Water Resources Systems
Prof. P.P.Mujumdar
Department Of Civil Engineering
Indian Institute of Science, Bangalore

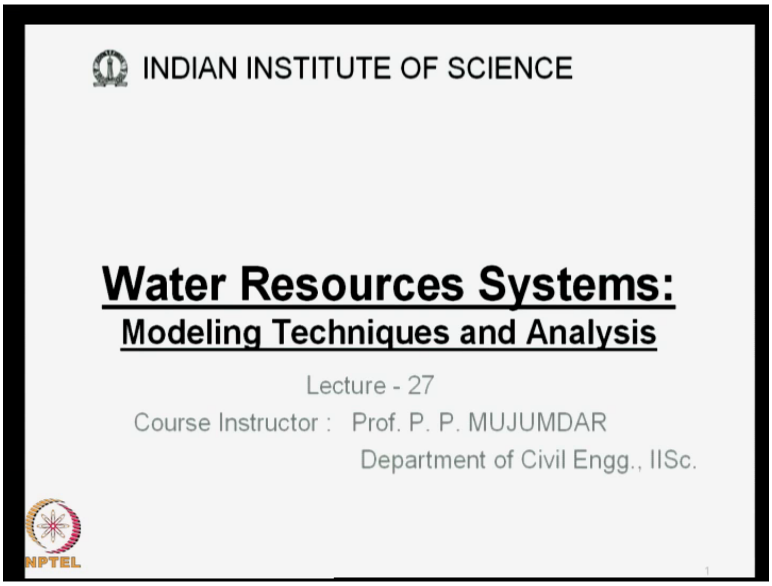
Module No # 06

Lecture No # 27

Basic probability theory (1)

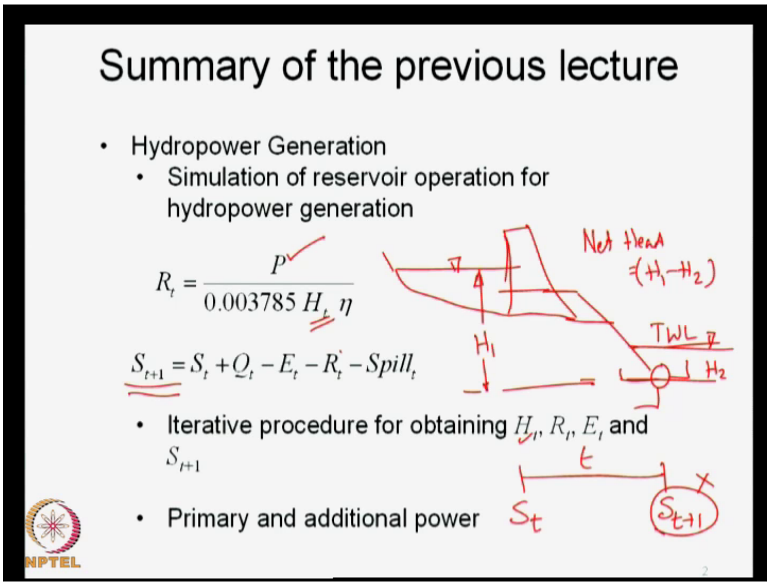
Good morning, and welcome to this the 27 lecture of the course, Water Resource Systems: Modeling Techniques and Analysis.

(Refer Slide Time: 00:18)



The over the last few lectures, we have been discussing about the reservoir operation problems, reservoir design problems and so on. in a deterministic framework, in the sense that the sequence of inflows is known and all the hydrologic variables that influence the decisions are suppose to be known with certainty and these framework, this type of framework is called as a deterministic framework. And in that deterministic framework, we have made several decisions or we have built models to make decisions on what is the minimum reservoir capacity that you need to provide to meet a certain sequence of demands or how do you operate the system optimally to meet certain objectives to maximize let us say, the hydropower generated or to maximize the returns that you expect and so on. Even in the case, where we considered the stationary policy or the steady state policy using the dynamic programming, the sequence of inflows was constant, was head constant. In the sense that, the same sequence keeps repeating year after year, if you have annual flows and if you have a larger sequence of flows that you would like to make use in the decisions. Let us say, a thirty years of sequence, you want to make the a make use of that particular sequence in making the decisions, you may build the models or those thirty years hydrologic variables, including the steam flow and so on. But the thirty years sequence still remains the same after thirty years, in the sense that we assume that essentially the same thirty year sequence keeps on repeating after the thirty year windows. Now, these are called as deterministic problems.

(Refer Slide Time: 02:21)



So, specifically in the previous last lecturer, we dealt with the hydropower generation problem and in that we dealt with both the ramp of the reverse system, as well as the hydropower generation for with reservoir systems. So, in the reservoir systems, recall what we did, you have a pen stock and then you have the turbine located somewhere here. So, this defines the head available for power generations, this is the water level and if you have a tail water level and let us say, this is H 2, as we used in the last class. So, this is total head corresponding to the storage and then you have the tail water level that is H 2. So, the net head, we define as H 1 minus H 2. So, this is the net head available for power generation after accounting for losses and so on.

As we saw in the previous class, to generate a given power P associated with the head net head H t and the overall efficiency eta, we obtain R t in terms of million cubic meters, this is the release to be made from the reservoir in million cubic meters and this is the expression, where P is in Mega Watts; H t is in meters and this we are talking about for a monthly time period, you can refer to the last lecture, this is the expression that we obtain and then we introduce a iterative procedure, because the net head depends or the total head depends on the storage. So, H t when we are talking about, which is the net head during the time period t and therefore, it will depend on the storage at the beginning of the time period t S t, as well as storage at the end of time period t plus 1 during time period t.

Now, if you look at the continuity relationship here, the S t plus 1 is not known when we are doing simulation from one time period to time period, this is not known and the head here H t and therefore, the R t here, because R t depends on H t and also E t, which depends on the area of water spread, which is changing from the beginning of time period to end of the time period, because it depends on the storage as you can see here, all of these need be determined simultaneously and we introduced a iterative procedure for obtaining H t, R t, E t and S t plus 1; all of them simultaneously, in the simulation procedure that we saw last time. In, addition, we introduce the concept of primary as well as secondary power, recall that we also define the firm power and the secondary power both in the case of run of the reverse system, as well as in the case of reservoir systems.

So, the firm power is the power that is available with 100 percent reliability, this is in fact, the minimum power that can be generated for a given hydrology and at a particular site with the turban capacity fixed and so on. So, that is the minimum power the that can be produced, the secondary power is available with 50 percent reliability; that means, at least 50 percent of the time, the power gen generated will be greater than or equal to that particular power. The primary power is the power that you would like to generate all through the year, to the best extent possible that is you will set a target power. And then you would maintain particular power all through the year.

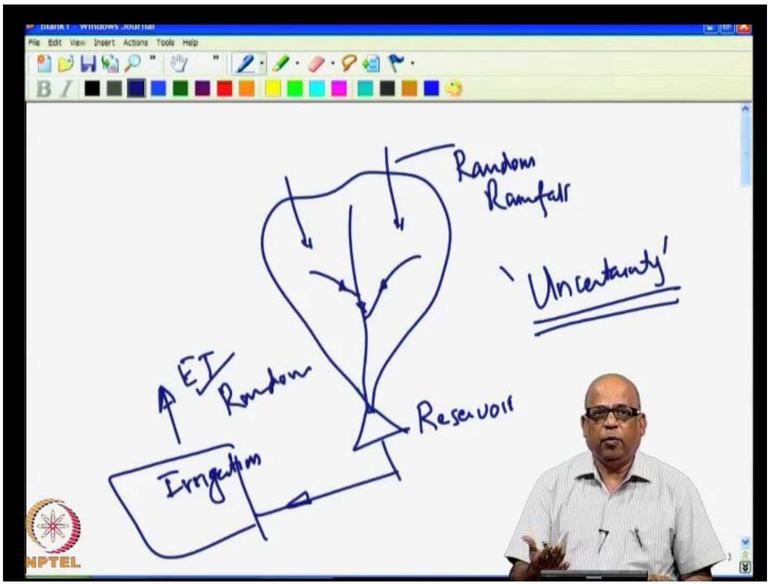
In the example that we studied in the previous lecture, we consider the primary power to be 73.5 Mega Watts. So, you would like to produce 73.5 Mega Watts constantly all through the year, whenever it is possible and when it is not possible, you will make the complete release when it is not possible, because of the water availability constraint, you will make the complete release and then produce whatever best power that you can produce, which will be less than the primary power.

Then the secondary power will correspond to the spills that are occurring; that means, the reservoir has reached its capacity and still you have excess water that has that is going out as spill in certain situations, you can use the spill water also for producing the power, at least in the planning paper, the planning stages you would like to see, whether you can actually utilize the spill at the maximum head that is possible corresponding to the capacity of the reservoir and produce the additional power.

So, you have the concept of primary power and additional power concept of firm power and secondary secondary power. Now, there will complete the deterministic modeling procedures, now we will see an important concept in water resource systems, which is that dealing with uncertainty associated with the hydrologic variables, primarily hydrologic variables there are also other uncertainty associated. Those things are typically covered in the advanced water resource systems courses.

As I have been mentioning you know, in all the procedures that we were dealing with the hydrologic variables primarily you look at the evapotranspiration or evaporation the stream flows. Now, these are the primary hydrologic variables that we use in making decisions, all these hydrologic variables are burdened with uncertainties associated with natural processes for example, you are looking at the stream flow at a particular location.

(Refer Slide Time: 08:55)



Let us say, you are looking at a catchment and then you have the reservoir at this point, this is the reservoir and it has its catchment and the stream flow here is governed by… this is the stream flow; is governed by the rainfall that is occurring in the catchment and the rainfall is typically, associated with randomness or I will say rainfall is a random variable and, because rainfall is random variable. The stream flow, which is depend on the rainfall, also becomes random variable and therefore, there is an uncertainty associated with the stream flow process.

Similarly, if you are looking at the evaporation at this point or let us say, you are taking the water and then you are applying for irrigation, this is the irrigation district or irrigation area. Now, in this there is evapotranspiration taking place, now evapotranspiration depends on straight away first the temperature, which is the random variable and then it depends on the soil moisture, it depends on the rainfall and so on. So, the evapotranspiration becomes random and evapotranspiration determines the demand that has to be met from the reservoir from time period to time period. So, most of the hydrologic variables that we deal with in water resources systems are burdened with uncertainties and the first level of uncertainty that we talk about is the uncertainty associated with randomness of the hydrologic variables.

So, when you are making decisions for the future, you need to account for the uncertainties associated with all the hydrologic processes for example, you were in our earlier classes, we determine the reservoir storage for a given sequence of inflows assuming that the same sequence will repeat year after year, it is well known that the same sequence will not repeat year after year; however, that may be certain probabilistic probabilistic characteristics of the stream flow sequence that may be preserved for example, the probability distribution may be preserved in the future.

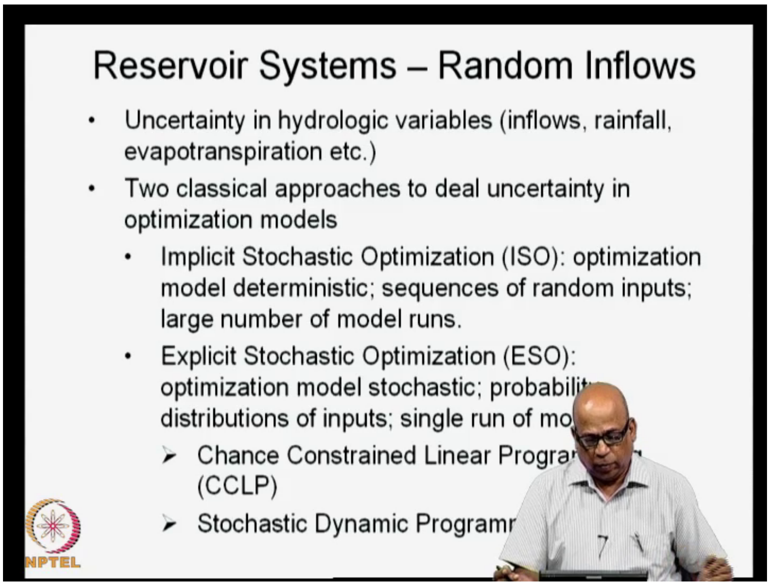
So, when you are making decisions based on the historical data, you may, you will warn to consider the uncertainties in terms of the probability distributions of this this various variables. So, you must have methodologies by which we make decisions in the phase of these uncertainties by incorporating the uncertainties and not by ignoring uncertainties. In the case of deterministic optimization and deterministic procedures that we developed earlier, we simply ignore the uncertainty, but uncertainty is the major component in any water resource systems and therefore in any water resource systems actually and therefore, we must have methodologies by through which we address.

(Refer Slide Time: 12:28)

So now onwards, we will now move on to methodologies incorporating the uncertainties. So, what we will be now talking about is the reservoir systems with random inflows. In this course, primarily we will consider uncertainties only associated with inflows, but I, as I mention the uncertainties are associated with most of the hydrologic variables that we deal with evapotranspiration is a uncertain, evaporation is the random variable, the soil moisture is random variable, the ground water levels are random, the ground water flow itself is random and so on. So, there are uncertainties associated with virtually any hydrologic variable that with deal with in water resource systems.

However the primary variable, which governs most of our decision is, in fact the inflow to the reservoir. So, we will start introducing, the concepts of uncertainties with respect to the inflows and specifically in this course, we will deal with reservoir systems, reservoir systems with random inflows.

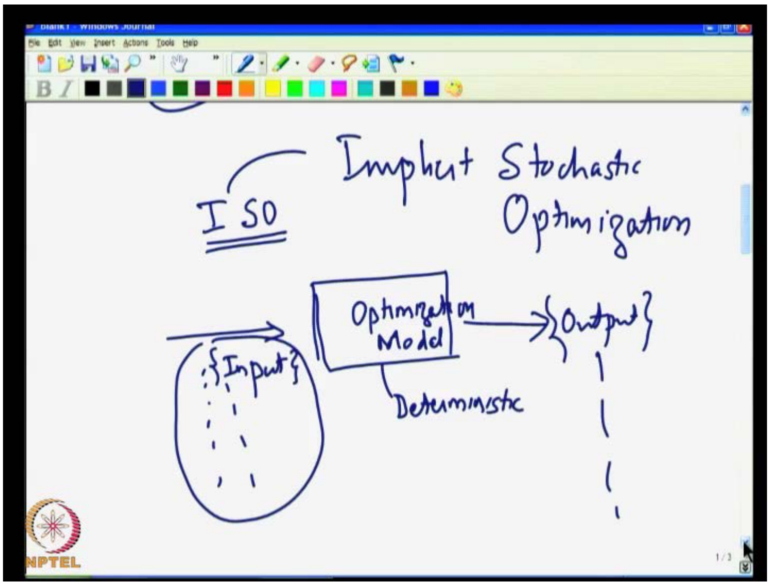
(Refer Slide Time: 13:34)



So, the uncertainty in hydrologic variables, leads us to the methods that we are going to develop now. We dealt with optimization techniques, to determine the size of the reservoir earlier, as well as to determine the operating policy of the reservoir. In which case, the sequence of stream flow was head constant as, R was treated as deterministic.

Now, because of the uncertainties associated with the stream flows now or the inflows, we start modifying the procedures that we have developed earlier for deterministic sequences to incorporate in some sense, the variations that are likely to occur in the stream flows or in the inflows. There are two primary optimization techniques that we deal with in such a situation, one is called as the Implicit Stochastic Optimizationand the other one is the Explicit Stochastic Optimization. Let me explain what we mean by that.

(Refer Slide Time: 15:00)



So, there are two ways in which we deal with the uncertainties in optimization. Let us say, you have the optimization model here.

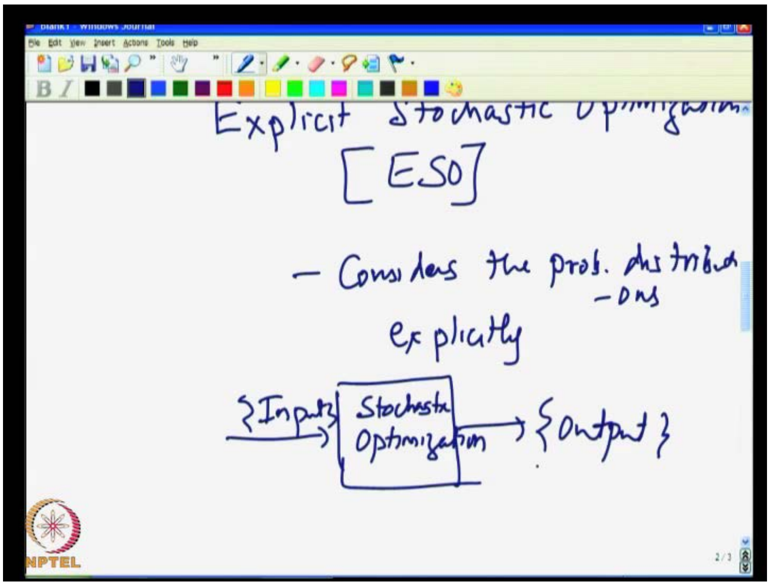
(No Audio From 15:05 To 15:14)

In the Implicit Stochastic Optimizationcalled, denoted as ISO Implicit Stochastic Optimization, the optimization model itself remains deterministic, but the sequences that are putting that we are using to run this optimization model. Let us say, you have the input sequence, these input sequences, we will run for several subsequences that is we take one inflow sequence.

Let us say, you have thirty years of flows and then you run this optimization model for that particular sequences, you get some output and change the input sequence, how do we change the input sequence or how we generate several sequences of flows and so on, is is slightly beyond this particular lecture, but in subsequent losses, I will discuss that also. So, there are ways of generating several sequences of inputs, inflows to the reservoir. So, run this optimization model, which is essentially the deterministic optimization, a deterministic optimization model for several sequences of inflows and therefore, we are considering the randomness associated with the stream flows, outside of the optimization model and therefore, it is a Implicit Stochastic Optimization.

Associated with each sequence of inflow input, you get one sequence of output. Let us say, you are talking about reservoir operating policy, you used yearly flows for that. So, one year flows you used and then got the operating policy, you change the sequence to make sure that the uncertainty associated with the inflows or incorporated, you change the sequence get another year inflows and then get another output and so on, or if you add a thirty year sequences, run this optimization model which remains constant, which remains constant in a deterministic framework, you run his for several sequences and therefore, you will get several sequences of output. And then you can do some analysis on the outputs that you produce to account for the uncertainties output. So, this is the way of addressing the uncertainty implicitly.

(Refer Slide Time: 18:26)



So, this is implicit stochastic optimization.

(No Audio From 18:04 To 18: 17)

But a more compact and more elegant mathematical, mathematically is the Explicit Stochastic Optimization

(No Audio From 18:27 To 18: 39)

Or the ESO, now in Explicit Stochastic Optimization, the optimization model itself has the probability distribution or the probability information or the uncertainty information associated with the input variables inbuilt into into it. So, the Explicit Stochastic

Optimization, considers the probability information, probability distribution, let us say, I will define, what are all the things presently distribution explicit. So, the probability model that we had earlier is itself a stochastic optimization model.

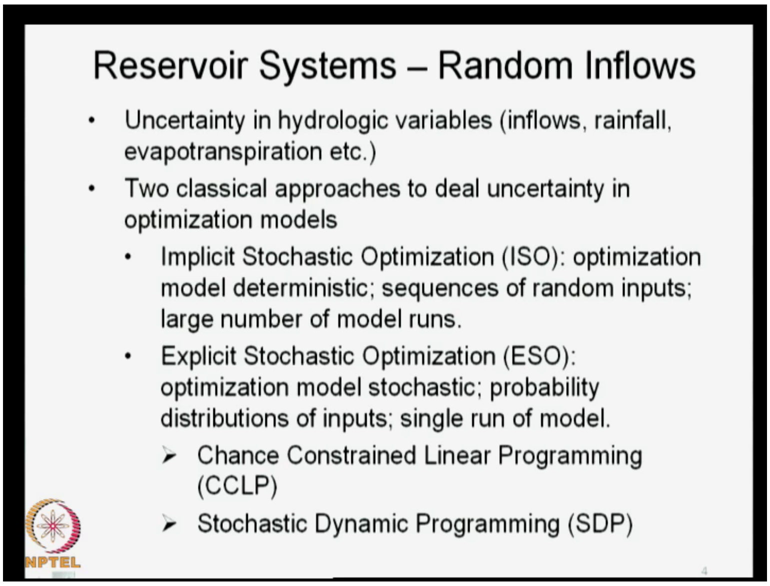
(No Audio From 19:32 To 19: 51)

And therefore, all the uncertainty information associated with the flows and all other hydrologic variables has been already incorporated in the model itself. You run the model only once and you get the output with the associated certainties defined. Output also we will have its one distribution. So, you run the stochastic optimization model only once unlike the Implicit Stochastic Optimization model, which you need to make for which, in which you need to make several runs of the optimization model, each associated with one sequence of inputs inputs. So, that is the primary difference.

So, in the Implicit Stochastic Optimization model, the model remains deterministic, you run the model which several sequences of inputs, all of which together address the uncertainties in inputs and therefore, associated with each of these sequence, you get one output sequence and therefore, you will get several sequences of outputs and then you do the analysis of the outputs to get the decisions incorporating the uncertainties.

Therefore, the optimization model is run several times whereas, in the Explicit Stochastic Optimization model, the optimization model itself considers the probability distributions and therefore, the optimization model is stochastic optimization model, you run it only once and then you get the associated decisions output with the uncertainty. So, you run it only once, input actually it is specified in terms of the probability distribution and so on. We will see that later. So, these are the two primary classes of the stochastic optimization.

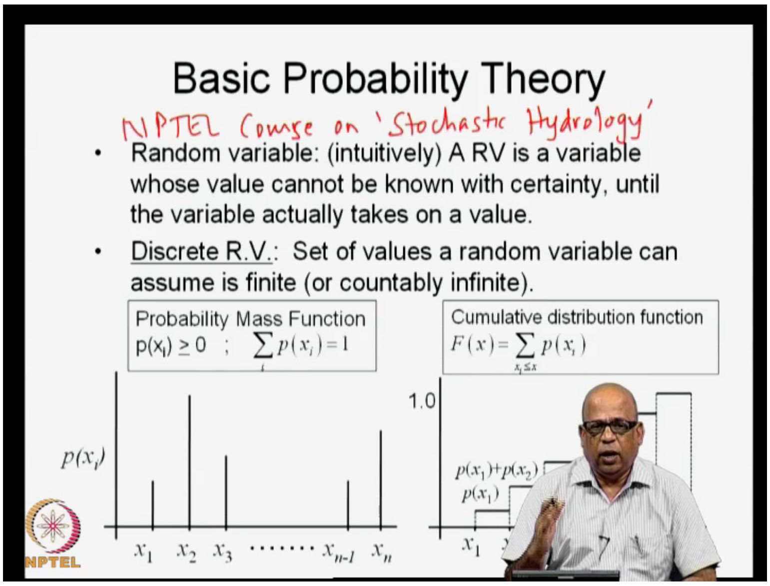
(Refer Slide Time: 22:02)



Now, in explicit optimization, we have two primary two major techniques that will be dealing with, one is the Chance Constrained Linear Programming is called as with this is denoted as (CCLP) Chance Constrained Linear Programming; another is Stochastic Dynamic Programming (SDP). The Implicit Stochastic Optimization, I I repeat does not need any special treatment now, because the optimization problems, itself remains the same as what we covered earlier in the deterministic optimization except that you run the optimization model with several sequences, how to generate the several sequences and so on. Perhaps will cover later or you can refer to other courses on stochastic hydrology to see, how you generate several sequences.

So, what we will do? now is that we will begin with some basic concepts of probability, I will quickly review the probability theory. So, that the students will get use to the concepts that will be using subsequently, I would encourage you to look at the NPTEL course on stochastic hydrology for this basic concepts of hydrology, basic concepts of probability theory.

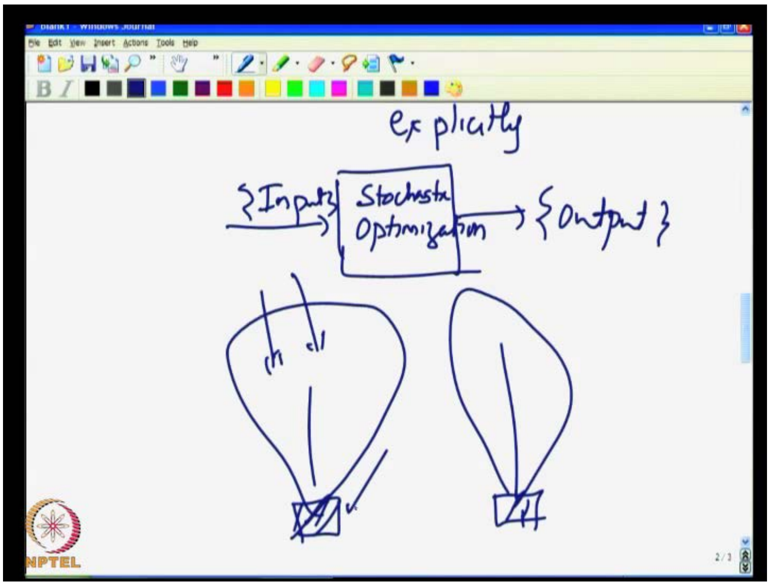
(Refer Slide Time: 23:23)



So, I will write here, this you can refer to NPTEL course on 'Stochastic Hydrology'. All the basic concepts have been covered in that particular course, but we will quickly review, what we need for this particular course.

So, there is a important concept of a random variable let us see, you look at the inflows, inflows are generated by precipitation typically, along with the other processes that take place in catchment, such as evaporation, such as depression storage there may be also infiltration that is taking place there is the interception that is take in place before that are starts and so on. So, there are a large number of hydrologic processes there are large number of natural processes that affect the stream flow generation of stream flow and all of these very randomly and that brings us to the concept of a random variable, you look at a newtonian type of equation where we are saying let us say, S is equal to U t plus half A t square, you specify the U which is the initial velocity initial velocity and the acceleration for a given time; the S which is the distance covered will remain always the same irrespective of at which place you apply this equation at what time you apply this equation and so on. Now, these all deterministic forms of model whereas, you look at the rainfall run off process itself

(Refer Slide Time: 25:31)

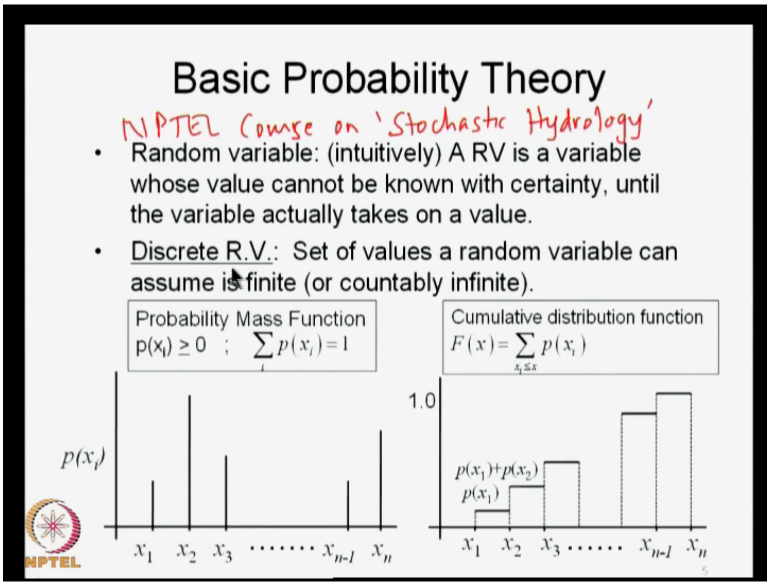


The rainfall run off process, for the same rainfall in a catchment here modelling let us say the stream flow, you are measuring the stream flow at this point for the same rainfall in the catchment, you may get different run off, during different times and if we if we have the same area of catchment at different location, you will get a different run off and so on. So, the rainfall run off relationship is a random relationship is a is a stochastic type of relationship, where there is a uncertainty associated with a run off associated with a given amount of rainfall.

So, the concept of random variable is is it is that particular variable in intuitively mathematical definitions and so on. You can refer to the stochastic hydrology course. So, intuitively what we mean by a random variable? it is that particular variable whose value cannot be predicted or whose value cannot be known in advance until it actually takes the value. What, I mean by that is again compare this with newtonian type of models S is equal to U t plus half A t square those plus 1, where you specify U, you specify t, you specify A, you can always get S, you can know S at priori whereas, the random variables you are looking at the run off at the particular location, even if you specify the rainfall, you will not be able to know certainly that this should be the rainfall, this will be the runoff presidely until it actually takes on the value. So, intuitively the random variable are those variables, whose value cannot be known in advance, until it actual whose value cannot be known until the value actually occurs.

This is the intuitive definition of course, there is a mathematical and more formal definitions of random variables and so on. You can refer to the stochastic hydrology course. Now, with this primary definition then we move on to slight mathematical treatment of this for which we define the probability distributions.

(Refer Slide Time: 27:54)



So, there is a concept of discrete random variable and continuous random variable. Now, the discrete random variables are those which assume only discrete values for example, the number of rainy days is a discrete random variables, it can only take on values such as 0, 1, 2, 3 etcetera, it cannot assume a value of 0.5 for example, if it rains it is a rainy day, if it does not rain it is not a rainy day. So, that is one of a example,

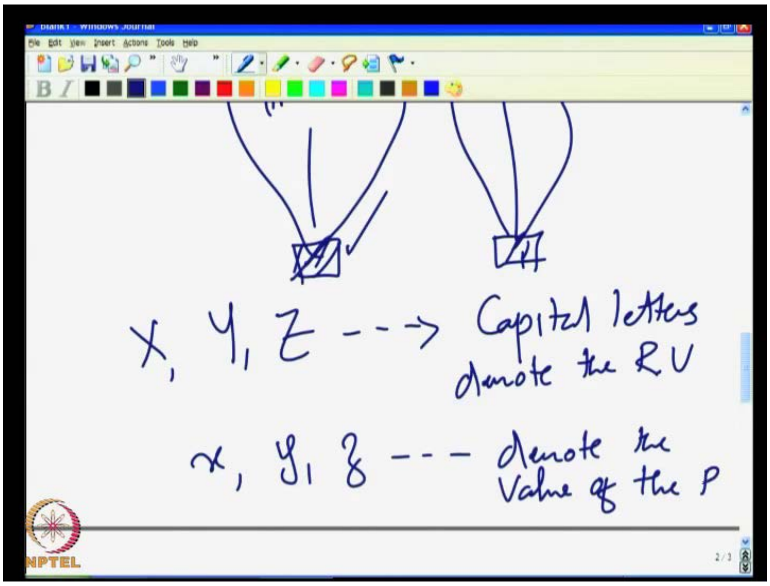
So, if a random variable takes on only discrete values, then it is called as a Discrete Random Variable. Now, associated with Discrete Random Variables, then we have the probe concept of probability mass function. So, let us say x is your random variable and it can assume x 1, x 2, x 3 etcetera x n; associated with each of these events that is probability of x being equal to x 1; associated with the event x is equal to x 1 there is the probability associated with the event x is equal to x 2 there is a probability and so on. And this defines, what is called as a probability mass function, for a discrete random variable, these probabilities are the non negative and the sum of all the probabilities over all the values that it can assume must be equal to 1, which means that exactly one of these values will is sure to occur.

Therefore, probability of x is equal to x 1 plus probability of x is equal to x 2 plus probability of x is equal to x 3 plus etcetera must be equal to 1, then we have the concept of Cumulative Distribution Function, it simply gives the cumulative probabilities for example, this is the probability of x is equal to x 1; probability of x is equal to x 2. So, this value here is probability of x is equal to x 1 plus probability of x is equal to x 2. So, this is the cumulative probability similarly, this value will give probability of x 1 plus probability of x 2 plus probability of x 3 and so on.

So, this is probability of x 1 plus probability of x 2 plus etcetera probability of x n, and the last value that you get here in the case of discrete distributions will be equal to 1. So, this is the concept of Cumulative Distribution Function and therefore, it is denoted as F of x, capital F of x is equal to summation over all x i less than or equal to x for this particular value of x probability of x i.

What what I mean by that is, let us say that you want probability x b less than or equal to a particular value x 4, you sum all the probabilities up to x 4 and that is what you get as probability of x being less than equal to x 4 and this is what is that particular value. In the case of discrete random variables, the mass function that we have thus defined gives the probabilities of discrete events, we now go to the continuous random variables and one more important point that you must remember is that we denote by capital letters, the random variables themselves and by small letters the value that the random variable takes.

(Refer Slide Time: 31:45)



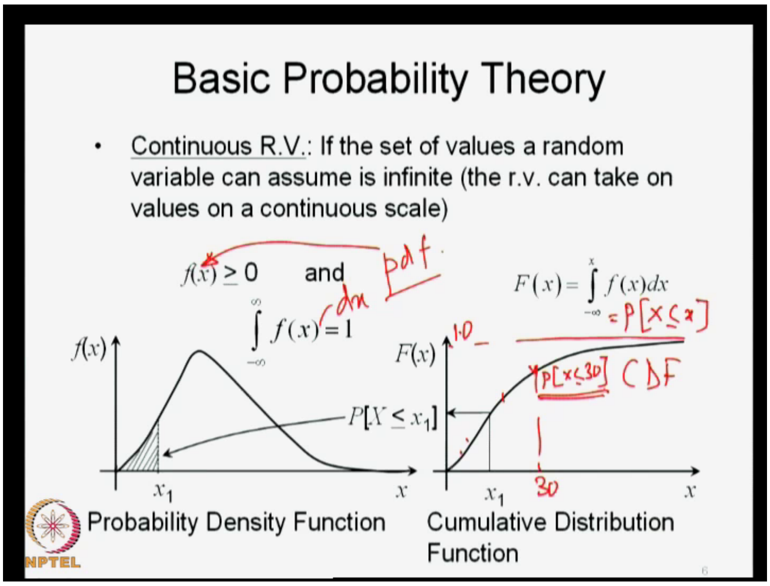
So, for example, X, Y, Z etcetera like this capital letters denote the random variable and the associated small letters x, y, z etcetera these denote the value of the random variable

(No Audio From 32:08 To 32:18)

For example, x may be stream flow, y may be rainfall and z may be evapotranspiration etcetera. Then we can say x is less than or equal to x that is stream flow is less than or equal to a given value of x. We may write x is less than equal to 30. So, this is small x denotes; that denotes the value that random variable takes. So, this is one convention that we use just keep this in mind, when we are covering the probability models. Then we move on to the continuous random variables, discrete random variables take on only discrete values 0,1,2,3 etcetera.

The continuous random variables, on the other hand take on a value along a continuum. For example, a line associated with between 30 and 40 let us say you are taking about talking about stream flow between 30 and 40, you have a line and it can assume any particular value along the line and therefore, 30.0235 is a possible value, 30.5 is a possible value etcetera unlike the discrete random variables. So, the continuous random variables assume infinite number of values whereas, the discrete random variables can assume only finite or countable infinite number of values, countably infinite in the sense, you may have 0, 1, 2, 3 etcetera, but it can go up to infinity, but they are simply they are still countably infinite number of values they are still discrete values.

(Refer Slide Time: 34:22)

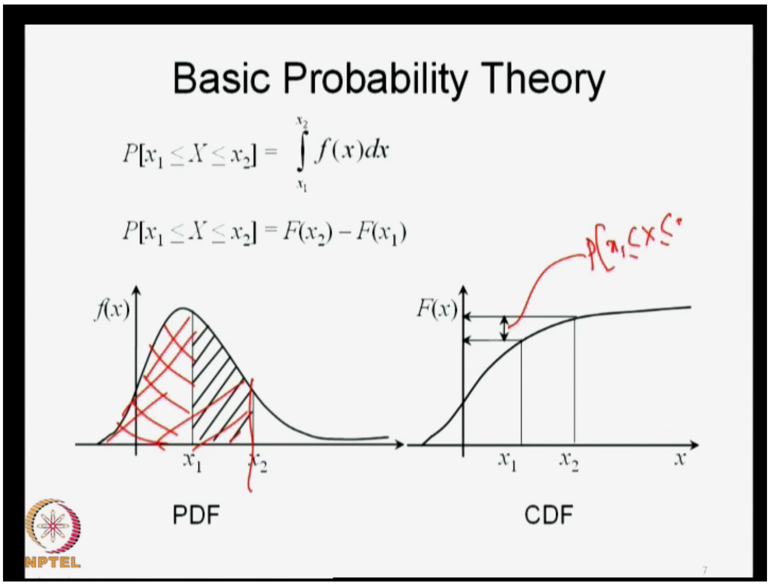


Now, we go to the continuous random variable much the same way as we define for the discrete random variable, we define now, what is called as the probability density function? So, any function which is non negative F of x greater than or equal to 0, which is non negative and the integral between minus infinity to plus infinity with respect to x is equal to 1. So, any function that satisfies this is a probability density function. So, we de denote this as PDF, Probability Density Function. Remember the probability mass function gave you directly the probability, but the Probability Density Function gives you the density and therefore, we can talk about the probability of x taking on a certain value has the area under the density, a density curve.

So, let us say, you are talking about this is the Probability Density Function, as an example and on the x axis is the variable, the values that the random variable can take and on the y axis is the pdf f of x, if you want the probability that x is less than or equal to x 1, just to relate to your hydrologic variables, this may be the distribution of flows and this may be a value of 30 let us say, and you are interested in what is the probability that the flow will be the less than or equal to 30. In such a situation, you just integrate the pdf up to that point and that will give you the probability of being less than or equal to x 1. The Cumulative Distribution Function, which is dented as the CDF. This is the CDF and this is the pdf, f of x is the pdf is simply minus infinity to x; that is f of x is equal to minus infinity to x integral f of x dx.

So, this gives you probability of x being less than or equal to x. So, look at this function now, this function every point here is obtained by integration of the pdf up to that particular point. So, that is what, we get as f of x and therefore, when you go to x is equal to infinity, what happens? you are talking then about minus integration of minus integration between minus infinity to plus infinity f of x dx which should be equal to 1. And therefore, the value approach is 1 or the maximum value x is equal to infinity will be equal to 1. Now from this CDF, we should be able to get the associated probabilities. Let us say that I have a value of 30 here, what does this value give? it gives the probability of x being less than or equal to 30.

(Refer Slide Time: 37:47)

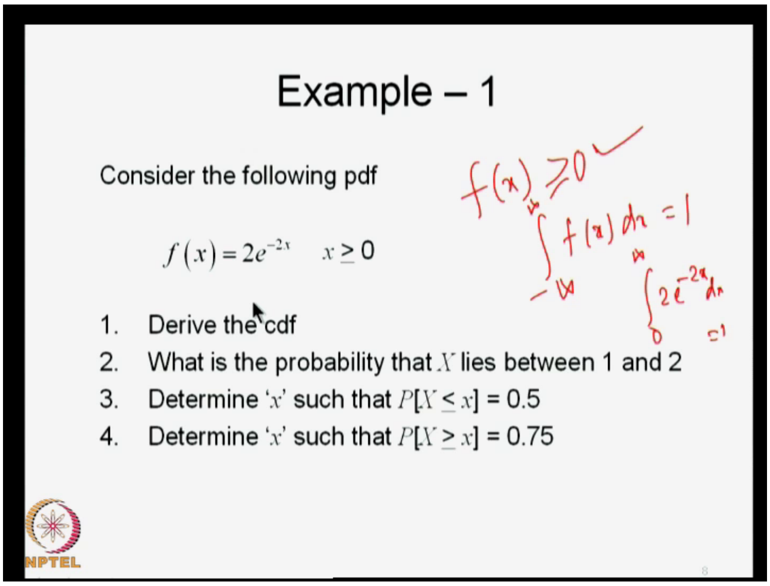


Now from this, it is obvious that probability that x takes on a value between x 1 and x 2. Let us say, you are talking about this region now, this is the probability density function f of x. So, to get any probabilities, you integrate the PDF over that particular region. So, probability that x takes on value between x 1 and x 2, you identify the region under PDF and then integrate that. So, that will between x 1 and x 2 f of x dx from this it follows that it will be equal to f of x 2 which is the area up to this point.

All of this area together is f of x 2 minus f of x 1, which is this area and that is how you get the particular area. So, probability of x taking on values between x 1 and x 2 is simply f of x 2 minus f of x 1 and this is in fact, this is f of x 2 here and this is f of x 1. So, this gives the probability that x takes on value between x 1 and x 2. So, that is the primary concept. So, in probabilistic analysis, you must remember the concept of probability mass function and the probability density function and of course, the Cumulative Distribution Function. So, both for discrete as well as continuous random variables, we have the concept of Cumulative Distribution Function.

In the optimization models that will be dealing with, we will be typically dealing with the continuous random variables, because for example, you you may talk about the stream flow probability distributions or the evapotranspiration probability distributions and so on. Now these are all continuous random variables. So, we will see some examples dealing with the continuous random variables.

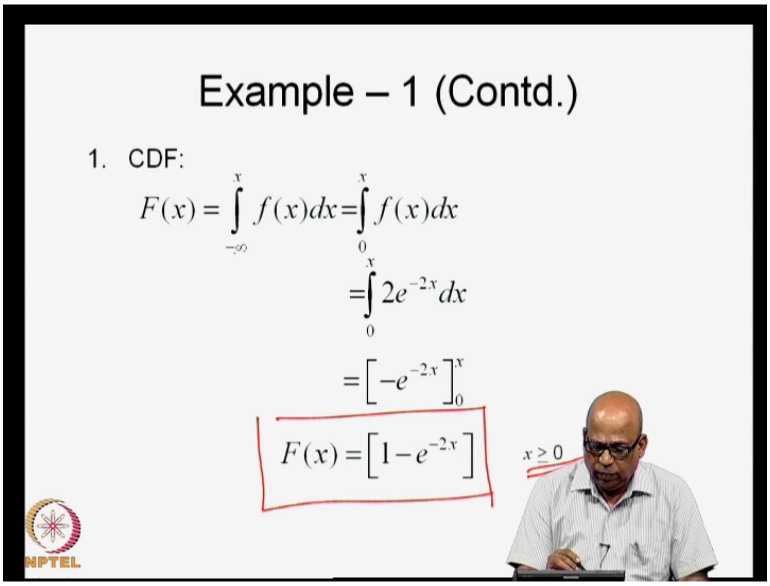
(Refer Slide Time: 39:46)



So, that the concepts are clearly understood. Let us say that you have a pdf defined by f of x is equal 2 e to the power minus 2 x for x greater than are equal to 0. For this now, we need to derive the cdf. First of all make sure that this the valid pdf to make sure, this is the valid pdf, what are two conditions that it has to you have to check, one is that it is nonnegative.

So, for x greater than or equal to 0, this is nonnegative that is a that is satisfied. The other one is that minus infinity to plus infinity f of x dx must be equal to 1. So, verify that between 0 and infinity to e to the power minus 2 x dx is in fact, equal to 1. And as we will see later on this is in fact, the exponential distribution that we are talking about. So, for this simple f of x which is the pdf, we will derive the cdf first and then we will look at the probability that x lies between 1 and 2. We will also determine x, such that probability of x being less than or equal to x 0.5. We will determine the value of x, such that probability of x being greater than or equal to x is equal to 0.75.

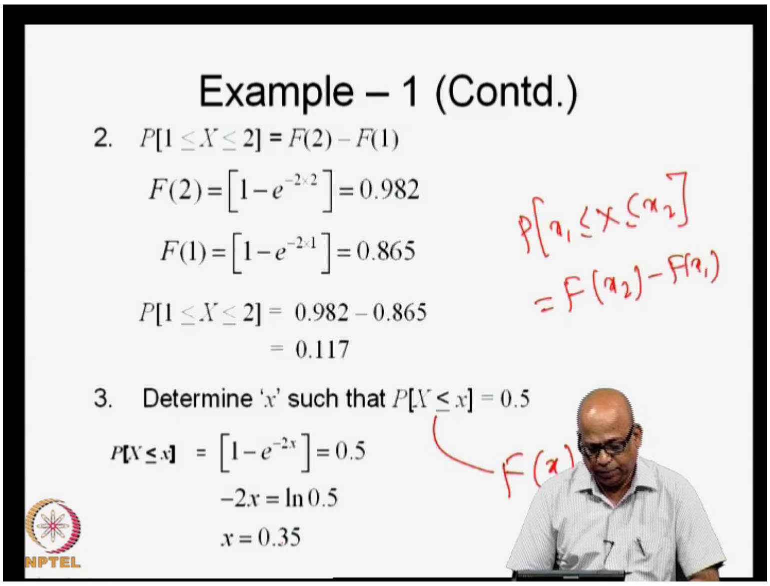
(Refer Slide Time: 41:13)



So, CDF by definition is recall capital F of x is equal to integral minus infinity to plus infinity I m sorry minus infinity to x f of x dx and in this case, it will be 0 to x, because minus infinity to 0 the f of x is 0, remember whenever we define it like this is equal to 0 elsewhere is understood always.

So, this f of x will have a value of 0 elsewhere and therefore, it turns out to be the same as integral between 0 to x f of x dx and it is a very simple integration, you can verify this ,you will get f of x is equal to 1 minus e to the power minus 2 x and this is defined for x greater than are equal to 0. So, this is the CDF, to get the CDF all you have to do is integrate the pdf between minus infinity to x

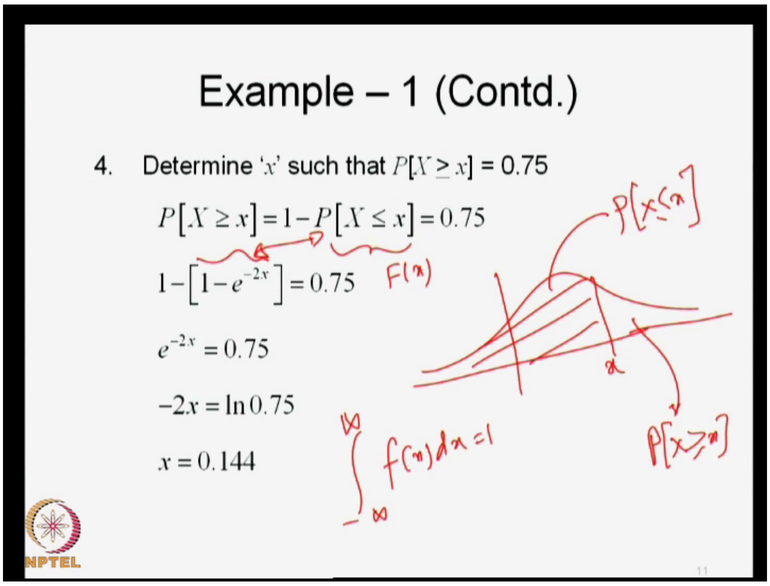
(Refer Slide Time: 42:15)



Then we want the of x taking probability that x lies between 1 and 2, recall that we just defined probability of x taking on values in a particular region between x 1 and x 2, it is simply F of x 2 minus F of x 1 and we have obtained the F of x already. So, in that we substitute x 2 and x 1 values. So, first we will determine F 2, which means x is equal 2 is what we are putting here and that also to be 0.982.

And F of 1, we put x is equal to 1 in this and that also to be 0.865 and therefore, the probability of X taking on values between 1 and 2 is simply F of 2 minus F of 1, which is 0.117. We look at this example now, we want that particular value such that probability of x being less than or equal to x is equal to 0.5. Now, such problem arise in hydrology, when we are talking about let us say, the low flows and we are interested in what is the probability that the flow will be less than or equal to a certain given value here. Especially, in water quality problems will be interested in probability of x being less than or equal to a given value the, where x is the stream flow. Now, the f of x is given. So, what we will do is, this is what probability of x being less than or equal to x is by definition simply f of x and f of x we know is 1 minus e to the power minus 2 x is equal. And therefore, we equate it 0.25 and from that you get x is equal to 0.35

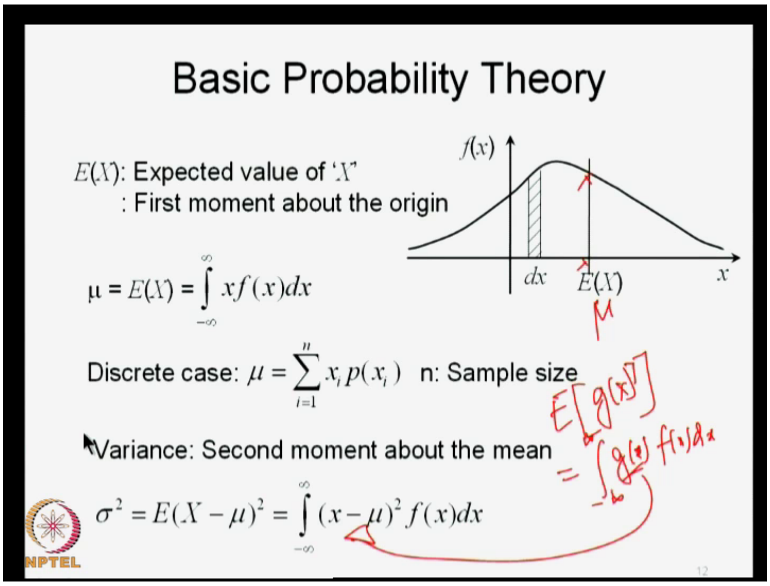
(Refer Slide Time: 44:14)



Then we determine x such that probability of x is greater than or equal to x is equal to 0.75. Now, these kind of problems arise, when we are talking about excidence probabilities typically, we use 75 percent exceeded values and so on. Where we are looking at what is the particular value of flow, which is exceeded 75 percent of the time. So, we are talking about probability of x being greater than or equal to x is equal to 0.75 we are interested in getting the value of x. So, we write this as probability of x being greater than or equal to x as 1 minus probability of x being less than or equal to x. For this, you just refer to stochastic hydrologic course, I will quickly explain this.

Let us say that this is the particular value of x, what does this area give? this area gives probability of x being less than or equal to x and what is this area? this area gives probability of x being greater than or equal to x. And the total area is 1, because minus infinity to plus infinity f of x dx is equal to 1 and therefore, the total area of the pdf is 1 and therefore, we write probability of x being greater than are equal to x as equal to 1, which is the total area minus probability of x being less than or equal to x and probability of x being less than or equal to x is nothing, but F of x. So, that F of x is what we write here, which is 1 minus e to the power minus 2 x. So, probability of x being greater than are equal to x is equal to 1 minus F of x, which is 1 minus 1 minus e to the power minus 2 x and that we equate it to 0.75, which is given and from that we get x is equal to 0. 144. So given the f of x, which is the pdf. We know now, how to calculate various probabilities associated with that.

(Refer Slide Time: 46:23)



Then there is a important concept of the excepted value. Now, all of these will be using subsequently in the optimization. So, I i will just go through quickly, the definition of these these various terms at will be using. Now, the excepted value is the first moment of the probability density function about the origin. So, you take the moment of the area about the origin, which is simply what we are doing minus infinity to plus infinity x f of x dx. So, this is the first moment of pdf about the origin and this is what we call as the excepted value and denote it as mean mu and it is variously called as the mean and the excepted value. In fact, when we have the sample estimate of this, we call it as simple arithmetic average we will come to that presently.

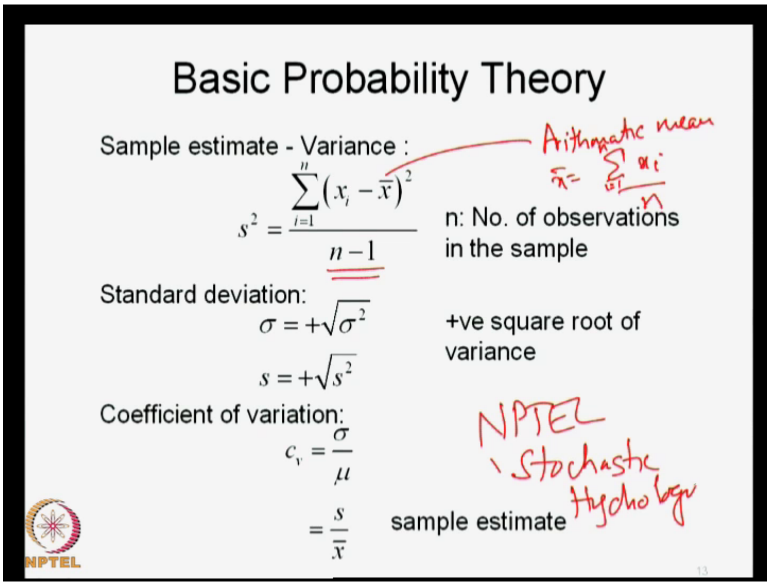
Now, in the discrete case, now this is for the continuous case. In the discrete case, mu is defined as i is equal 1 to n summation x i into p x i. So, x i is that particular value of the random variable corresponding to i is equal to 1; p x i is the probability that x is equal to x i; and i is equal to 1 to n; n is Sample size. If we are dealing with the discrete variables, then we have the concept of variance, which is simply the second movement of the pdf about the mean. So, first moment we took it about the origin like that you can take second moment, third moment etcetera also about the origin, but now onwards we start taking moments about the mean itself.

So, let us say that this is the excepted value of x or equal to mu and then we start taking moments about this excepted value itself. So, the second moment about the mean is in fact, the variance we denote it as sigma square and that is excepted value of x minus mu the whole square. Now, you first define the moment here first, so the second moment about the mean is integral minus infinity to plus infinity x minus mu the whole square we are talking about the second moment f of x dx.

So, this defines the variance and this we can also see that it is excepted value of x minus mu the whole square, there is a small concept you must know here, expected value of g of x is equal to minus infinity to plus infinity g of x f of x dx; that means, what we are saying is excepted value of a function defined on the random variable x is equal to minus infinity to plus infinity that function in to the pdf with respect to x the integration for example, you look at this excepted value of x, where the function of x is x itself. Therefore, this is x f of x dx. The excepted value of x minus mu the whole square, where the function is x minus mu the whole square is equal to g of x f of x dx.

So, the variance is written as excepted value of x minus mu the whole square. Now, these are called as the population moments, but typically what you will have is you will have a sample of observed values, the population is the set of all possible values of a random variable for example, you may have a thick book consisting of one thousand pages let us say, this and you are interested in picking up randomly certain pages. So, the page in a book is a random variable, the collection of all the pages, which is 1000 pages is the population and any subset of the population for example, you may you may consider pages in a chapter, in a particular chapter of that particular book, which is the subset of the population and that is the sample. Now in hydrology typically, what happen is that you will have observed data, now this observe data constitutes a sample and therefore, the moments that we just define for example, the excepted value the variance and so on. We need to determine for the samples and we now give the sample estimates of the particular moments.

(Refer Slide Time: 51:44)



So, the sample estimate for the variance is denoted as S square is equal to i is equal to 1 to n; n is number of observation in this sample; x i minus the x bar the whole square; x bar is the arithmetic mean and x bar is given by sigma x i over n over all i; i is equal to 1 to n. So, this is simply the average, arithmetic average divided by n minus 1. We use n minus 1 instead of n, because we are interested in what is called as a unbiased estimate. Now all these details, you can go to this NPTEL stochastic hydrologic course in in that all these concepts are clearly are elaborately discussed.

(No Audio From 52:42 To 52:50)

So, given the sample then you will be able to determine S square, which is the variance. The standard deviation is defined as the positive square root of the variance. So, sigma is equal to positive square root of sigma square and for the sample, we denote it as S as plus root of S square, then we have the concept of coefficient of variation, because both the mean mu and the standard deviation sigma will have the units of the original random variable itself and we want to have a unit less quantity to examine the variation and that is why we defined the coefficient of variation as sigma over mu or S over x bar, as you can see this will be a unit less quantity. So, we have the concept of expected value of x.

Now, we have the concept of probability density function and the probability distribution function, we have seen the concept of the mean and the standard deviation, as well as coefficient of variation. Now, with these background now, we should be able to consider the probabilistic information contained in the observed data explicitly in our optimization model, which means what, if you look at your reservoir operation problem inflow q was specified, now instead of specifying q as a deterministic value there, we must specify q in terms of the probability distribution of q in the optimization models and use the probabilistic information on the stream flow in the optimization model that is what we called as Explicit Stochastic Optimization.

Now, when we go to probability distributions there are a certain number of most commonly used probability distributions for example, you used normal distribution, lognormal distribution, gamma distribution and so on exponential distribution and so on. Depending on how well the data fits those particular distributions. Now, these common distributions will quickly revise or review in the next class, but I again repeat that for all these topics on probability theory, you can refer to the stochastic hydrology course on NPTEL where these have been discussed more in more detail. So, in today’s classes essentially we started with the important topic of reservoir operation or reservoir systems with random inflow.

So, this is the first time that we have started we have forehead in to the area of uncertainty, we are now talking about the hydrologic variables being random and not deterministic and we have started with the concept of the probability, we have started reviewing the probability concepts, but before that I just explained what we mean by Implicit Stochastic Optimization and Explicit Stochastic Optimization. In the Implicit Stochastic Optimization, you recall now that in the Implicit Stochastic Optimization, the model itself remains deterministic and this model you run over a large with a large number of inflow sequences and thus you address the uncertainty, because rather than running it with one sequence, you are running with several sequences all of which are equally lightly to occur and therefore, you will address the uncertainty associated with the inflows by considering several sequences.

And therefore, you generate several sequences of outflow output of the model. So, the model you do not touch, simply run it for different sequences whereas, in the Explicit Stochastic Optimization, the model itself considers explicitly the probabilistic information contain in the inflow sequence through the probability distributions and therefore, only one single run of the model is all that is necessary and then we defined what is the discrete random variable - what is the continuous random variable - what is the probability mass function - probability density function. We also looked at the Cumulative Distribution Functions and then examined, how we determined the probability associated with various events and towards the end of the class, I just introduce the concept of the excepted value of x and the variance.

Now, all of these concepts will be using when we going to Explicit Stochastic Optimization, there are two class of models that will be two classes of models that will be talking about one is called as a chance constraint linear programming; the other one is the stochastic dynamic programming. However in the next lecture, we will introduce the certain probability distribution functions that are commonly used in hydrology typically, we will start with normal distribution go on to lognormal distribution and exponential distribution. Thank you very much for your attention.