu is given as 300, sigma is 150; and therefore, I get x is equal to 108 million cubic meters. Now, the interpretation of this is that the flow, which will be exceeded the magnitude of the flow, which will be exceeded 90 percent of the time is 108 million cubic meters, using the normal distribution for that.

So, what we did here is that we essentially always convert x into z, z is the standard normal distribution, the standard normal variant; and then talk about the probabilities on z, because the capital F of (z), which is cdf of z is tabulated, and we use the tabulated values of cdf on z; and then convert them into the probabilities on x.

(Refer Slide Time: 46:40)



Now, as I mention the normal distribution has the important property; then there is a non zero probability associated with negative values of x, and most of the hydrologic variables are non negative; for example, stream flows cannot be negative, the rainfall cannot be negative and so on. So, when we are applying normal distribution we always get a finite non zero probability associated with negative values of the random variable, also the normal distribution is a perfectly symmetrical distribution; whereas, most of the hydrologic processes, hydrologic variable may have non un symmetric distributions, they may have a (( )) to right or (( )) to left and so on.

Now that is, when the other distribution become handy. So, we will introduce a lognormal distribution, which is also commonly used especially for stream flows and rainfall. We say y X follows log normal distribution, if the transformation Y is equal to ln X; that is the log X to the base e, natural base follows normal distribution, I repeat that we say X follows log normal distribution, if Y is equal to ln X follows normal distribution.

So, whenever I use a random variable X and indicate that it follows log normal distribution. The transformation Y is equal to ln X follows normal distribution that is all. Then you can transform the original variable in to ln X, Y is equal to ln X and then use the normal distribution on Y.

So, we will say ‘X’ is set to be log-normally distributed, if Y is equal to ln X is normally distributed. This is the useful way of addressing the normal distribution; and the pdf turns out to be, now we are referring to the variable Y as equal to ln X. So, whenever I am writing sigma y, mu y etc. It refers to the variable Y here.

So, f of (x) which follows log normal distribution will be equal to 1 divided by root 2 pi x sigma y e to the power minus ln x minus mu y whole square by 2 sigma y square, I am talking about the pdf of x, which follows a normal a log normal distribution now, and you have the moments of moments on mu y and sigma y; that is this parameters are associated with the variables Y is equal to ln X, what what I mean by that.

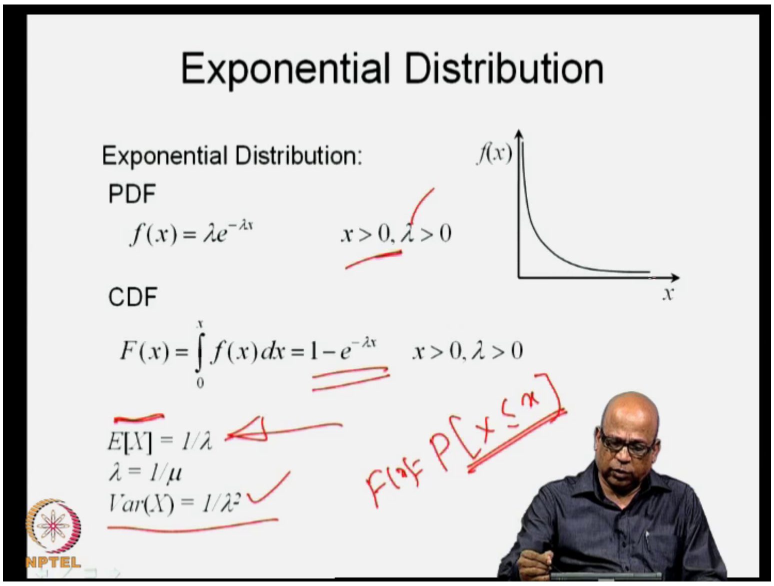
Let us say that you have the stream flows, and stream flows we are indicating by the distribution, by the variable X, capital X; and then we have transform the capital X as Y is equal to ln X; and we are writing the pdf on the variable x. So, this is the pdf on the variable X, we use sigma y and mu y on Y, which is Y is equal to ln X.

And therefore, because we are defining it on x, we will define the limits also on x, but, we are define we are having the parameters on y. So, we define the limits on y. So, this is the pdf; however, for applications we will not actually use this pdf, what we typically do is simply transform the X into Y is equal to ln X, and use the normal distribution.

Now, for this now, mu y and sigma y in hydrologic literature, there is there is a simple method available to transform the parameters on X, which is your original random variable to the parameters on Y, what I mean by that is let us say, your original random variable is X stream flows that you have transform that you have transformed to another variable Y, Y is equal to ln of X, then you can obtain mu y and sigma y square by using these expressions. Now, these expressions are given in chow 1988; which is a applied hydrology chow etal in fact, there is a “Applied Hydrology” textbook.

Now, S x is given, this is the standard deviation on x, x bar is given which is the mean on x, these two can be got from your original random variable; therefore the C v can be got. Once you get C v, you can get mu y and sigma y square using these expressions. Once, you get the mu y and sigma y, because you know that y is equal to ln X is normally distributed, and the normal distribution is completely defined. Once, you defined the mu and the sigma of that particular variable; and therefore, you can go with these two parameters directly to log normal directly to the normal distribution; and then like we did just now, you can talk about probabilities on Y. So, this is about the log normal distribution.

(Refer Slide Time: 52:11)



Then we also have the exponential distribution, which is f of (x) which is the pdf f of (x) given by lambda e to the power lambda x. Now lambda is the only parameter here, and lambda is estimated by excepted value of X is equal to 1 by lambda. Form this now, this is the pdf, you get the cdf, which is the cumulative distribution function, because it is defined for x greater than 0; the capital F of (x) will be integral 0 to x f of (x) d x; and then, you can verify that this transferred to be 1 minus e to the power minus lambda x, this has only one parameter lambda unlike the log nor normal and log normal distributions, which had two parameters mu as well as sigma.

Given the excepted value of X; let us say that you have the sample, from the sample, you have determine the mean or the sample mean from that you should be able to estimate lambda, as excepted value of X is equal to 1 over lambda or lambda will be equal to 1 over mu; and the property of the exponential distribution is also that variance of x turns out to be 1 over lambda square, because f of (x) will be given by this expression now, 1 minus e to the power lambda x, recall that f of (x) is simply probability of X being less than or equal to x.

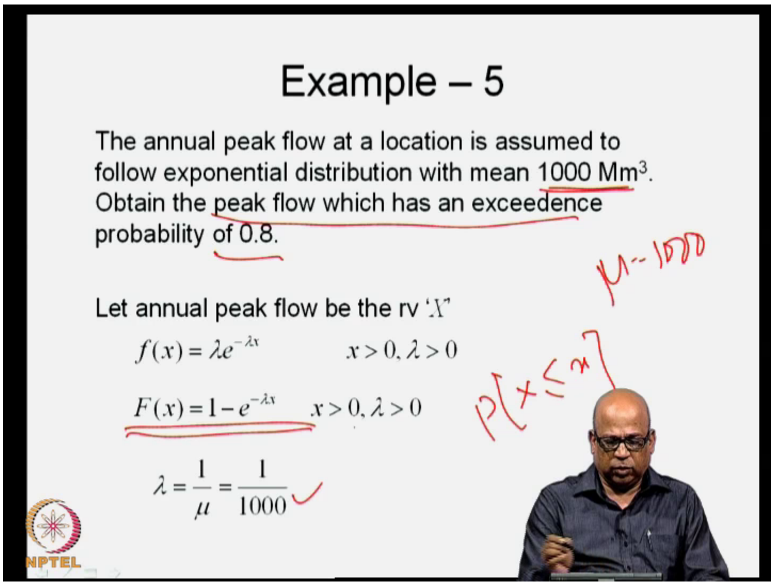
And therefore, you can determine various probabilities associated with this, if you have the sample on x; let us say, that you have a stream flow that a particular location, and the stream flows, you want to you want to fit in exponential distribution to that typically normal stream flows stream flow do not fit exponential distribution, but what may perhaps fit the exponential distribution is the maximum values of the such inflows ,annual maximum that you are talking about or you are talking about intervals between flood deviance, given flood deviance; these may typically fit the exponential distribution.

The exponential distribution looks like this, it is in fact exponentially decaying with respect to X; and the total area under the curve must be equal to 1. So given any distribution, you now know, how to calculate the probabilities associated with the random variable X.

Let us say, we will quickly do another small example on the exponential distribution itself; on normal distribution we have done a example. So, you know how to calculate the probabilities associated with the normal distribution, then we went on to the log normal distribution, for the log normal distribution, if you have the variable X following log normal distribution, you simply convert that into Y is equal to ln X; and Y is equal to ln X follow the normal distribution; and therefore, you work with normal distribution still.

Now, for the exponential distribution, if you have the sample, which follows exponential distribution, you can determine the parameter lambda as (( )) equal to 1 over mu, mu is the mean, mean you can calculate from the sample. So, you have the sample estimate for the mean; and therefore, lambda gets fixed, lambda is the parameter for the exponential distribution, once you get the lambda you have a cdf f of (x), which is given by 1 minus e to the power minus lambda x; and therefore, you should be talk about probabilities.

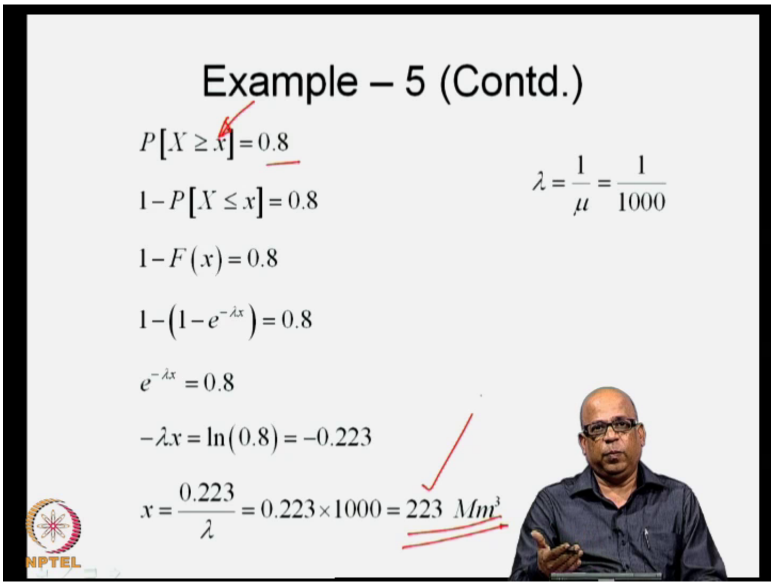
(Refer Slide Time: 55:39)



We will quickly do a Simple example, now annual peak flow at the location is assume to be an exponential distribution, assume to follow exponential distribution, which mean of 1000 million cubic meters. So, mu is given as 1000 million cubic meters, and we want that particular flow, which has an exceedence probability of 0.8.

Now, look at this point here, F of (x) is equal to 1 minus e to the power lambda x. So, this gives the probability of X being less than or equal to x and lambda is 1 over mu therefore, 1 over 1000.

(Refer Slide Time: 56:23)



From this now, we can get, what we are ask to find out this is this particular value of X, which will be exceeded with the probability of 0.8. So, we want to obtain this value; and therefore, we write probability of X being greater than or equal to x is equal to 0.8, and that I write as 1 minus probability of X being less than or equal to x equal to 0.8, because probability of X being greater than or equal to x is 1minus probability of X being less than or equal to x.

From this now, we get x is equal to 223 million cubic meters, this is the fairly straightforward calculation, you can verify that x is 223 million cubic meters, which means, if those peak flows follow exponential distribution with a mean of 1000 million cubic meters, we get the particular flow, which will be exceeded with 80 percent probability or probability of X being greater than or equal to x is equal to 0.8 is in fact 283 million cubic meters. Now, these kinds of calculations will be useful, when you are talking about protection against flood flood peaks and so on.

So, essentially in today's lecture, we have introduce the normal distribution, and did some example associated with the normal distribution; and then we went on to introduced the log normal distribution and the exponential distribution. Given any distribution; that means, given any f of (x) small f of (x), which is the pdf, we should be able to talk about probabilities on capital X, which is the random variable. We talk about even such as, what is the event such as the peak flows exceeding the certain value or the flows being less than or equal to certain value etcetera. Now, these are important a concepts, when we talk about water resource systems, planning and operation.

Also, we will now start introducing the important concept of chance constraints or the reliability constraints. In the next lecture now, we will progress to formulating the linear programming problem; that we developed earlier for determining the reservoir capacity, using the concepts of probabilities; that means we start introducing the uncertainties associated with the inflows, and convert the original deterministic optimization problem into a stochastic optimization problem; in fact, an explicit stochastic optimization problem, where we will use the probability distributions on the inflows directly into the optimization model. So, we will continue the discussion in the next lecture. Thank you very much.