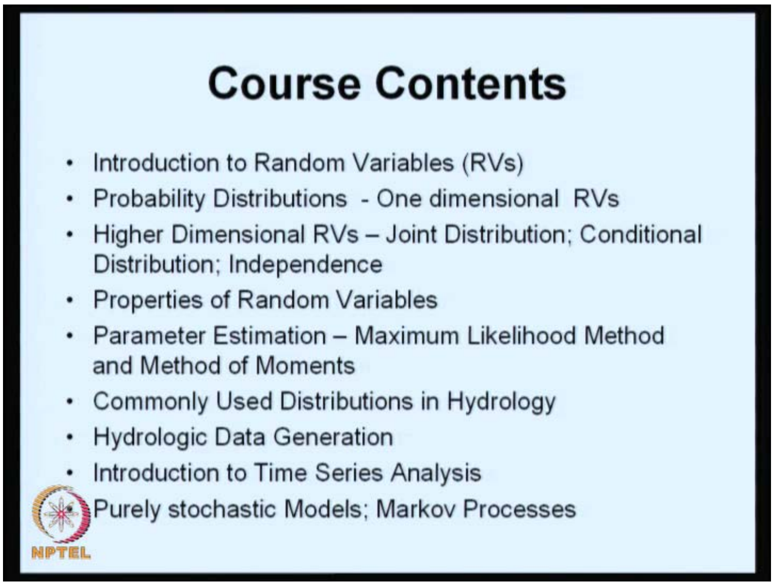
Stochastic Hydrology

Prof. P. P. Mujumdar
Department of Civil Engineering
Indian Institute of Science, Bangalore

Lecture No. # 01
Introduction

Good morning. Welcome to this course on stochastic hydrology; this is a lecture number 1. I take this opportunity to thank NPTEL for giving me this opportunity to teach this course. My name is Professor P. P. Mujumdar; I am at the department of civil engineering at the Indian Institute of Science; I work on hydrologic processes and water resource systems. So, this course, we will essentially introduce analysis of uncertainty in hydrologic processes.

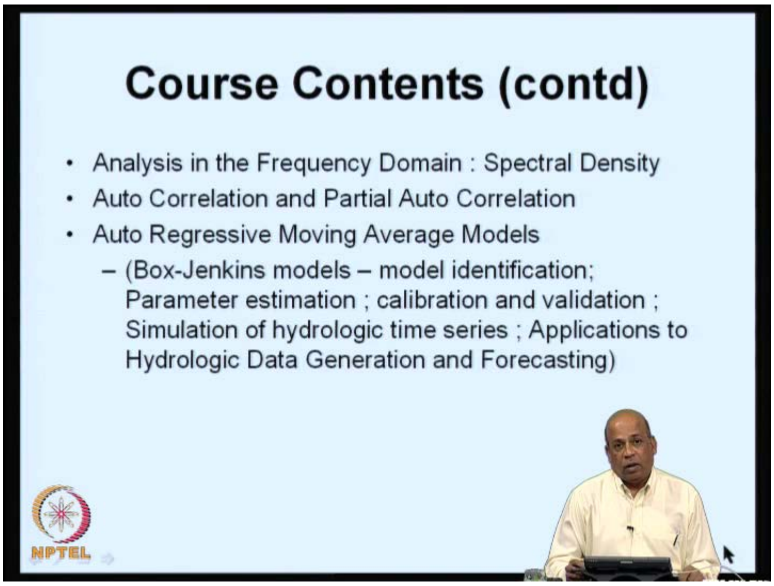
(Refer Slide Time: 00:44)



So these are the course contents. We will first introduce the concept of random variables. Then look at the one-dimensional random variables associated probability distributions. Then we will move on to higher dimensional random variables, where we talk about the joint distributions, conditional distributions and independence of random variables.

We will cover properties of random variables including the various moments of the distributions, and then parameter estimations using maximum likelihood method and method of moments. Then we will cover some commonly used distributions in hydrology; for example, the Gaussian distribution, log normal distribution, exponential and the extreme value distributions such as the Gumbel distribution. We will then see some specific applications for hydrologic data generation, and then we move on to time series analysis just a brief introduction to the time series analysis, in which we will also cover purely stochastic models and Markov process.

(Refer Slide Time: 02:09)

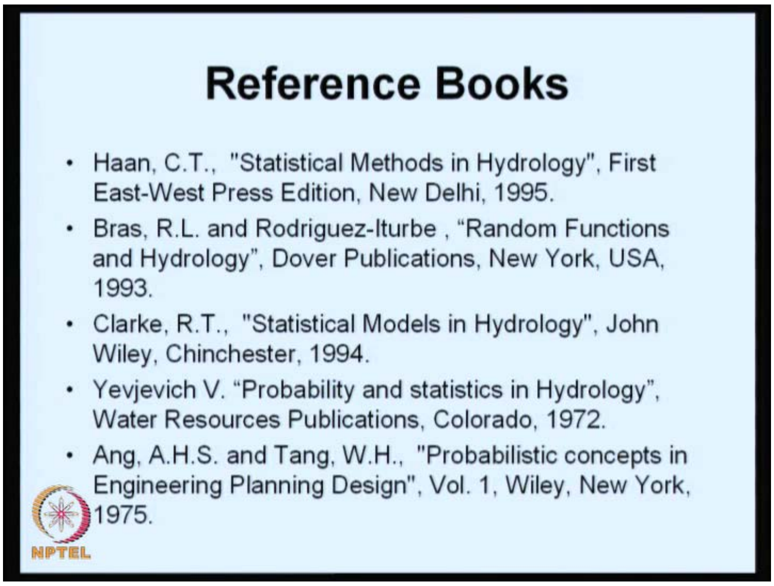


Then, we will introduce analysis in the frequency domain specifically the spectral analysis using spectral density and auto correlation and partial auto correlation. We will cover in this course some specific time series models called the ARMA and ARIMA models Auto Regressive Moving Average models and Auto Regressive Integrated Moving Average models. In this, we will introduce the Box-Jenkins models specifically covering methodologies for model identification, parameter estimation, calibration and validation, simulation of hydrologic time series as examples and applications to hydrologic data generation and forecasting.

As I mentioned, the broad aim of this course is to introduce to the students the various uncertainty concepts that are used in hydrology. So, I would expect that the students would have a background on physical hydrology. So the various physical processes that are involved in hydrology I expect the students to know and it is desirable, but not essential that the students also have some you know preliminary background on probability and statistics; especially, the set theory and so on which we will assume that the students have enough background.

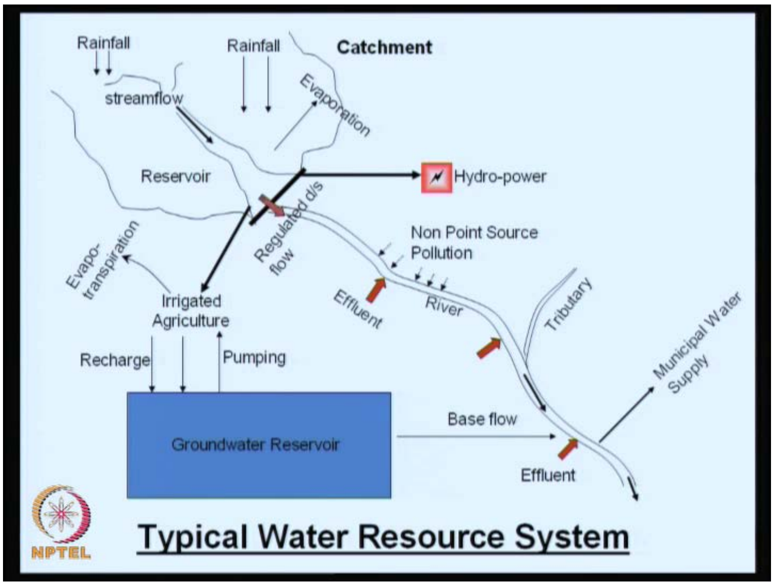
In this particular lecture the first lecture, we will introduce the various components of water resource systems where uncertainties exist, and the importance of stochastic hydrology in making decisions both design decisions as well as operational decisions in water resources systems.

(Refer Slide Time: 04:20)



So, we will start with the reference books that may be of use to the students. We will be specifically using the classical C. T. Haan book statistical methods in hydrology, then Bras and Iturbe random functions and hydrology, Clarke statistical models in hydrology. Then, Yevjevich probability and statistics in hydrology and Ang and Tang probabilistic concepts in engineering planning and design.

(Refer Slide Time: 04:49)



We will look at typical water resource system where there is a continuous interaction among large number of hydrologic variables; for example, you are seeing here a reservoir; this has a catchment its own catchment and in the catchment the rainfall that occurs within this catchment contributes to the stream flow and the stream flow adds to the storage in the reservoir. Then, there is a regulated flow from the reservoir which goes as downstream flow in the stream. At the reservoir itself the water is being used for irrigated agriculture and from the irrigated area there is a evapo-transpiration that is taking place. From the reservoir storage there is a evaporation that is taking place apart from other loses such as infiltration, seep agents such other things.

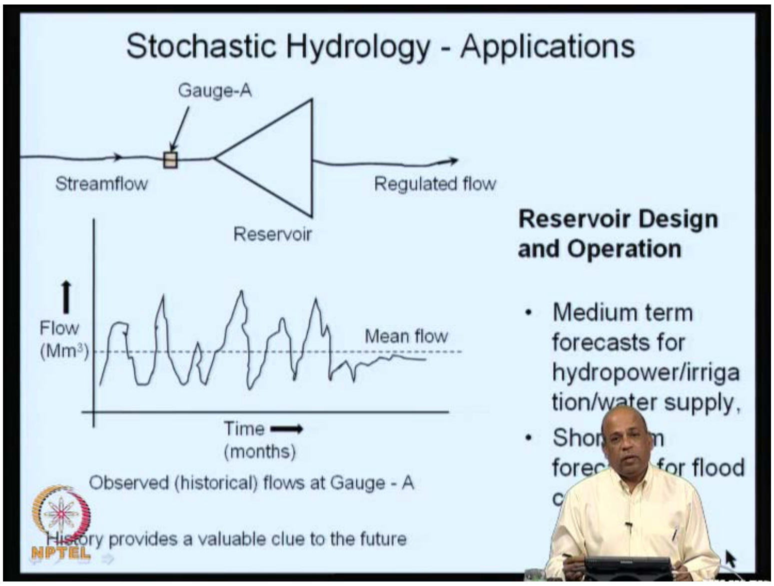
We are using the water also for hydro power generation and we are using this storage also for flood control. If there is a flood we use this storage as a flood absorption storage. Now, at the downstream of the reservoir, because this is a regulated flow, you may have non point source pollution from the agriculture run of typically which the run of picking up fertilizers and pesticides etcetera and then adding on to the stream here. Typically, the non point source pollution is an uncontrollable source. In addition, you may also have industrial and municipal effluents joining the stream and there may be some tributaries that add to the flows of the stream here. We may want use the water for municipal and industrial water supply.

Then, there is also a ground water reservoir and there is a continuous interaction between the ground water reservoir and the stream flow surface stream flow. The ground water also contributes in general as base flow to the stream flow here. And there is a recharge that is occurring from the irrigated area as well as from directly from the rainfall. And from the ground water storage we are also supplying water through pumping to irrigate agriculture. Now, this is a typical water resource system where there is a interaction among large number of hydrologic variables. As you can appreciate, the rainfall in the catchment area is governed by some random variations and therefore, the stream flow that joins the reservoir that contributes to the inflow of the reservoir also becomes a random process random variable.

And then, there are other natural processes such as evaporation, evapo-transpiration, the recharge and then the non point source pollution which is governed by the overland flow. All these add to uncertainties in the water resource systems and we need to make several decisions on such water resource systems including; for example, reservoir design including the magnitudes of the floods that are expected to come and the water quality transport in the stream given that it is governed by natural processes such as those stream flow and anthropogenic interventions through effluent discharges and so on.

Now, because of these uncertainties we need to evolve methodologies through which we address these uncertainties and make several decisions on water resource systems in the phase of these uncertainties.

Refer Slide Time: 09:28)



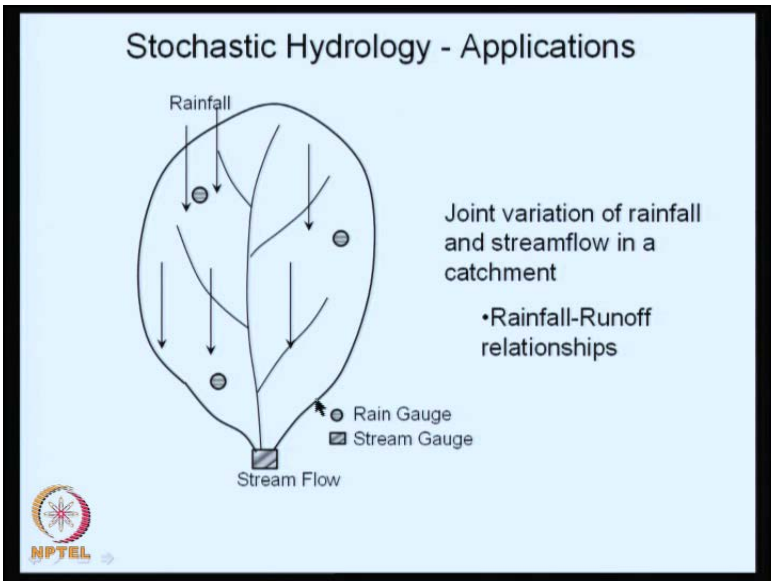
And the basic aim of this course on the stochastic hydrology is to introduce such methodologies that are useful in addressing uncertainties in the hydrologic processes. So, we will just go through some of the typical applications so that the students will know the usefulness of this course in actual applications. Say for example, you have a stream on which you would like to build a reservoir. The first question that we ask is how much of storage is necessary from the reservoir or at the reservoir to meet certain demand patterns?

Now, the only basis we have for this decision to be made is the historical flows that we have observed. Now, this shows the time series of the historical flows. There may be a gauge site here at which every month we have observed the flows and then we plot this and we get the time series plot of the flows. A basic premise in most of the methods that we use in water resource system decisions and also in addressing uncertainties is that the history provides a valuable clue to the future. So, we base all our decisions and all our uncertainty analysis primarily on the observed data which is typically historical data that we have on various variables such as rainfall, stream flow, etcetera.

Now, with this given data then we start addressing uncertainties for making decisions for the future, because when we build the reservoir, the reservoir is supposed to serve us for next about 50 years, 100 years and so on. So, we use the historical data to make decisions about the future and that is where we start addressing uncertainties associated with the future flows associated with the future hydrologic processes and so on. Now, within this broad framework of problems, we may be interested in medium term forecast for hydropower irrigation water supply and so on; for example, let us say we want to operate this reservoir over a 10 day period, 15 days periods and so on.

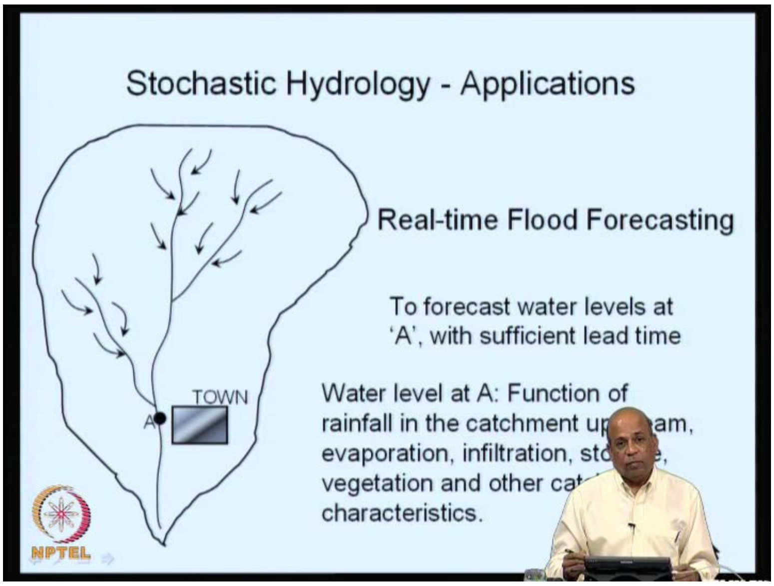
So, we would be interested in getting what is a likely flow to this reservoir for the next 15 days so that we regulate the storage. We would also be interested in short term flood forecast especially during the flood season so that we can use the reservoir storage as a buffer to accommodate that kind of storage that kind of a flood volumes. So, these type of problems will require probabilistic analysis which we will cover in this particular course. Then, we may have another classical problem where we are interested in getting the relationship between the rainfall in the catchment area and the stream flow at a particular location.

(Refer Slide Time: 12:39)



So specifically, we may be interested in obtaining rainfall runoff relationships so that we can use the rainfall data to predict the stream flow at a particular site. Now, this is a classical hydrologic application where we use methods of probability probabilistic analysis to generate such relationships. Then, we have a important problem where these methods would be useful. We may be interested in real time flood forecasting.

Refer Slide Time: 13:10)

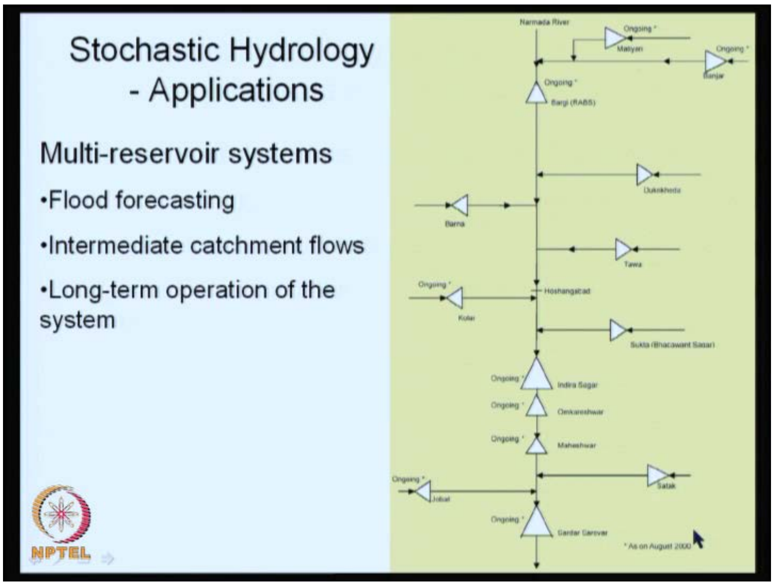


Let us say there is a town here which is located just adjacent to the stream and then, we know that high intensity rainfall is occurring somewhere upstream of this town, and then the rainfall converts itself into flood hydrographs, and then moves along the streams. So, we must be able to predict the water levels at this particular point which has implications on flooding of this town, and we should be able to do this with large time lags or sufficient lead times so that we will be able to take actions here.

So, if there is a high intensity rainfall that has started upstream of this, we must be able to provide what is the likely water levels rise in water levels at this point over the next let us say 6 hours, 10 hours and so on. Now, as you can see the water level at this particular point is a function of several catchment characteristics; for example, what kind of vegetation you have, what kind of rainfall intensity that is happening at these locations and what is the drainage pattern that you have, and the infiltration the depression and retention storage that you may have in the catchment and so on.

So there are a large number of physical factors that govern this. But most of these physical factors also are governed by other process; for example, if we are talking about infiltration and depression storage, evaporation, etcetera, they are also governed by some random fluctuations and therefore, the water level at this point becomes a becomes a random variable and we must evolve methodologies to relate what is happening upstream of this particular point to the water levels at these locations.

(Refer Slide Time: 15:29)

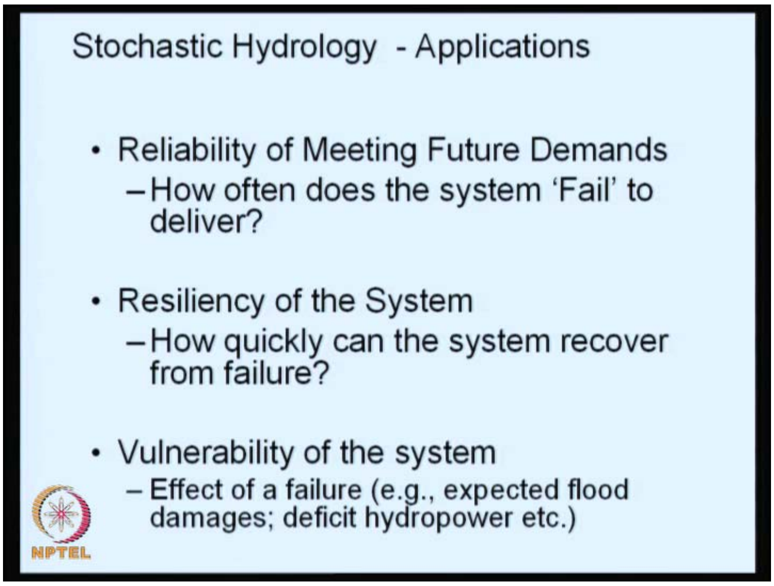


And the methods that we cover in this course will be of so far such applications. Then, we also have applications related with multi-reservoir systems; this shows a typical multi-reservoir system. In fact, this is the classical Narmada system that we have. There may be large number of reservoirs; for example, Bargi is located here; Indira Sagar is located here and Sardar Sarovar is located here and then on the tributaries there are also small and medium reservoirs as shown. Now, in these kind of large systems one of the important problems is that of flood forecasting; for example, we may have a important town at this location and there is large amount of intermediate catchment. So this intermediate catchment contributes to the flows at these locations.

So, we must have mechanisms by which we relate the catchment processes with the water levels at this location using methods of stochastic hydrology. Then, we also look at the long term operation of this system, entire system so that the system becomes sustainable in some sense in terms of its ability to meet the demands; in terms of its ability to retain the environmental integrity hydrologic integrity and so on. So, we are essentially looking at how the system is likely to perform in future. Given that it has certain inputs observed inputs over the past. Let us say we have observed data at observed flow data at several of these locations for the last about 50 years. Then, we how we use this last 50 years data to make decisions about the future is the scope of this particular course.

In such situations we ask the questions, how reliable is the system? That is we have the observed data and then we are making predictions on, how reliable will be the system in terms of its ability to meet the demands. It is not just the demands in terms of the quantity of water supplied, but also in terms of the quality of the water that is available. So, the first question that we would ask is how often does the system fail to deliver?

(Refer Slide Time: 18:00)

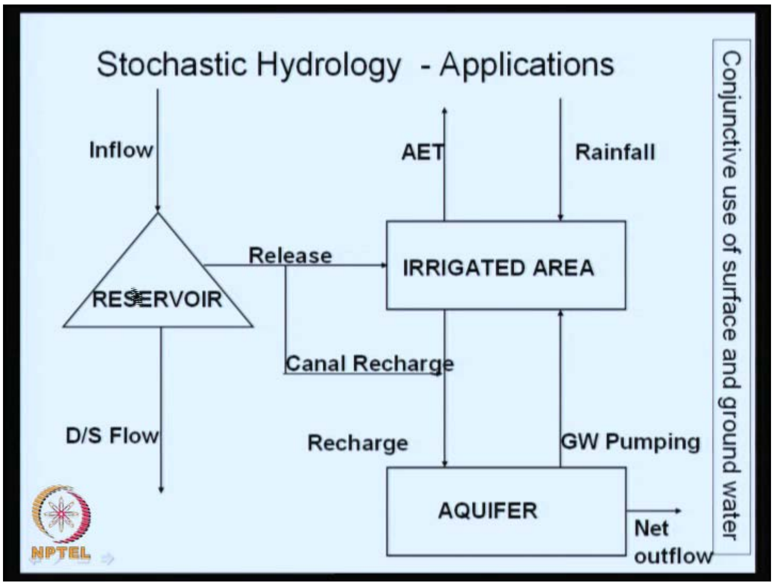


So, it is a measure of the failure of the system. In fact, a reliability and risk are related with each other. Then, how resilient is the system in terms of its ability to recover from failure once the failure occurs. So, we will be looking at questions such as, how quickly can the system recover from failure? Once it is known that the system has gone into a failure system, a failure state. Then, what is the effect of the failure itself?

So, we address the vulnerability of a system. Let us say that a failure occurs in terms of let us say its inability to protect against the certain magnitudes of floods, then what are the implications, what are the physical losses you suffer because of such a failure; or you are interested in hydro power production at a particular reservoir, and because of the flows being not adequate or the head being not adequate, because of low storages; the system fails to deliver, and then what is the deficit hydropower that you have at that particular occasion and so on.

So, the reliability concerns itself with the ability to meet the demands. The resiliency concerns itself with the capacity of the system to recover from failure once the failure occurs, and the vulnerability lead deals itself with the effect of the failure once the failure occurs. Now, these questions, these important questions we should be able to answer using the methodologies that we develop in this particular course. Then, another classical problem especially in drought prone countries like ours is the one related with conjunctive use of surface and ground water where you have a surface reservoir here.

Refer Slide Time: 19:54)



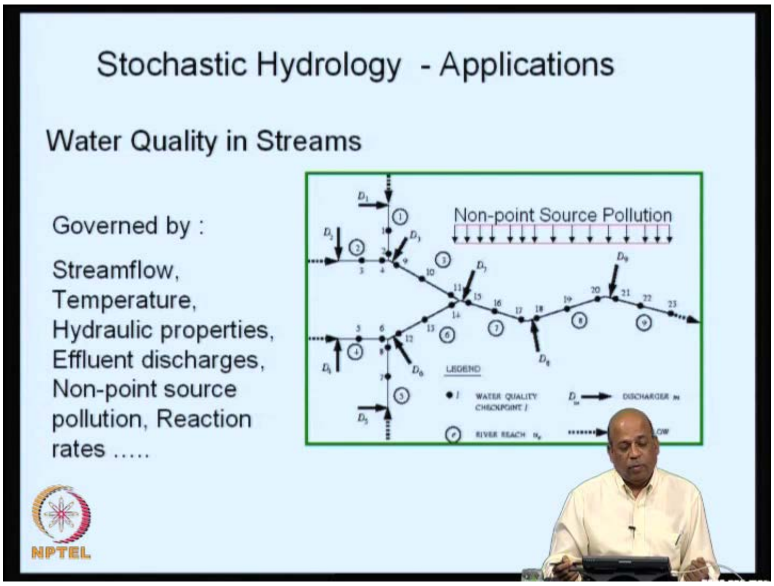
And then, you also have a ground water aquifer. You would like to meet the demands at an irrigated area, the agricultural demands at an irrigated area by using water from surface source as well as from the ground water source.

Now, the ground water source itself is governed by the aquifer characteristics and there is also a recharge that is taking place from the canals as well as from the rain fall and there is also a recharge that is taking place from the water that is applied to the irrigated area. At the irrigated area you have the actual evapor-transpiration that is taking place. Then, at irrigated are you also have the rainfall. Similarly, at the surface reservoir this inflow is governed by the rain fall that is happening in the catchment. Then, there is a reignited flow that is goes downstream and so on.

Now, this entire system when we want to analyze and arrive at decisions as you can see there are large number of uncertainties associated with several variables; for example, the inflow itself is governed by a rain fall in the catchment area which is on certain phenomenon and then the actual evapor-transpiration depends on large number of factors; for example, it depends on the soil moisture, it depends on the type of crop, it depends on the timing of crop growth stage and so on and so forth. Similarly, there is a rain fall that is occurring in the command area or the irrigated area which itself is uncertain and then the the recharge is drawn by a large number of uncertain parameters. The ground water flow is governed by large number of uncertain parameters and so on.

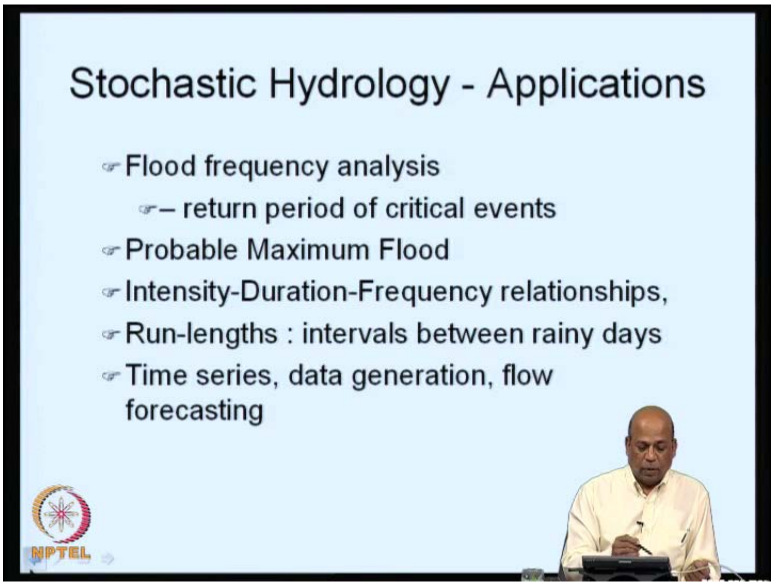
So, when you want to make decisions one such large system, you need to address uncertainties associated with all these variables.

(Refer Slide Time: 21:57)



Then, another classical problem that we deal with which requires analysis of uncertainty is the one dealing with water quality in streams. You look at this general example. There is a river here and then these are tributaries, and you have industrial and municipal effluents joining the stream at various locations. In addition, you also have non-point source pollution joining the stream, because of over land flow, because of sediment deposition sediment deposition into the stream and so on. As you can see, the water quality at any particular point; let us say you are interested in getting or using the water at this particular location for municipal and industrial use. So you would like to maintain a certain amount of water quality at this location.

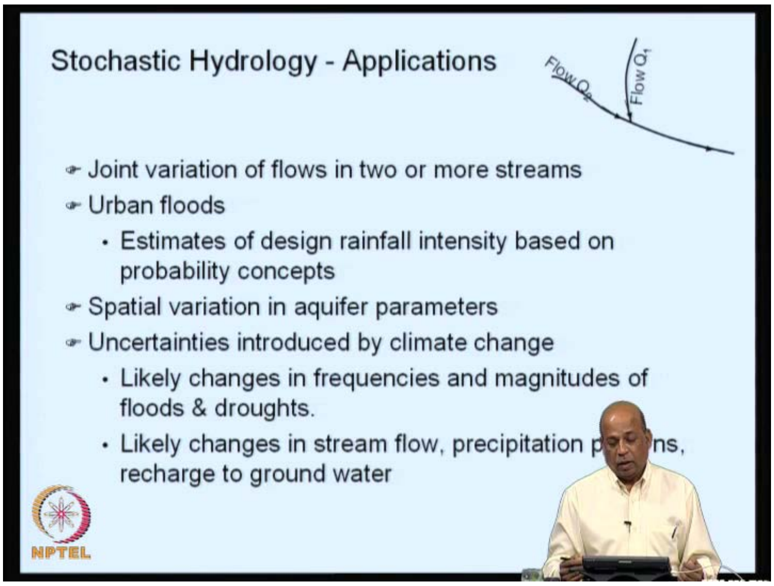
(Refer Slide Time: 22:54)



But this is governed by the water quality at this location is governed by the stream flow that is coming from upstream location and it is also it depends on the temperature. It depends on the hydraulic properties; for example, what is the time of flow between this point to this point, and the effluent discharges that are coming here and the non point source pollution and so on. In addition, there are also reaction rates; for example, the dissolved oxygen re-aeration rates and decomposition rates and so on.

Now, all these introduce uncertainties in decision making for water quality in streams. Then, there are a large number of other problems that we will be interested in; for example, flood frequency analysis, we will be interested in a return period of critical events then, we will also be interested in estimating the probable maximum flood especially when we are designing the reservoirs. Then, intensity duration frequency relationships for making decisions on flood protection measures then run lengths. Typically, we will be interested in intervals between rainy rainy days.

(Refer Slide Time: 24:25)

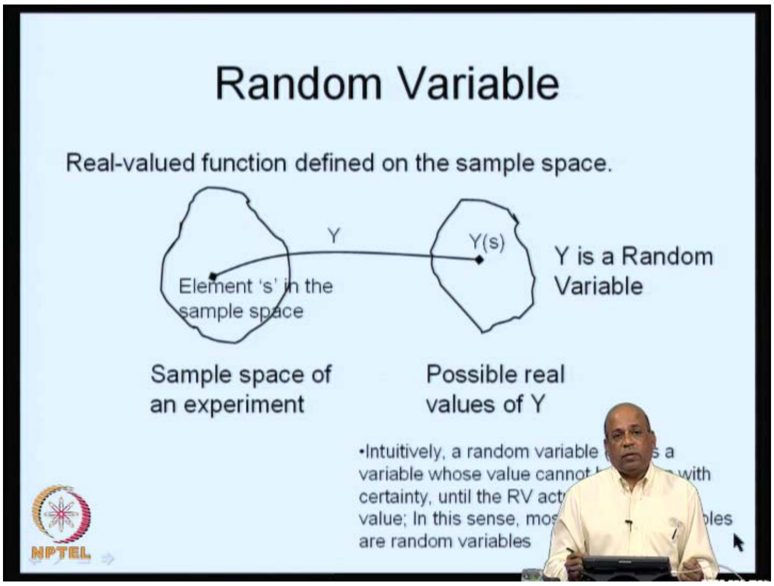


Then, time series, data generation, flow forecasting, all these are applications of stochastic hydrology. Then, you are also interested in joint variation of flows in two streams; let us say there is tributary here which is contributing to the main stream and for many decisions that we would like to make on this stream. We will be interested in looking at the joint variation between the flow q 1 and the flow q 2 here.

And when we come to the urban flooding; typically, we will be interested in getting the estimates of designed rain fall intensity and again we will have to do analysis of the historical data, and we use the probability concepts to get the design rain fall intensity and the spatial variation in aquifer parameters. This introduces a significant source of uncertainty when we are making decisions on decisions related to ground water. So, we must have a ways and means of addressing these uncertainties to make decisions on ground water ground water utilization. And the recent of interest in climate change also brings along the associated uncertainties; for example, we will be interested in looking at in the context of climate change, how the flows in a particular stream are likely to change or they going to increase or they are going to decrease in the event of climate change and likely changes in frequencies and magnitudes of floods and droughts.

Now, when we address questions such as these, we need to look at the uncertainties and the methodologies that are available to address these uncertainties. And the the primary core primary purpose of this course is to introduce the methodologies available for handling such uncertainties.

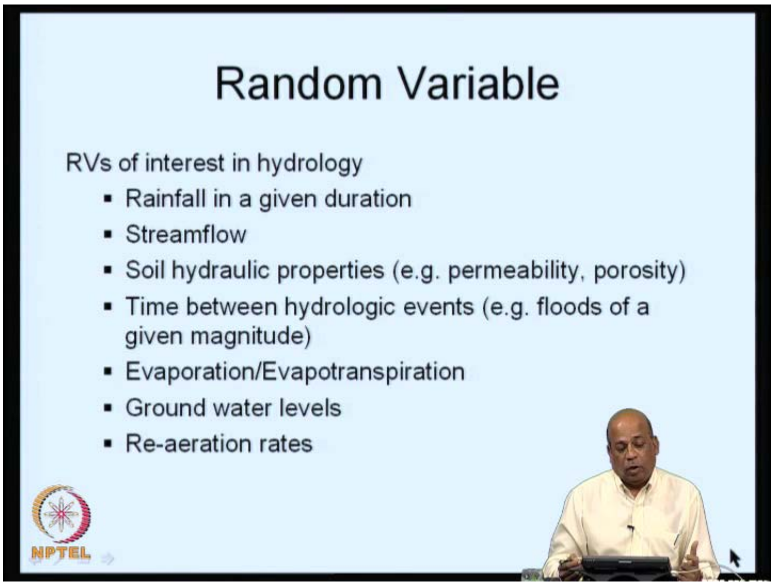
(Refer Slide Time: 26:33)



So, we start with the definition of a random variable. Now, intuitively random variable is the variable whose value cannot be known in advance until it actually takes on a particular value. So, we associate probabilities to the specific values that a random variable takes. But mathematically, we define more formally, we define the random variable as a real valued function defined on the sample space.

For example we may have an experiment in which we you have tossed a coin so the possible outcomes of such an experiment, are head or tail. Now therefore, this sample space may consist of h and t head h for head and t for tail. Now, this we map it on to a map it through a real valued function and then let us say we denote h by 1 and t by 0. So y is then y becomes a random variable; y is the function that relates every element in the sample space to a real value, so this is a formal definition. However, we will not be really concerned with this formal definition as we go into the hydrology, because most of the variables that we will be dealing with are in fact real valued variables directly and this is a called as the range space typically. So, from the sample space, we map it on to the range space.

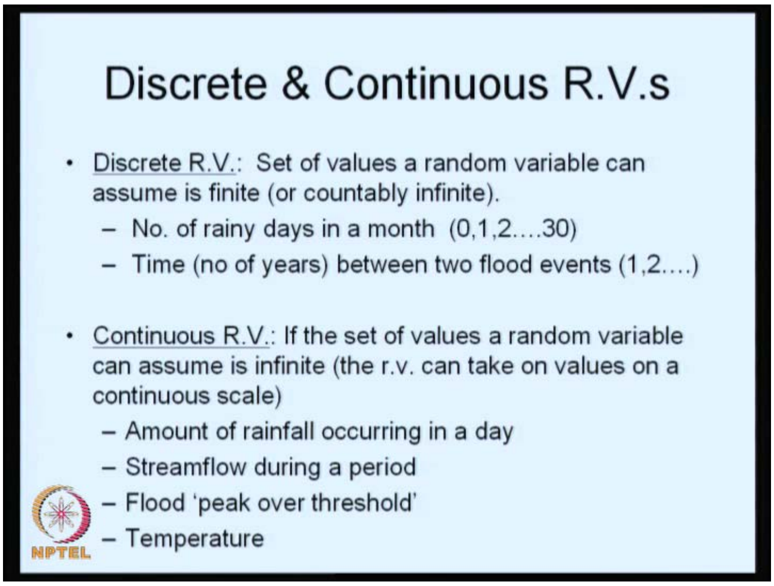
(Refer Slide Time: 28:03)



So, if you look at the variables in hydrology or let us say rainfall in a given duration, this becomes a random variable; stream flow becomes random variable. Soil hydraulic properties for example, permeability, porosity or and other properties like storage coefficient and so on in in the aquifers these become random variables. Time between hydrologic events; let us say you are interested in time between floods of a given magnitude this becomes hydrology random variable. Evaporation and evapo-transpiration, which are governed large number of natural variations they become random variables, ground water levels are random, then re-aeration rates are random and so on. So in fact, most of the variables of interest in hydrology are all random variables, because they are all governed by natural processes. So, we now start introducing more formally the analytical methods that we use for a in dealing with the random variables.

So, once we know what is a random variable, a function that is defined on a random variable also becomes a random variable; for example, if x is a random variable, then z is equal to g of x, where g of x is a function of x also becomes a random variable. We follow the convention that we use capital letters; for example, capital X, capital Y,-capital Z etcetera to denote the random variable itself and the associated small letters to denote the value that the random variable takes; for example, capital X may be rainfall; the small x may indicate 30 millimeter for and then similarly, capital Y is stream flow and small y may be the value that the stream flow takes. And we define events on the random variable, for example, we may say x is equal to 30 is an event and y lies between a and b is an event. And we associate probabilities to occurrence of such events and we represent these as; for example, probability of X being equal to 30; probability of Y taking on values between a and b and so on.

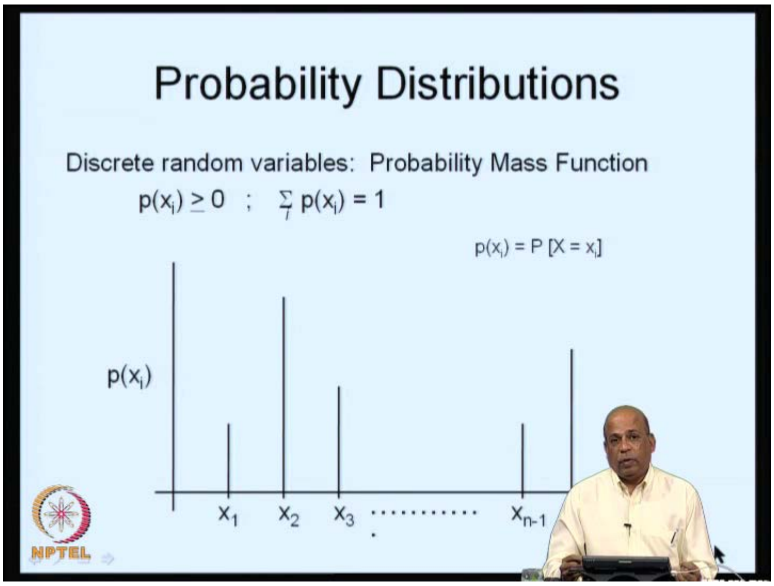
(Refer Slide Time: 30:41)



We have the concept of discrete and continuous random variables. If the random variable can take on values only discrete values; for example, it can take on values which are finite in number or perhaps countably infinite, then the particular random variable is called as a discrete random variable. Let us say as an example you are interested in number of rainy days in a month consisting of 30 days, now the number of rainy days can be either 0 or 1 or 2 etcetera up to 30. So this is a finite number of values that the random variable can take where the random variable is number of rainy days in a month; or we are interested in time in terms of number of years between two flood events. This can be 1, 2, 3 etcetera it can go on. But, they are still discrete number of values that it can take.

So this is such random variables are called as discrete random variables. On the other hand, the continuous random variables can take on values on a continuous scale. That is they the number of values that they can assume is infinite. Let us say you are talking about amount of rainfall occurring in a day. So you are talking about the actual quantity of rainfall. So, even if you know the range; let us say you know that the rainfall may be between 10 millimeters to let us say 30 millimeters or something, the number of values that it can take is on the line joining 10 and 30 on a real scale. So it can virtually assume infinite number of values and stream flow during a period flood peak over a particular threshold, temperature and so on.

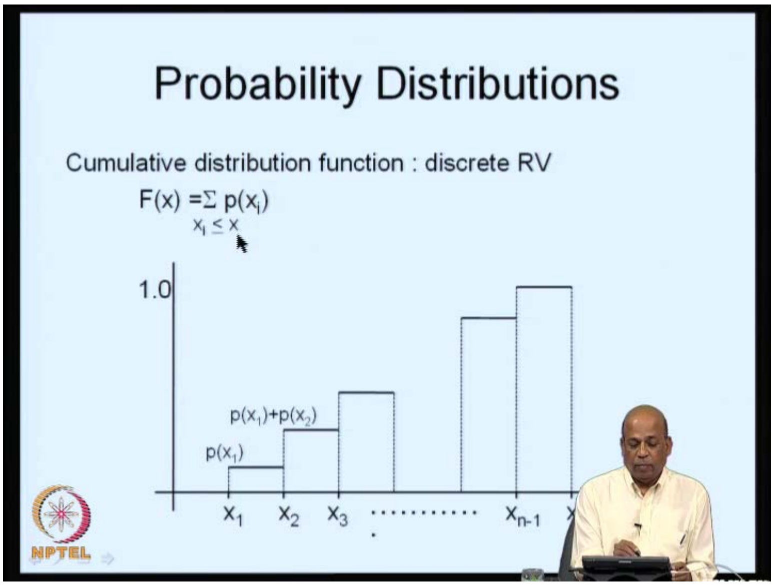
(Refer Slide Time: 32:45)



So most of the variables where we are dealing with amounts or the quantities in hydrology they all become continuous random variables. Once we know whether a random variable is a discrete random variable or continuous random variable, we start assigning the probabilities associated with these random variables associated with the events on these random variables. In the case of discrete random variables, we denote the Probability Distribution by Probability Mass Function. We call it as probability mass function, where the random variable X can take values discrete values X 1, X 2, X 3 and so on up to X n, it can take values up to X n.

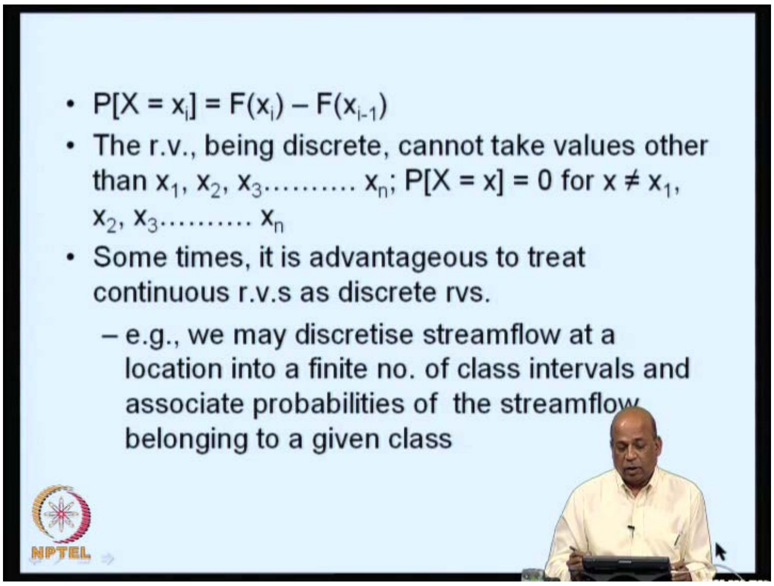
And probability of X i is indicated on this axis. So there is a finite probability of X being equal to X 1; X being equal to X 2; X being equal to X 3 and so on. Now, because X has to take at a exactly one of these values the sum of these probabilities is equal to 1. So the probabilities are non negative numbers and the sum of the probabilities over all the possible values that the discrete value can take is equal to 1. So, we indicate probability of X i as probability of X is equal to X i in the case of discrete random variables.

(Refer Slide Time: 34:15)



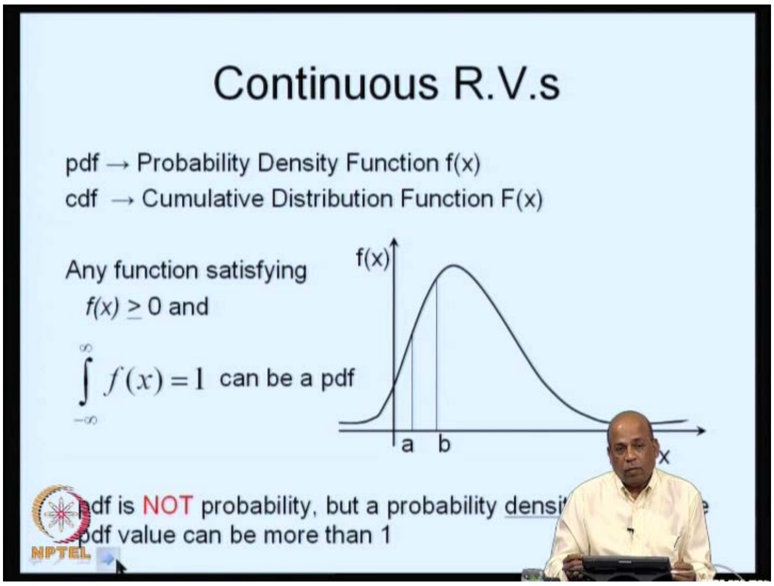
We also then talk about the cumulative distribution where we are interested in probability of X probability of the random variable taking on values less than or equal to a particular value. Let us say you are talking about probability of the random variable taking on a value less than or equal to X 3. So this is probability that takes on value X 1 plus the probability that it takes on value X 2 plus the probability that it takes on value X 3 and more formally we write it as such. So the capital F of X indicates the cumulative distribution function.

(Refer Slide Time: 34:54)



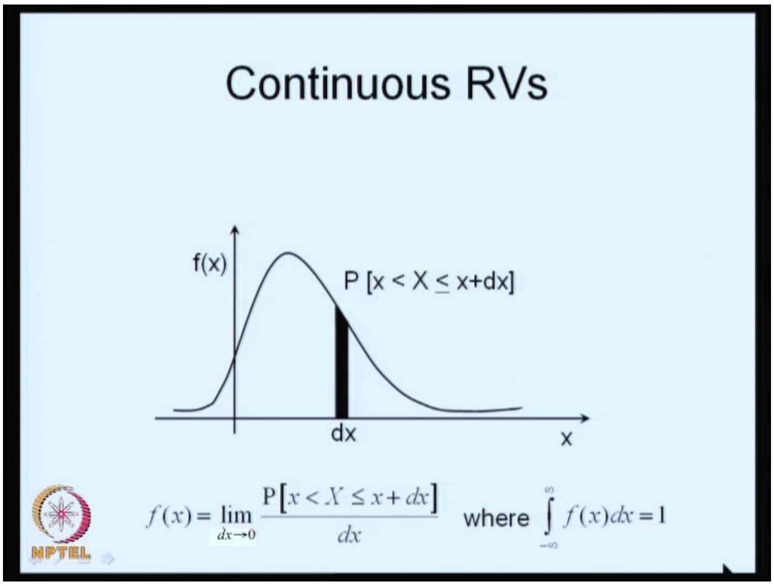
In the case of our discrete random variables then say for example, we are interested in probability of X is equal to x i. If you know the cumulative distribution it will be equal to F of x i minus F of x i minus 1. Look at this; for example, if you are looking at probability of X being equal to x 2 this is equal to this minus this. Then, the random variable being discrete cannot take on values other than the values that it can assume for example, x 1, x 2, x 3 etcetera x n, which means probability of X is equal to x a specific value of x for x other than x1, x 2, x 3 etcetera x n must be equal to 0; that means, it cannot take on any values other than these values. In hydrology, sometimes it is advantageous to treat continuous random variables as discrete random variables. In fact, in water resource systems analysis many times we use this. Let us say we are interested in stream flow. Stream flow is actually a continuous random variable, but we may discretise this stream flow into a number of class intervals and then treat the stream flow as if it was a discrete random variable, so it can belong to 1 of these classes.

(Refer Slide Time: 36:23)



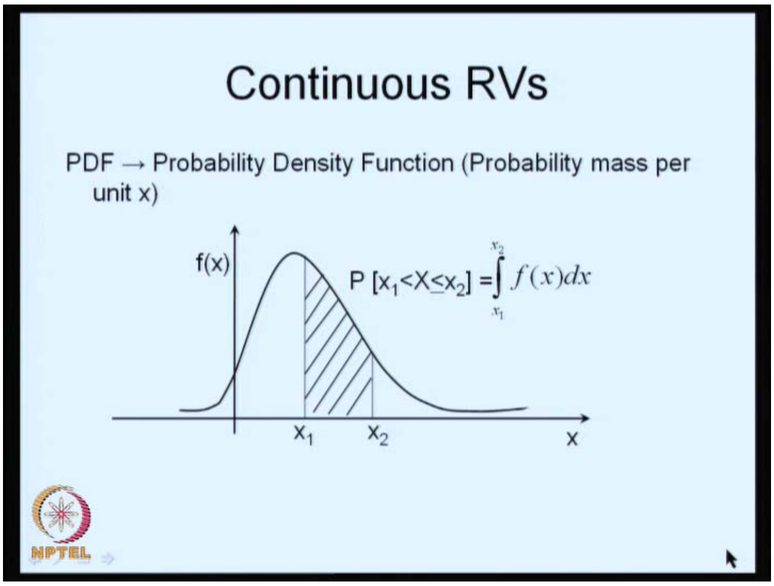
Then, we move on to the continuous random variables. Analogous to the discrete random variables, we also define the probability functions here. But, for the continuous random variable we define the Probability Density Function pdf denoted by small f of x, and the associated cumulative distributive function denoted by capital F of x. Any function that satisfies f of x any function f of x that satisfies f of x being greater than or equal to 0 that is non negative for all values of x and integral in the entire region minus infinity to plus infinity f of x equal to 1 can be a pdf. Note that pdf is not probability; unlike in discrete random variable where we defined the Probability Mass Function. Probability Mass Function directly gives you the probability, whereas the Probability Density Function is not probability, but a probability density.

(Refer Slide Time: 37:39)



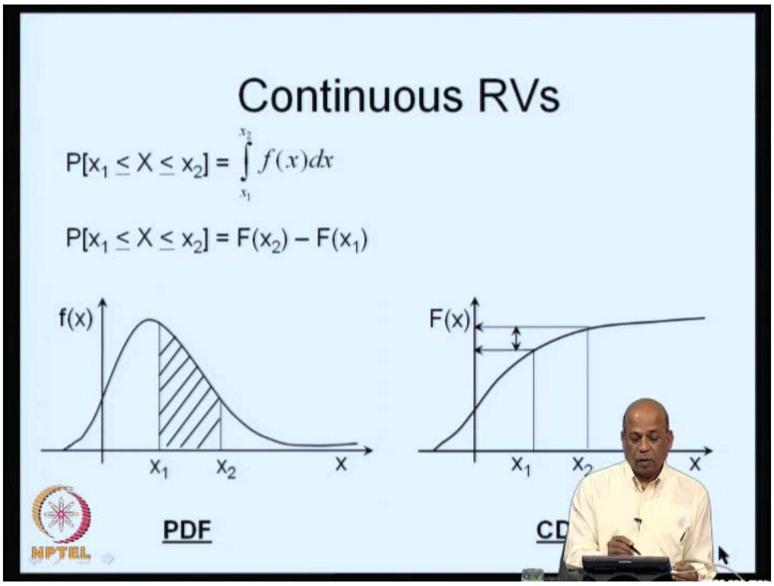
And therefore, the pdf value can in fact be more than 1. So for continuous random variable then, this is your pdf as an example. And because this being a density, we talk about the probabilities of x taking on a value in a small range here d x and this is how we get probability of X taking on a value between x and x plus delta x x plus d x in this case is the simply the area under this curve; area under pdf for this small strip here and that is how we get f of x as the limit we define f of x as the limit, and minus infinity to plus infinity f of x dx must be equal to 1 here. So, we use the pdf to obtain various probabilities.

(Refer Slide Time: 38:37)



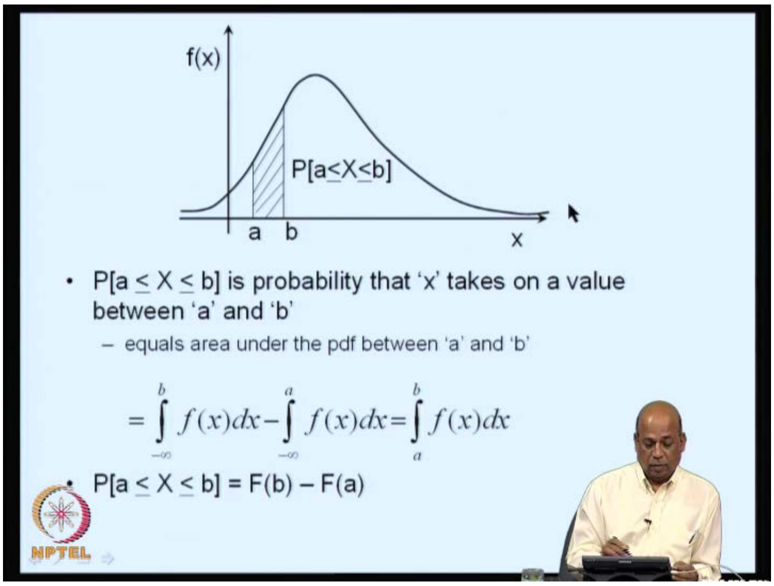
Let us say we are interested in getting the probability of X taking on values in a certain region between x 1 and x 2. So, we identify the region and then integrate the pdf over this region to obtain the associated probability; for example, probability of X taking on value between x1 and x2 is equal to integral x 1 to x 2 f of x dx.

(Refer Slide Time: 39:13)



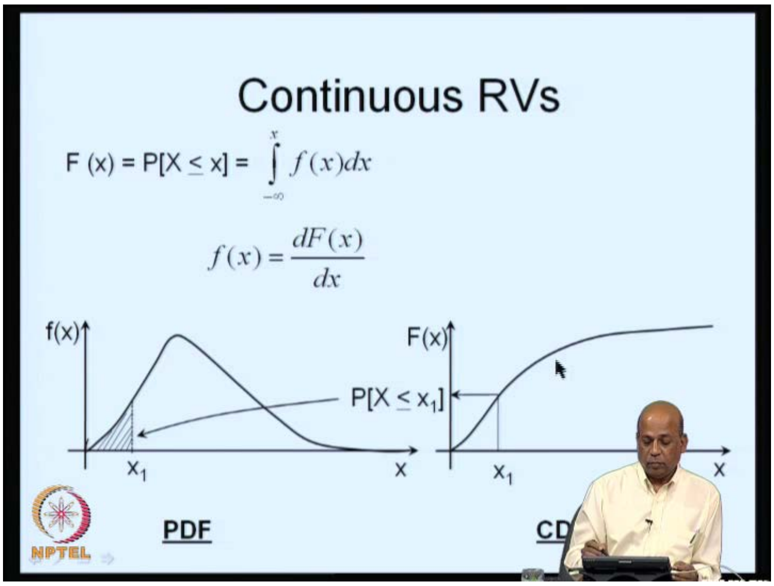
Similarly, we now come to the cdf's which is defined by minus infinity to x I am sorry this is not written here. Let us say we will write cdf as probability of X being less than or equal to x.

(Refer Slide Time: 39:30)



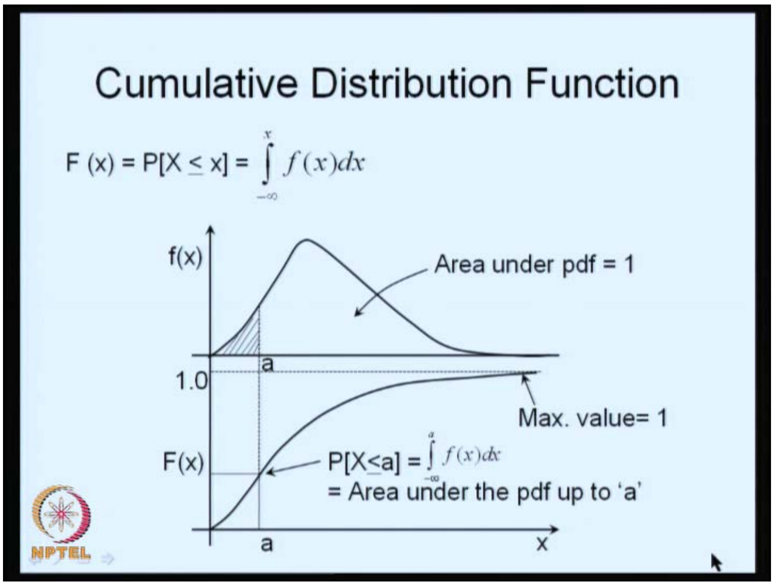
Let us say you are talking about the probability that X takes on a value less than or equal to a given value a specified value x 1.

(Refer Slide Time: 39:34)



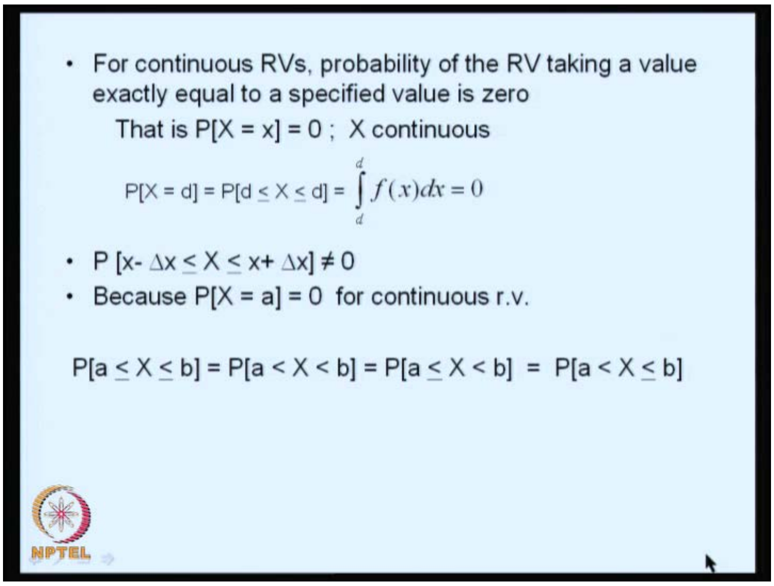
So, we identify a region over which x is less than or equal to x1 and integrate the pdf over that. By the definition of the cdf, cdf directly gives you the probability of X being less than or equal to x and therefore, if you have the cdf this value is directly equal to probability of X being less than or equal to x 1.

(Refer Slide Time: 40:29)



Because, cdf is a integral between minus infinity to x. Given the cdf you should be also able to get the pdf by differentiating the cdf or this only reinforces what we just talked about. So, we are talking about probability of X being less than or equal to a and from the cdf you can obtain this directly. So minus infinity to a of f of x dx will give you this area.

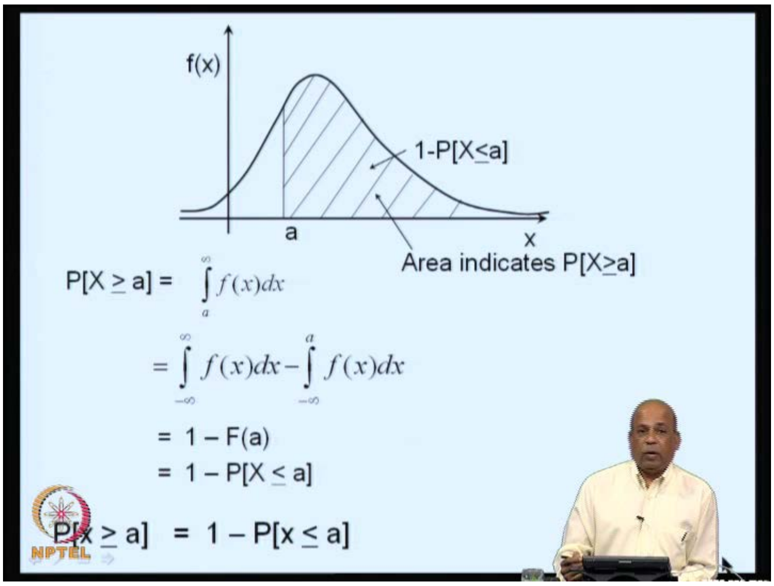
(Refer Slide Time: 40:50)



Now for continuous random variables, you must remember that the probability of the random variable taking a value exactly equal to a specified value is 0. That is what we are saying is probability of X being exactly equal to x is equal to 0, because as you can see; let us say probability of X is equal to d we are interested in. This we write it as probability of X being less than or equal to d and greater than or equal to d; which means you are integrating between d to d f of x dx which has to be 0. However in general, probability of X taking on a value in a small region x minus delta x to x plus delta x is in general not 0.

Now, because probability of X is equal to a, a given number is equal to 0 for continuous random variables, we can use interchangeably less than or equal to and less than and equal to and greater than or equal to and greater than as we have written here; for example, probability of X a less than or equal to x, less than or equal to b is also equal to probability of a less than x less than b equal to probability of a, less than or equal to x less than b and so on.

(Refer Slide Time: 42:09)

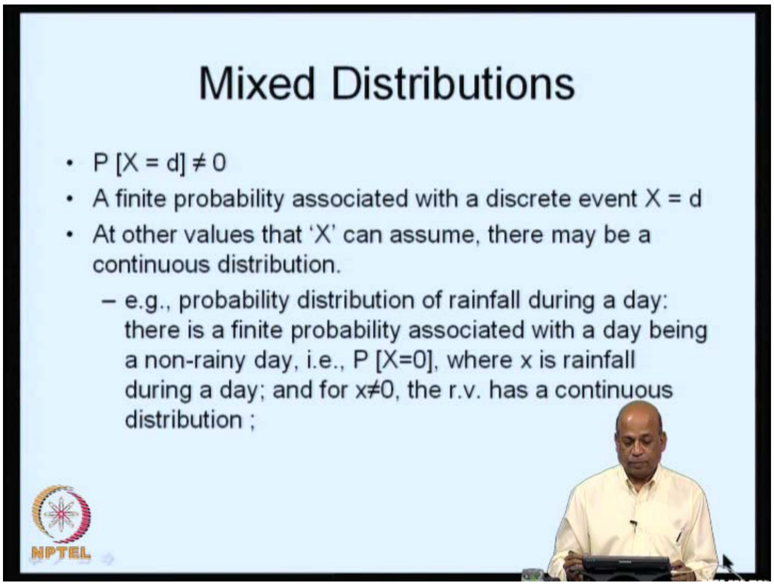


So, we will now look at how we use the pdf to obtain a various probabilities. As I mentioned earlier, if you are interested in getting the probability of the random variable taking on a certain value certain a values on a certain region, we identify the region and then identify the region under the pdf and integrate the pdf over that particular region. Let us say you are interested in probability of X being greater than or equal to a, a specified number.

This is the region in which x is greater than or equal to a. Now, because the total area under the pdf is 1 this region will also indicate 1 minus probability of X being less than or equal to a which we will presently see. So, we write probability of X being greater than or equal to a as a to infinity, because we are talking about this region here; a to infinity f of x dx. So you are integrating the pdf with respect to x over this region. Now this I can write as this entire area minus infinity to plus infinity f of x dx minus the area up to this point, that is minus infinity to a f of x dx.

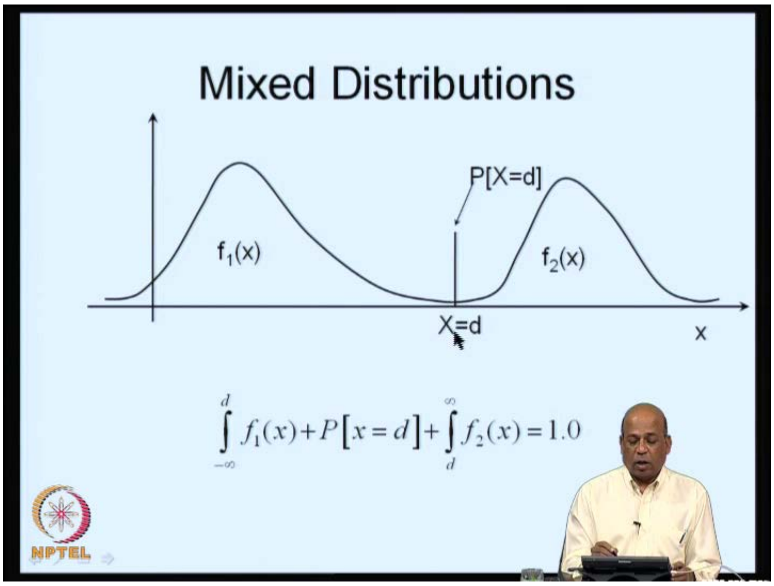
Now this area we know this integral is 1 by definition of pdf, so this is 1 minus by definition of cdf this is f of a, because the cdf is defined as f of x is equal to minus infinity to x integral minus infinity to x f of x dx. So we write this as 1 minus f of a, what is F of a? This is probability of X being less than or equal to a. So 1 minus probability of X being less than or equal to a; and therefore, we arrive at this conclusion that probability of X being greater than or equal to a is equal to 1 minus probability of X being less than or equal to a. This result is quite useful in many situations.

(Refer Slide Time: 44:19)



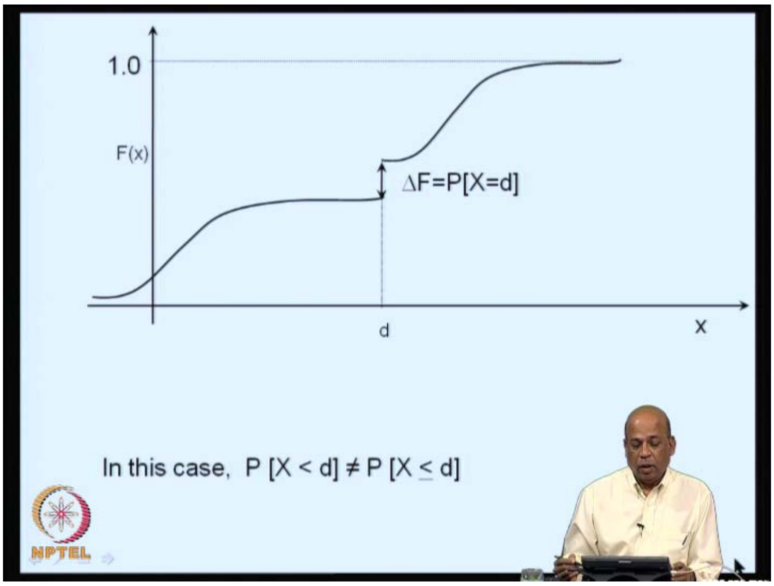
Then, we also have mixed distributions, where we may have some finite probability associated with a specific value. Let us say x is equal to d; probability of X is equal to d is not 0. There is a finite probability associated with this, then at other values other than x is equal to d we may have continuous distributions; for example, we may be interested in probability distribution of rainfall during a day. Now, there is a finite probability associated with a day being a non rainy day; that means, there is a probability of X being equal to 0 which is finite and then when it does rain, that is x not equal to 0 the amount of rainfall may have a continuous distribution. So these are called as mixed distributions.

(Refer Slide Time: 45:20)



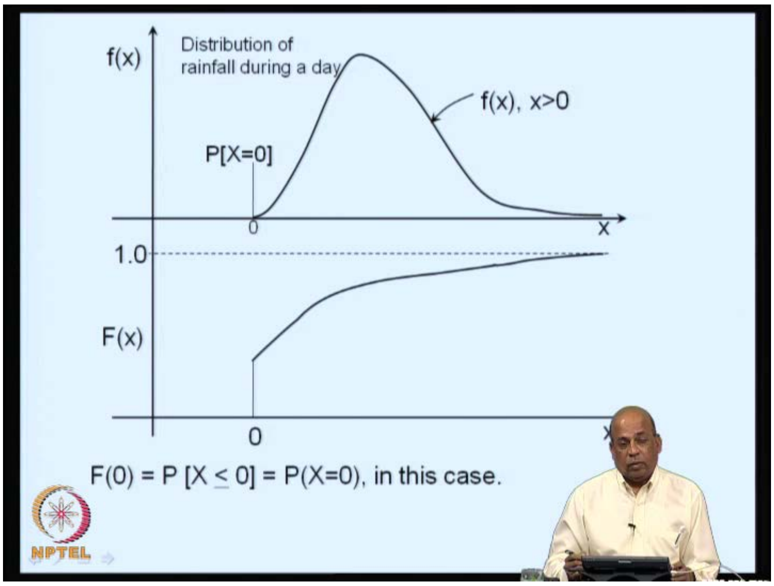
Now in mixed distributions; for example, we are talking about there is a finite problem we are talking about x being equal to d being associated with a certain probability, probability of X being equal to d. Then when x is not equal to d, there may be a continuous distribution f 1 of x for x less than d and f 2 of x for x greater than d. In such situations, the total probability that you get out of in the entire region here will be minus infinity to d that is this area integral f 1 of x with respect to x this is dx is missing there. Then plus probability of X is equal to d plus integral f 2 of x dx should be equal to 1; that means, the total probability must be equal to 1.

(Refer Slide Time: 46:14)



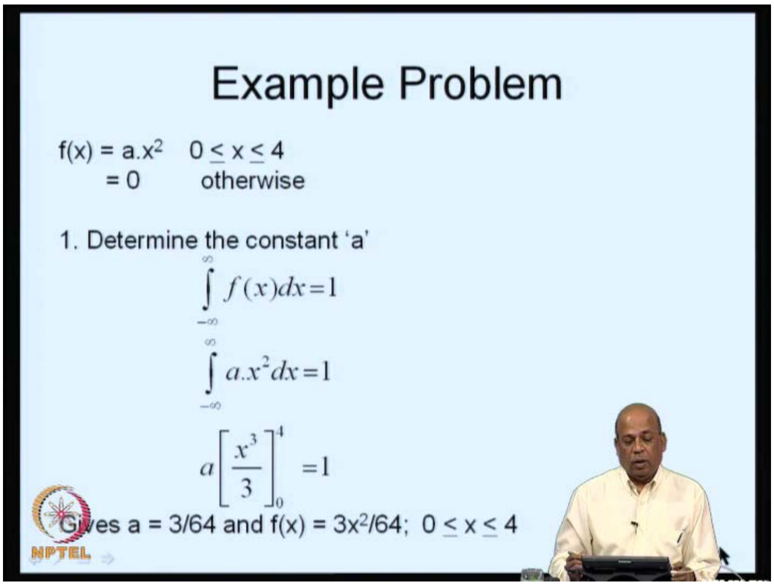
So in such a case, the cumulative distribution function will look something like this. So for less than d it goes up to this point, then there is a finite probability associated with x is equal to d. So there is a jump here of delta f is equal to probability of X is equal to d, then this is the cumulative distribution function associated with f 2 of x. So the final value of the maximum value of this cdf will be equal to 1. Note that in this particular case probability of X being less than d is not equal to probability of X being less than or equal to d, because there is a finite probability associated with the event x is equal to d.

(Refer Slide Time: 46:59)



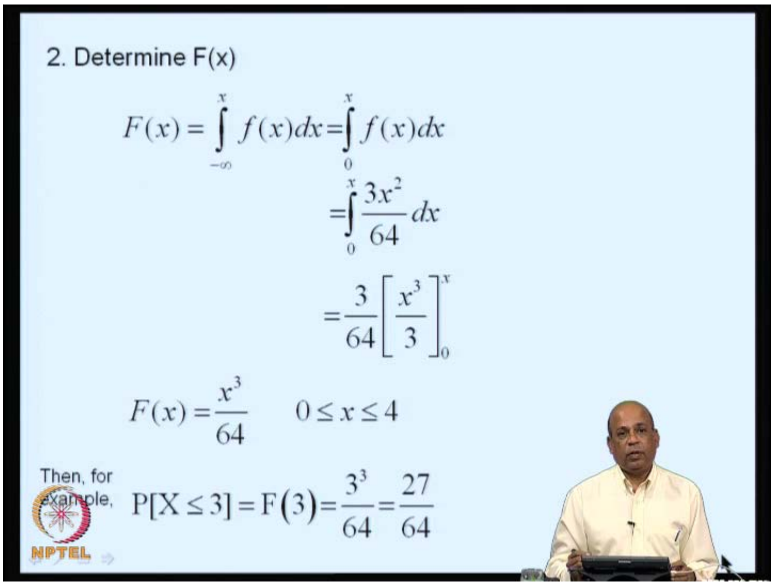
In the in the case of rainfall during a day, we may have probability of X being equal to 0 and then non 0 values may have a certain distribution of this type, and the associated probability cumulative distribution function will appear something like this. Again f of 0 is equal to probability of X being less than or equal to 0 is equal to in this particular case probability of X is equal to 0.

(Refer Slide Time: 47:28)



So, we have covered now the basic concepts of a random variable and the pdf and the cdf. Let us look at a few numerical examples so that the concepts are driven home. Let us say we are talking about a pdf f of x is equal to a x square over the region 0 to 4, that is x taking on values between 0 and 4 and it is 0 elsewhere. First, let us determine the constant a so that f of x becomes a pdf. Remember that for x to be f of x to be pdf first of all it has to be non negative.

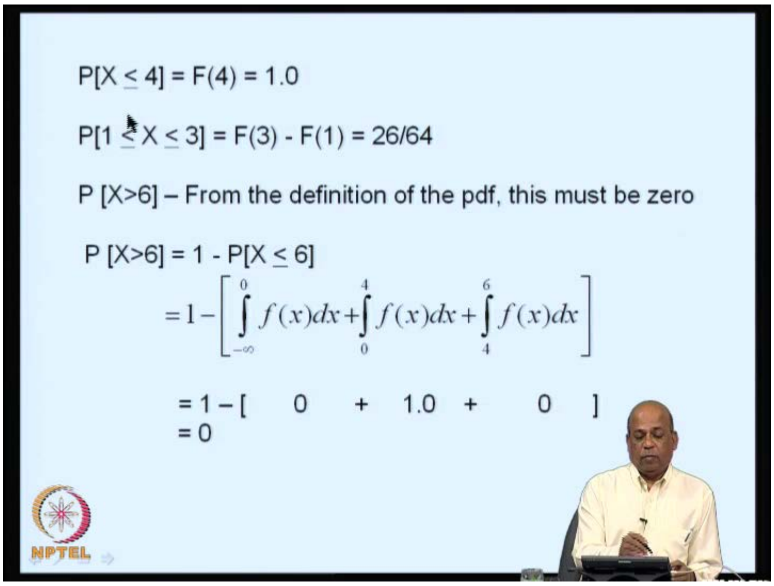
(Refer Slide Tim e: 48:46)



And then, the integral over minus infinity to plus infinity of f of x with respect to the variable x should be equal to 1. So, we will integrate this a x square d x over the entire region between 0 to 4 and then we arrive at a is equal to 3 by 64 and therefore, we get f of x is equal to 3x square by 64 here for this particular region.

Then, we can determine the cdf the capital F of x is the cdf. So minus infinity to x f of x dx which is between 0 to x f of x dx we integrate that and then get the cdf. So, cdf is defined as x cube by 64 over this range. Once we know the cdf we should be able to talk about various probabilities; for example we may be interested in probability of X being less than or equal to 3 and by definition of F of x we write this as f of 3 which will be 3 cube by 64 which is 27 by 64.

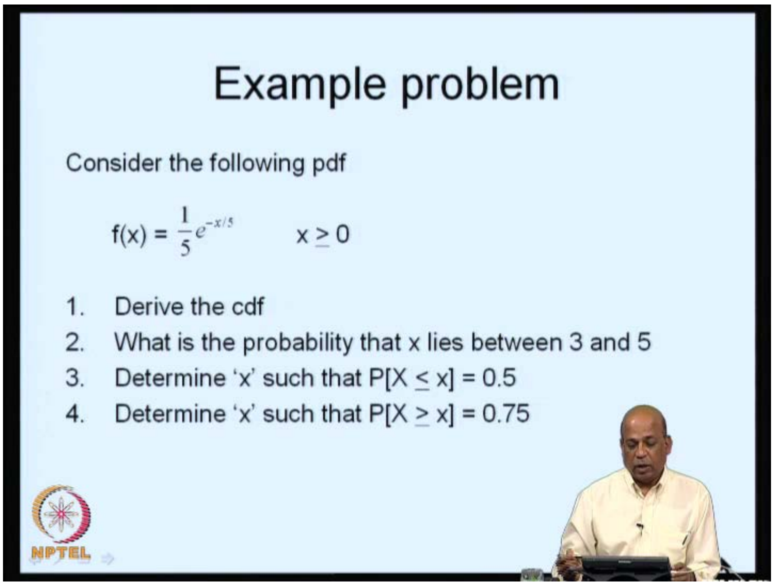
(Refer Slide Time: 49:27)



Then, we may be interested in probability of X being less than or equal to 4 which is F of 4, this should be 1 because x varies between 0 and 4. As you can see from the definition of the pdf your x varies between 0 and 4 and therefore, F of 4 capital F of 4 must be equal to 1. And let us say we are interested in probability of X taking on value between 1 and 3, this we it as f of 3 minus f of 1. This is from the result earlier we obtained; probability of X lying between a and b is equal to f of b minus f of a and this comes out to be 26 by 64.

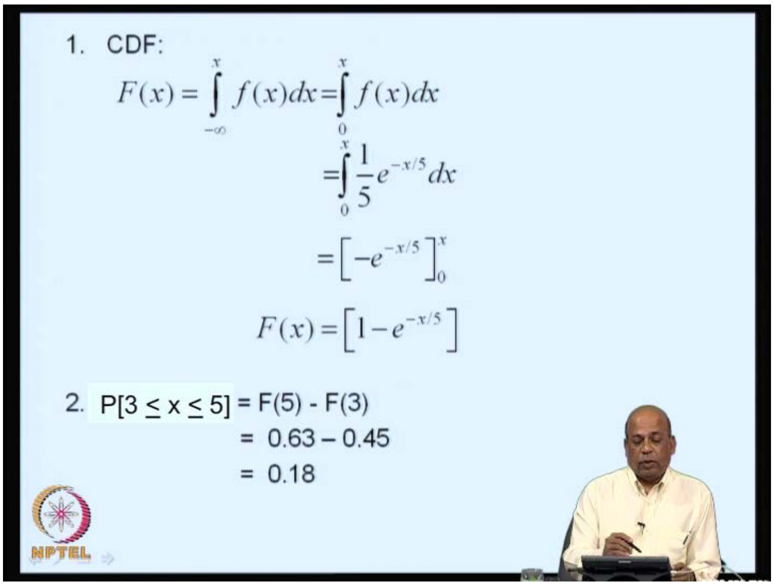
Let us say we are interested in probability of X taking on a value beyond the range of values for which f of x is defined. Let us say probability of X being greater than or equal to 6. This we write it as 1 minus probability of X being less than or equal to 6, which is 1 minus, minus infinity to we come up to 0, then 0 to 4 and 4 to 6. Now from minis infinity to 0 because f of x is 0 there you get a 0 there and 0 to 4, we are covering the entire region of the entire range of the random variable and therefore, this integral should be equal to 1. Then from 4 to infinity 4 to 6 your f of x is 0 again and therefore, this is 0 and therefore, the probability of X taking on a value beyond the range in which the pdf is specified is equal to 0.

(Refer Slide Time: 51:12)



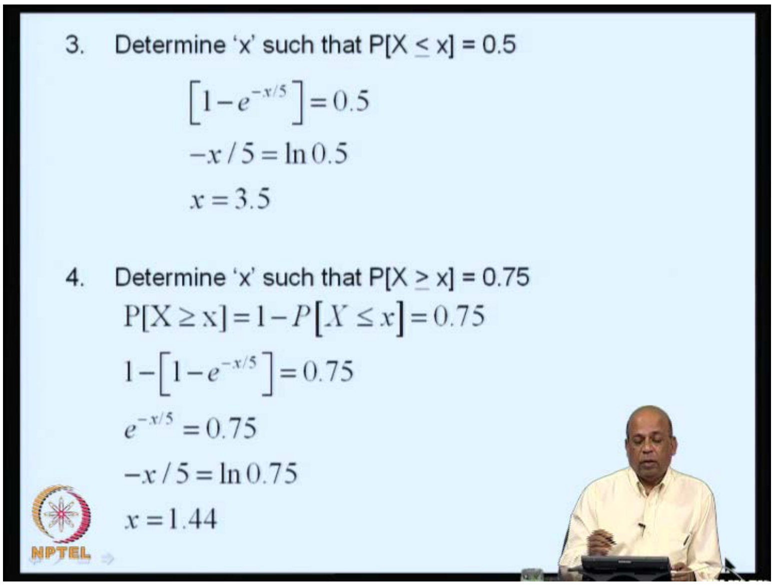
We will quickly see another example problem where your f of x is defined as 1 by 5 e to the power minus x by 5 for x greater than 0. We will derive the cdf for this, then we will also check the probability that x lies between 3 and 5 and we will determine the value of x such that probability of X is less than or equal to x is equal to 0.5. And similarly, probability of X being less greater than or equal to x is equal to 0.75.

(Refer Slide Time: 51:46)



So, CDF will go with minus infinity to x integral of minus infinity to x f of x dx, we get from this 1 by 5 e to the power minus x by 5 dx. So this will be minus e to the power minus x by 5 0 to x. Therefore, we get f of x is equal to 1 minus e to the power minus x by 5. Once we know f of x we should be able to talk about probabilities such as this. Probability of X; this is the capital X here; X taking on values between 3 and 5, so this will be equal to f of 5 minus f of 3 that is equal to 0.18.

(Refer Slide Time: 52:33)



Then, we look at what is that value of x for which probability of X being less than or equal to x is equal to 0.5. So probability of X being less than or equal to x is nothing, but our cdf so 1 minus e to the power minus x by 5, this will be equal to 0.5 and that turns out to be x is equal to 3.5. Similarly, for probability of X being greater than or equal to x equal to 0.75, we write this as x being greater probability of X being greater than or equal to x is equal to 1 minus probability of X being less than or equal to x, this is equal to 0.75 and 1 minus this is your cdf. So this should be equal to 0.75 from which we get x is equal to 1.44.

So in summary then, we have in the first class looked at the various applications of stochastic hydrology and we have introduced the concept of random variables. We have looked at the probability mass function for the discrete random variable and probability density function for the continuous random variable, the associated cumulative distribution functions and the way to arrive at the various probabilities that the random variable. Various probabilities associated with the events defined on the random variables and we have also solved two simple examples. So in the next class, we will introduce two dimensional random variables and see how the joint density functions are defined, how the cumulative, how the joint cumulative distribution functions can be derived. And then, we move on to look at conditional probabilities marginal probability distribution and so on. Thank you very much for your attention.