INTRODUCTION TO PROBABILITY THEORY AND STOCHASTIC PROCESSES

Prof. S Dhramaraja
Mathematics
IIT Delhi

INDEX

Week 6

Week 7

Week 8

Week 9

Week 10

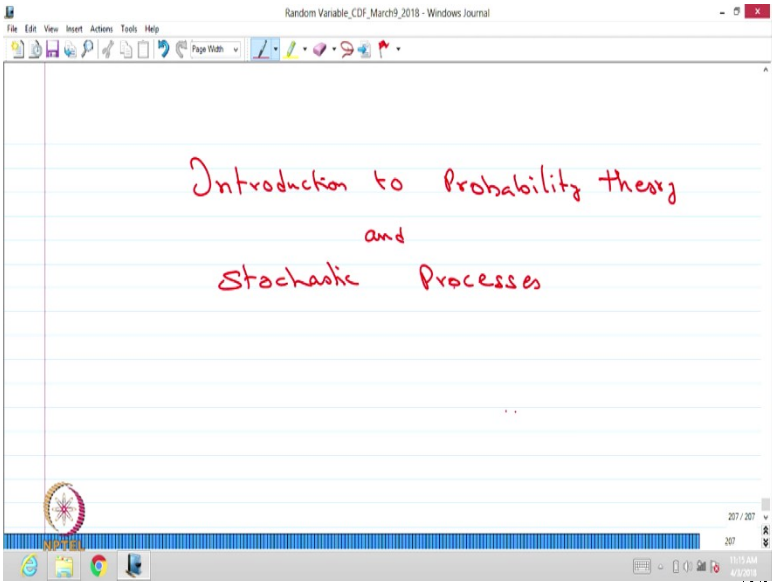
Week 11

Week 12

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module – 01
Basics of Probability
Lecture – 01
Sample Space and Σ-Field

(Refer Slide Time: 00:23)



This is a course on Introduction to Probability Theory and Stochastic Processes is a course on Introduction to Probability Theory and Stochastic Processes. This is an undergraduate level course. In this course, we are going to discuss the two topics; one is a probability and the second topic is stochastic processes.

Since, it is undergraduate course, we are going to discuss probability and the stochastic processes in the introductory level; that means, we can always have a post graduate course level on probability separately and there will be a separate course on stochastic processes whereas this is undergraduate course on Introduction to Probability Theory and Stochastic Processes. The word theory is important in this course because we are going to cover some of the theoretical aspects of probability as well as some problems on the probability as well as the stochastic processes.

So, we are going to discuss the theory part of probability as well as some theory part of a stochastic processes also in this course. The whole course is a divided into 12 models, each model is roughly of a 3 hours lectures. So, these 12 models; out of 12 models, 8 models- we are covering for probability theory and the remaining 4 models for the stochastic processes.

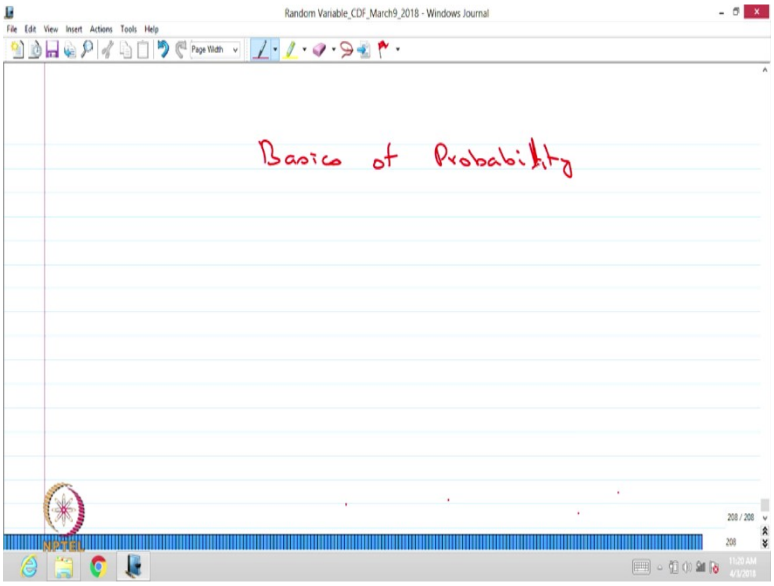
Out of 8 models in probability theory, we started with the first model of basics of probability, second model we discuss the random variable, third model we have a movements and inequalities, fourth model standard distributions. So, basically the first 4 models; we discuss about the 1-dimensional random variable that is basically sort of very elementary level of probability, then the fifth model; we started with the random vectors; that means, multi dimensional random variables.

Then in the sixth model, we discuss the functions of several random variables, in the seventh model, we discuss the cross movements and eighth model we discuss the limiting distributions that is a very important topic in the probability theory limiting distributions. So, with that 8 models we cover the probability theory, then from ninth model onwards we have stochastic process.

So, there also, we disuses only the introductory level; that means, we started ninth model with introduction to stochastic processes in which we define definition, properties some common random processes or stochastic processes and some important properties, then in the tenth model we discuss the discrete timer coaching in detail and the eleventh model we discuss the continuous time Markov chain, in the twelfth model we discuss the simple Markovian queuing models simple means the underline stochastic process is going to be a birth death process.

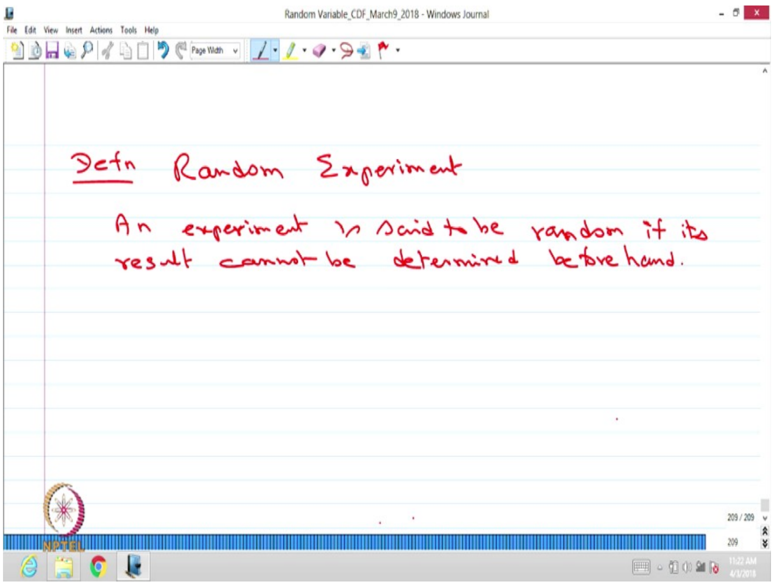
So, we discuss stochastic processes in 4 models and probability theory in 8 models. Let me start with the first model that is basics of probability.

(Refer Slide Time: 04:41)



In this model, we are going to discuss about the random experiment, then sample space, then axiomatic definition of probability and probability space in the first lecture and the second lecture, we are going to discuss the conditioning probability and independent of events. And in the third lecture, we are going to discuss the total probability rule, multiplication rule and Bayes theory with that we are going to complete the first model that is called basics of probability.

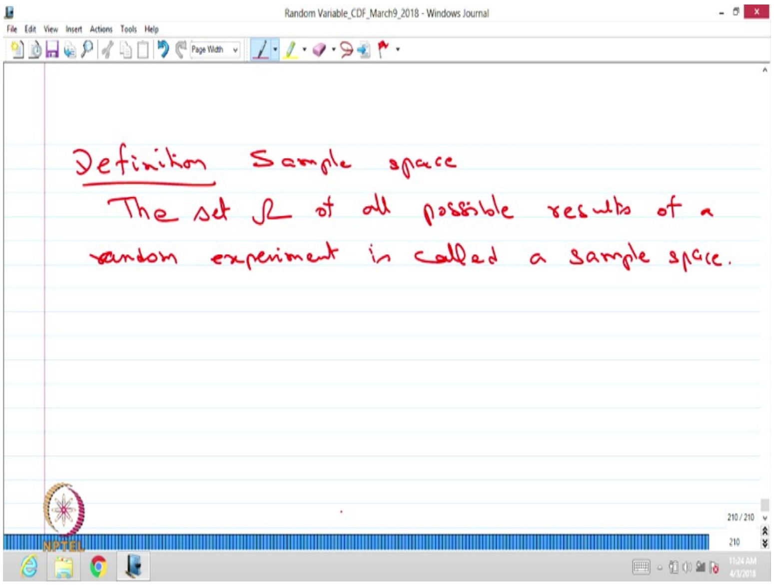
(Refer Slide Time: 05:37)



Let me start with the definition of a random experiment, the first definition that is called random experiment. An experiment is said to be random if its results cannot be determined beforehand, whenever in any experimental, result cannot be determined beforehand, then we will say that experiment is a random experiment.

In the whole probability theory, we started with the random experiment; that means, we have an experiment in which the results cannot be determined beforehand, we can think of many examples over the course, we are going to discuss one by one examples, therefore, I am going to give the examples little later.

(Refer Slide Time: 07:11)

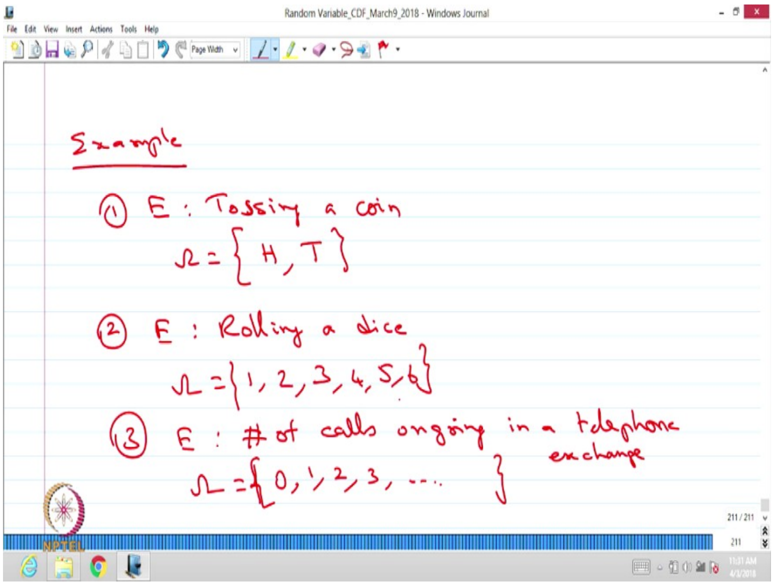


The next definition; sample space, the set Ω of all possible results of a random experiment is called a sample space. The sample space is nothing, but the collection of all possible results or outcome of a random experiment; that means, we have a random experiment in which the results cannot be determined beforehand. If you able to collect the all possible outcomes of a random experiment that collection or that set of all possible results that is going to be call it as a sample space; each possible outcome that is called the sample either possible results or possible outcomes one and the same.

So, each possible outcome or each possible result that is going to be called it as a sample and collection of all possible results or all possible outcomes that going to call it as a sample space since we have a random experiment of a many kind, the set could be consisting of a finite number of elements or it could be countably infinite number of elements in it or it could be uncountably many elements. The set Ω may consist of finite elements or countably infinite number of elements or uncountably many elements not only that; those elements could be numerals or non-numerals. You may have some random experiment in your mind all possible outcomes or results of a random experiment, if you make a collection that collection could be numbers or it could be non-numbers.

Similarly, it could be finite or uncountably infinite or uncountably many elements.

(Refer Slide Time: 10:14)



So, I will go for few examples for the sample space. Example 1; the easiest example in the random experiment, I use a notation E for random experiment, random experiment is the suppose you have a very simplest example tossing a coin, tossing a coin that is a random experiment in the set of all possible outcomes.

In this random experiment is going to be I use a notation H for if I get the head, I use a notation H for obtaining head. So, only I use the notation T for getting a tail; that means, the only two possible outcomes are head or tail.

So, this is the collection of all possible outcomes in this random experiment. Let's go for another example, the random experiment is rolling a dice here the all possible outcomes, I show the each side that number, it is a number which I am going to get either the number 1 or I am going to get the number 2 or 3 or 4 or 5 or 6, these are all the possible outcomes in the dice whenever you roll, the possible outcomes going to be either 1 or 2 or 3, 4 or 5 or 6.

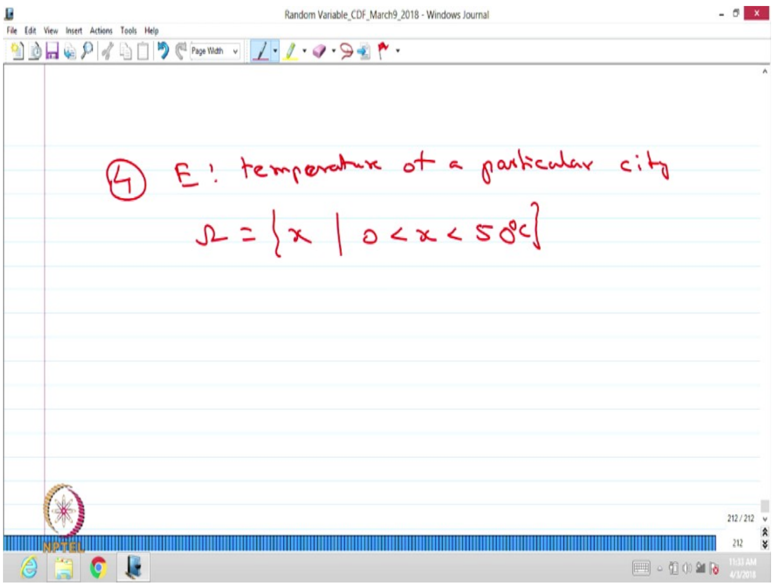
These are all the possible outcomes. So, the Ω consisting of all possible outcomes that all is very important; you should collect you should make a set consisting of all possible outcomes.

So, first one has a only 2 elements, whereas, a second one has 6 elements. Third example; random experiment is a number of calls telephone calls ongoing in a particular telephone exchange; number of ongoing calls, calls ongoing or ongoing calls in a telephone exchange the Ω all possible outcomes. There is a possibility no calls or there is a possibility only one call or there is a possibility two calls and so on; even though it may a finite number of calls.

So, you can just putting a dot dot dot; that means, it is a huge number. So, it could be 1000 or it could be 2000 or whatever be the number. So, I am just writing 0, 1, 2, 3 and so on. So, so many calls are going this is going to be a all possible outcomes of tis random experiment it could be finite also. So, this is going to be the collection; collection or set, you see the first example; it is not numbers whereas, the second example we have a numerals third example we have a numerals.

So, the way you have a random experiment, the collection of all the possible outcomes could be numerals or non-numerals, and the collection could be finite number of elements or countably infinite number of elements or uncountably many elements also, suppose I give a fourth example in which the random experiment is temperature of a particular city.

(Refer Slide Time: 14:33)

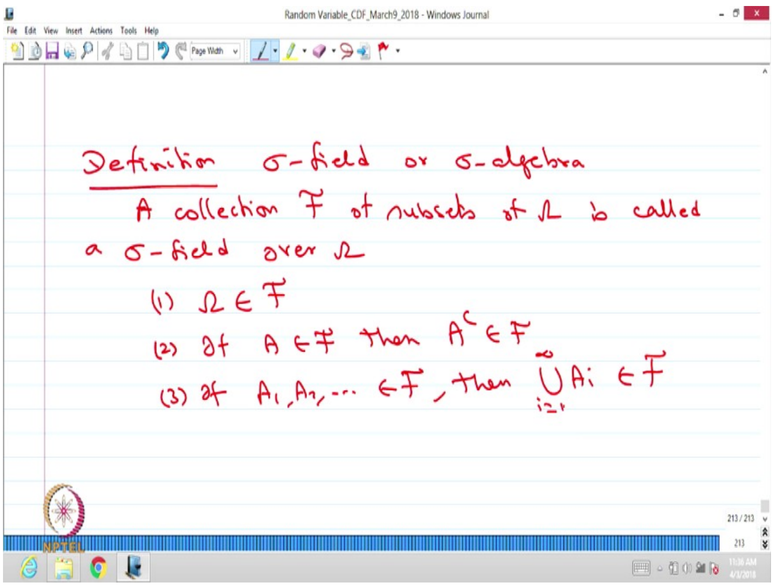


In this random experiment, the Ω is going to be the temperature can vary from some range, for example may range from collection of x such that in that particular city the temperature is a most of the time say lies between 0 to 50 degree centigrade. Suppose, I make the assumption, the temperature always goes from 0 to 50.

So, the Ω is going to be a collection of values lies between 0 to 50. So, it is a real number, therefore, a the elements of Ω is uncountably many between the interval 0 to 50 degree, 50 degree centigrade, I have given the example in which it is going to be numerals or non numerals, similarly, it could be finite or countably infinite or uncountably many elements also.

Now, we move into the next definition that is called σ-field.

(Refer Slide Time: 16:10)



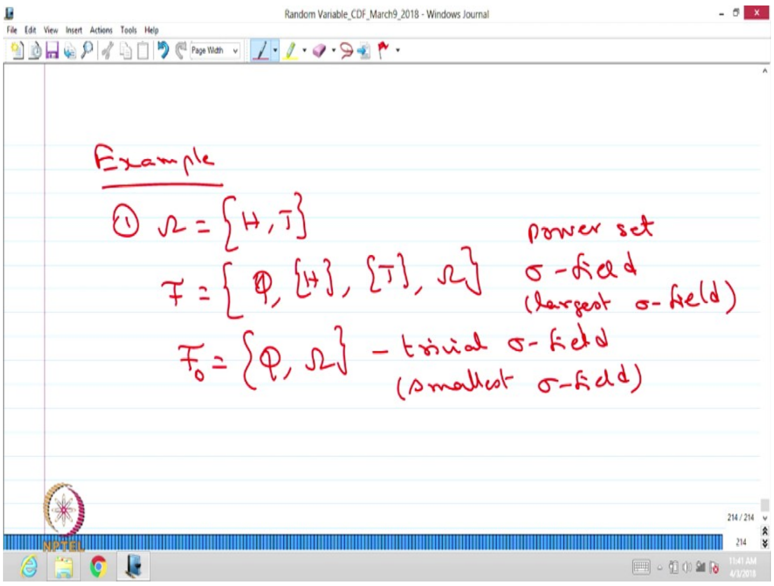
Or there is another name called σ-algebra, a collection that is denoted by F of subsets of Ω; that is called a σ-field over Ω, either, you can use over Ω or on Ω; the Ω is a nonempty set such that the Ω is belonging to F; second one if A is belonging to F then A^c that is also belonging to F. A collection F of subsets of Ω is called σ-field over Ω or on Ω, whenever it satisfies these 3 conditions, the first condition is Ω belonging to F.

If A belonging to F, then A^c is also belonging to F the third condition, if A_1, A_2 and so on; that is belonging into F, then the $\cup_{i=0}^{\infty} A_i$ that is also belonging to F whether you take a finite elements in the F or countably infinite elements in the F, the union of those elements also belonging to F.

If any collection F of subsets of Ω satisfying these three conditions, then we call that collection is called σ-field on Ω. So, Ω has to be a nonempty set then satisfies these 3 conditions of a collection of subsets of Ω, then it is called the σ-field or σ-algebra. This is going to be a very important to define a probability, let me give examples for the σ-felid, then we will go to the definition of probability.

Now onwards, I won't discuss the random experiment, I will start with the sample space.

(Refer Slide Time: 18:54)



Suppose the sample space is the same, example head, tail, we can create the σ-field or σ-algebra on Ω that is F empty set, this is empty set singleton element H, tail and Ω, you can verify whether all three conditions of the σ-field is satisfied. empty set is one element, compliment is the whole set both are belong in to F, the third condition, if I take a few elements union is also belonging to the same F. Suppose, if I take H as well as empty set union is going to be again H that is also belonging, if I take tail and the Ω the union is going to be Ω and if I take a head and tail union, then that union is going to be Ω that is also belonging that is a third condition, second condition, if I take any one element the compliment is also belonging to the F.

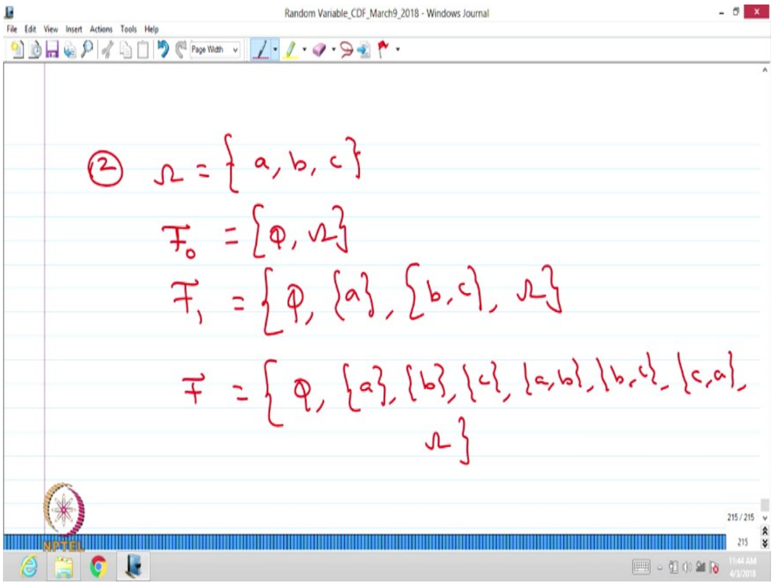
So, if I take empty set complement is Ω that is also belonging and the $H^c = T$ that is also belonging to F and $T^c = H$ that is also belonging to F, Ω complement is empty set are also belonging to F. Therefore, all those 3 conditions are satisfied by this collection which collection is a subsets of Ω, therefore, this is going to be a σ-field or σ-algebra, this is not the only σ-field which satisfies the three conditions, we can have a one more σ-field that is empty set and the whole set.

This also satisfies all the three conditions of a σ-field because Ω is one of the elements, compliment of empty set and the whole set is there and the union is also belonging to the collection therefore, this is also σ-field. So, this is called a trivial σ-field. For any Ω, you can always make a σ-field that is a consisting of empty set and the whole set that is a trivial σ-field, even some books, they use the notation that is F_0 if it is F_0 that is a trivial σ-field.

So, not only this trivial σ-field, this is a smallest σ-felid also, since I use the word smallest σ-field the other one which has a singleton element starting from empty set a singleton element and the whole set.

So, this is the largest σ-field, since it has a two elements Ω consisting of two elements, if you go for 2^2 that is 4. So, number of elements in this σ-field F as a exactly 4 elements, therefore, this is the power set also; this is the power set which is the largest σ-field. In this example, we have a trivial σ-fields one is the smallest one and other one is the largest one that is a power set in between we are not getting any other σ-field, we are getting only two σ-fields because it has only two elements; one is if the smallest, other one is the largest. Now, I am going to discuss some example in that we will land up and non trivial σ-fields other than smallest and the largest.

(Refer Slide Time: 23:28)



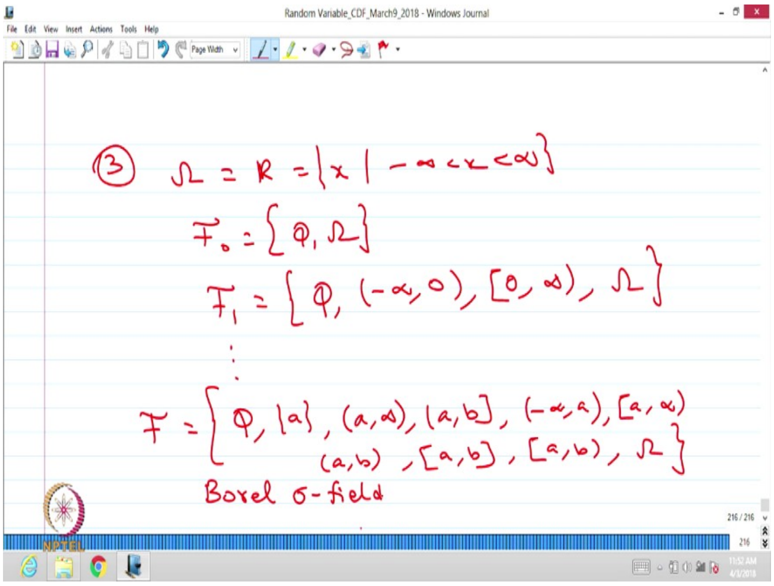
Second example, I would not bother about what is the random experiment, directly I give the sample space.

So, the sample space consisting of three elements, sample space consisting of three elements. Now, you will go for creating the σ-field, the smallest one always empty set and the whole set the another σ-field which I can create empty set I take a singleton element {a}, {b} and {c} together therefore, $\{a\}^c = \{b, c\}$ and $\{b, c\}^c = \{a\}$and union of {a}, {b} and {c} that becomes the whole set.

You can verify; you can verify F_1that is a σ-field because it satisfies all the three conditions of the σ-field, I can go for similarly, I can go for another σ-field keeping b separately and a and c together. Similarly, I can go for another σ-field keeping c separately, a and b together; that means, I can create two more σ-field of similar kind, I am going for the largest σ-field satisfying all the three conditions that is empty set, singleton element, any two element a, b; b, c and c, a and all three elements together count 1, 2, 3, 4, 5, 6, 7, 8 elements number of elements in the σ-field is 3. So, 2^3 is 8. So, this is the largest σ-field nothing, but the power set. So, in this example we have created the trivial σ-felids like the smallest one and the largest one and in between, we have created some more σ-fields, I have created one, similarly, you can create two more σ-fields.

That means, for a non empty set Ω based on the number of elements. You can always able to create many σ-fields including the trivial σ-fields and the σ-fields is going to be a play important role in the probability that we are going to discuss in the later part the first example and the second example, we have a finite elements though Ω consisting of finite elements, we can go for creating the σ-field when Ω is uncountably many elements.

(Refer Slide Time: 26:47)



Let us go for third example that is Ω is same as the real line; that means, the elements is from -∞ to ∞ that is a collection of x such that x lies between -∞ to ∞ real numbers uncountably many, Ω consisting of uncountably many elements in it.

Now, we are going to create the σ-field on Ω, the Ω consisting of uncountably many elements. Obviously, here also that trivial σ-field is going to be empty set and the whole set. We can go for non trivial σ-field F_1 consisting of a empty set, (-∞,0), the other element is [0,∞].

Then the whole set; you can verify whole set is one of the element and the second condition is a complement of every element is also belonging to the same collection, complement of empty set is whole set $(-\infty,0)^c=[0,\infty]$, the closed form and $[0,\infty]^c=(-\infty,0)$, complement of a whole set is empty set and if you make a union of any 2 elements, are any 3 elements, are all 4 elements that is also belonging to F_1, therefore, F_1 is also going to be a σ-field. The way I break -∞ to ∞ at 0 one side open interval, other side 0, you can think of any real number between the interval -∞ to 0, you can partition the whole interval into two pieces at any real point, then you can make a 2 elements.

Then you will have many σ-fields. This is a easiest σ-field one can create by partitioning the interval into 3 pieces, 3 subintervals, then it is going to be little difficult, you are to go for creating a union and the complement of everyone interval also belong into F and so on, that is little TDS job, but still you can partition the interval minus infinity to infinity into finite number of subintervals then accordingly you put some more elements. So, that all the three conditions are satisfied like that you can create many σ-fields. But we are going to have a largest σ-field, you can just visualize what are all the forms of intervals or elements going to be the element of F or the collection of a subsets of **Ω**.

So that, F is going to be σ-field in which it is going to be elements of it start the with empty set it is with the singleton element {a} and it is going to be having the element of (a,∞) form, a can be a real number and it will be of the form (a, b}, b is also real number and it is going to be of the form (-∞,a) and it is going to be of the form closed interval [a, ∞) and it is going to be of the form (a, b), [a, b], [a, b), here both a and b can take any real number. So, these are all the all the combinations of different real numbers we will land up.

Ok one more element that is **Ω** also. So, these are all the different forms will form a collection that is going to be the largest σ-field on or over **Ω** were **Ω** is a real line. Special name for this σ-field that is called Borel σ-feild this is Borel σ-field on the real line. Suppose, I have a **Ω** which is not the whole real line. It is a in the sub interval 0 to 1 closed interval, then you can always a create a F which is Borel set on real line intersection with the 0 to 1. You no need to have a whatever Borel σ-field, you created on the real line, suppose your **Ω** is going to be a subinterval between minus infinity to infinity.

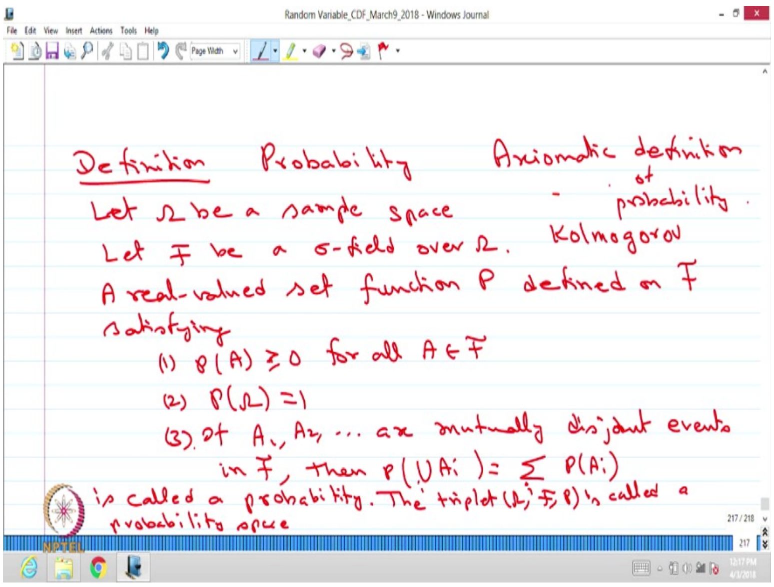
Then you can always make a σ-field on **Ω** which is a subinterval of real line by intersecting that interval with the Borel σ-field, therefore, now we have discusss how to create the σ-field. Whenever we have **Ω** having a finite number of elements and uncountably many elements; obviously, when you have a countably infinite number of elements you can do the similar exercise for creating the σ-field.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 02
Kolmogorov Axiomatic definition of Probability

Now, you are moving in to the definition of probability; definition of probability.

(Refer Slide Time: 00:05)



It is a very important definition because the probability theory goes with the definition of probability. Let **Ω** be a sample space, let F be a σ-field, a σ- field over **Ω**. A real valued set function that is P defined on F satisfying the first condition that is P(A) ≥ 0; for all A *belonging* ¿ F.

The first condition is P(A) ≥ 0 for all A belonging to F. The second condition P(**Ω**) is one, third condition if A_1, A_2, and so, on are mutually disjoint events, I have not said what is the meaning of events; I will discuss events later mutually disjoint events in F, then $P\left(\cup_i A_i\right)=\sum_i P(A$¿¿$i).$¿

Basically, a real valued set function P defined on F satisfying these 3 conditions that is called a probability, it is called a probability. The triplet that is (Ω, F, P); that is called a probability space. This definition is very important we have random experiment you know what is the meaning of random experiment; that means, it is experiment in which the results are not known in advance.

If you collect the all possible outcomes of random experiments that collection, all possible outcomes that collection is going to be call it as sample space that is Ω. From Ω that is a non empty set obviously; for from there we created the σ- field. So, you create a σ- field, one σ- field; it could be trivial or nontrivial you create σ- field over nonempty set Ω.

Now you are going to create set function that is real valued function; that means, it is mapping from F to R, P function, it is a set function from F to R, F is σ- field which consisting of subsets of Ω satisfying the 3 conditions of the σ- field. Therefore, the elements of F that is nothing, but the subsets of Ω subsets since it is subsets of Ω that is called events.

So, here the A when I make the first satisfying the condition $P(A) \geq 0$ for all A belonging to F. So, each element of F that is called event because that is subsets of Ω; that means, few collection of possible outcomes that is going to be the event; even single element is also single possible outcome or any possible result or each sample that is also called a event or subsets of possible outcomes that is also going to be call it as a event.

So, here the element of the F is called event whereas, element of Ω that is called samples. Repeating again the elements of Ω that is called samples, element of F is called event. We are defining the set function P on F which is the real valued function therefore, it is mapping from F to R; satisfying the three condition. See the first condition $P(A) \geq 0$; that means its real valued function and it is ≥ 0. Therefore, the values are possible value of range of P is going to be from 0 to ∞ from the first condition. But if you see the second condition, $P(\Omega) = 1$; if you recall F consisting of element starting from empty set, singleton element and all the possible elements and so on, finally you have Ω.

The $P(\Omega) = 1$ makes the real valued set function defined from F to R, which has the range from 0 to 1; it is real valued function that is means it could be $-\infty$ to ∞. Since the $P(A) \geq 0$ that restrict from 0 to ∞. The second condition $P(\Omega) = 1$; that is makes lies between 0 to 1; therefore, this domain is F range of P is between 0 to 1 closed interval, it can start from 0 it land up 1 therefore, it lies between 0 to 1.

Now, we will come to the third condition, the F consisting of many elements; starting from empty set with the many elements, the last element is Ω. So, if you take mutually disjoint events what is the meaning of mutually disjoint? If you take any 2 elements, if you go for intersection that is going to be the empty set, that is called mutually disjoint event if you take any 2 elements in the F, make intersection that is going to be the empty set.

So, if you take such events mutually disjoint events; it could be finite elements or countable infinite elements, if you take those elements and you verify $P\left(\cup_i A_i\right)$ make an union of those elements; that is also belonging to F because of the σ- field. And the right hand side if you make $\sum_i P(A_i)$ if both the values are going to be same if this condition is satisfied by taking any mutually disjoint events in the F, then you can conclude the said function P which has a domain F, range is 0 to 1 that real valued function we call it as probability, that is a probability.

It is very important because the Ω be a sample space for random experiment whereas, you can have more than one σ- field on the same Ω. For fixed F, we are creating set function that is one set function satisfying these three condition therefore, it is a probability; that means, we can always create some other set function on the same Ω and F, it may satisfy all the three conditions then we can have another probability for the same Ω and F. So, many more probability can be defined by satisfying the, these three conditions for as fixed Ω and F.

Therefore, this triplet; (Ω, F, P), this triplet is called the probability space. Because you can have many probability for a same Ω and F therefore, you can have many probability space can be created for the same random experiment; that is very important concept in probability. Whenever we come across, whenever we make real valued problem using the probability theory and we are trying to solve it; we should have one probability space to solve that problem.

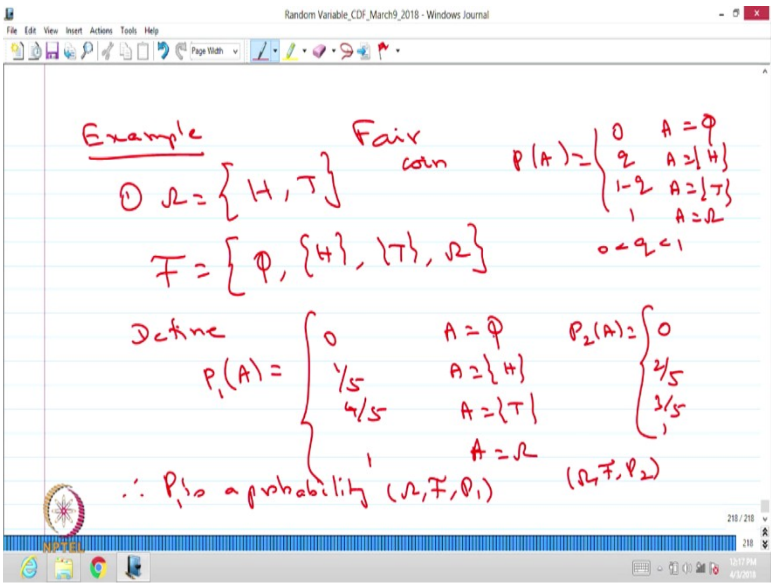
So, that probability space comes from collection of all possible outcomes for fixed F, then for fixed P, you can have different F, you can have different P therefore, you will have different probability space. That means the same problem can be solved in different probability space; that is going to be very important topic in a stochastic calculus or in a financial mathematics, people will solve the problem in different probability space therefore, there results are going to be different based on different probability space.

So, here we have Ω and fixed F fixed P therefore, its one probability space. There is a one more observation since when the P is set function satisfying the first condition and the third condition therefore, this P is called measure. You can always define measure on non empty set with σ- field, you should have non empty set Ω and you create σ- field on and on empty set; then you can define measure satisfying the first condition and the third condition; then that is going to be call it is measure.

Any set function satisfying the first condition and the third condition is going to be call it is measure and here the P is going to be measure because it satisfies it coincides the definition of measure. Therefore, the probability the word P is the probability we say then P is measure also.

Not only that, by seeing the second condition it is additional condition while seeing the concept of measure; it has the value it cannot cross more than 1 therefore, it is called normed measure. Whenever the measure has finite value then it is called normed measure. So, here the finite value is 1 therefore, it is special case of measure which is a normed measure. So, the probability is measure, probability is normed measure because of the second condition.

(Refer Slide Time: 13:30)



We are going to see different probability for the different Ω and F as examples. As a first example, we will take the easiest example that is Ω consisting of H, T; H or T; for F will go for the largest one that is empty set, singleton element and the whole set. I have to create function set function P such that it satisfies the all the three condition of the probability, then

I can conclude that set function is probability. Define P(A) such that when A is going to be empty set; I am going to make it 0 and when A is going to be H, I am going for 1/5, when A is going to be tail it is going to be 4/5.

When A is going to be whole set; 1. You can verify whether all the three conditions of probability definition satisfies. It is always ≥ 0; that is satisfied for all elements A belonging to F that is first condition, second condition P(Ω) = 1 that is satisfied.

If I take a mutually exclusive events in F, then the $P\left(\cup_i A_i\right)=\sum_i P(A$¿¿$i)$¿. Suppose I take the element H and T; so, if make a union. So, that union is going to be **Ω** and P(Ω) is 1; in the right hand side if I take P(H) = 1/5, P(T) = 4/5 if make an addition that is going to be 1.

Not only this, if I take any 2 elements if I make $P\left(\cup_i A_i\right)=\sum_i P(A$¿¿$i)$¿. For any 2 or 3 or any collection of the mutual exclusive events; it satisfies. So, since it satisfies all the three conditions P is probability. Now, we can think of is there any other function also satisfies the same 3 conditions therefore, will have one more probability. So, here this (**Ω**, F, P) that is a probability.

So, let me go for making notation called suffix one; that means, P_1 is probability therefore, (**Ω**, F, P_1) it is probability space. By seeing this numbers, we can think of, we can change the numbers; still you can have a probability that is I will go for immediately P_2¿A); that is going to take the value 0, 2/5, 3/5, 1; the same way A is equal to H, A is equal to tail, A is equal to 1; that means, this is also satisfying the probability conditions.

Therefore, this is also probability space; is there only these 2 we can go for many combination in which I can go for P(A) such that I takes the values 0 with some q; 1 - q and 1 where q is lies between. So, this is for A is equal to empty set and A is equal to H, A is equal to tail and A is equal to **Ω** where q is open interval 0 to 1, where q is open interval 0 to 1; that means, I can have a many probability.

Therefore, I can have a many probability space for this example; whatever we discuss in the elementary level, we always go for fair coin or unbiased coin. In that case the probability of getting head or probability of getting tail or P(A), when A is could H or A is equal to tail; these values are going to be 1/2 and 1/2 then it is corresponding to the random experiment of tossing a fair coin or unbiased coin.

So, whatever we discuss in the school days are the elementary level having a probability 1/2 and 1/2 that is one of the probability space which we have discussed, but in general if the coin is not fair or bias coin; we can have a many more probability space for the same example, by going for different P. And one more observation by different F also we can have a different P's then you can go for different probability space.

I have fixed a F here I am going for different P, but you can change the σ- field then also you will have different P satisfying the condition therefore, you will have different probability space. I forgot to mention this definition is called axiomatic definition of probability, this definition is called axiomatic definition of probability.

This is developed by the probabilistic person name Kolmogorov, is Russian mathematician who contributed quiet a lot for probability theory. Therefore, this definition is called Kolmogorov axiomatic definition of probability and this definition is valid for whatever be the situation you have for random experiment and this definition is valid.

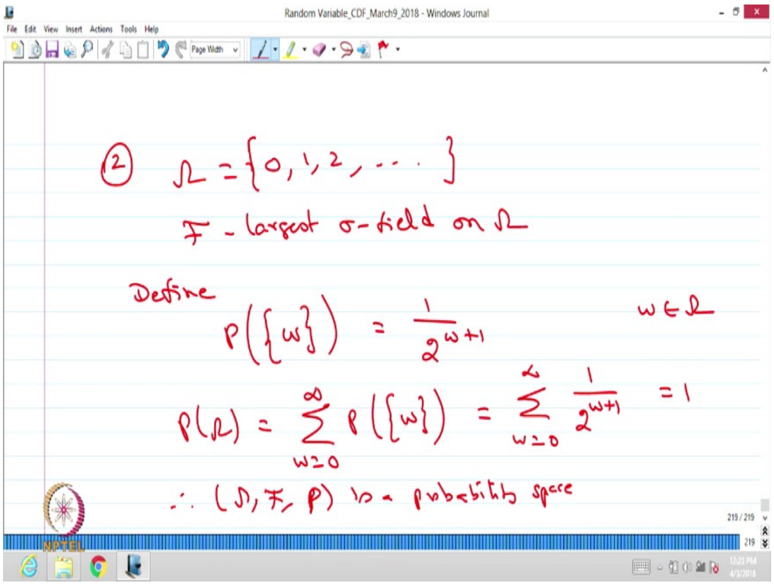
Whereas, there is classical definition of probability which is the special case of axiomatic definition of probability that we are going to discuss after few more examples. This axiomatic definition of probability does not have any functions on random experiment on Ω and F and so, on.

Whereas, the classical definition of probability which is a special case of axiomatic definition of probability that we are going to discuss after few more examples this axiomatic definition of probability does not have any assumptions on a random experiment on Ω and F and so on whereas the classical definition of probability which has some assumptions with that we will go for that easiest definition of probability that is the classical one whereas, this one does not have any assumptions therefore, this is in generally it is true for any set of situations where we have going to apply the probability.

So, this is a very important definition called axiomatic definition of probability; these 3 axioms are called Kolmogorov axioms. These three axioms or condition are called Kolmogorov axioms and the definition of called axiomatic definition of probability.

So, this is easiest example in which I have introduced a many probability; therefore, we get the many probability space for simple example itself. When it is fair coin then the P(A) is going to be 0, 1/2, 1/2, 1; for this element empty set H, tail and Ω.

(Refer Slide Time: 22:46)



Now, we will go for the second example in which we will have a Ω consisting of countably infinite elements. So, Ω is going to have 0, 1, 2 and so on countably infinite elements. F is the largest σ- field on R over Ω. I am not going to list out the elements; that means, it started with an empty set and all single elements; any 2 elements, any 3 elements and so on. So, that is the largest σ- field on Ω.

Now, we are going to define P of singleton element of samples in the Ω itself that is going to be $\frac{1}{2^{w+1}}$, where w is belonging to Ω. Even though the set function is defined from F to 0 to 1; I am defining for this each singleton element in the Ω that for any element in the F nothing, but the subsets of Ω that is a union of a few elements of Ω that is going to be the element of F. Therefore, the same set function can be define it further; P(A) where A is subsets of F. So, this is going to be $\frac{1}{2^{w+1}}$.

Let us have a example; suppose w is equal to 0 then the P(0) is $\frac{1}{2^1}$; $\frac{1}{2}$. Suppose w is equal to 1, then P(1) is $\frac{1}{2^2}$; that is $\frac{1}{4}$. Suppose w is going to be 3, then it is going to be 3 then it is going to be $\frac{1}{2^4}$. We can verify whether this is going to satisfy all the 3 axioms of axiomatic definition of probability.

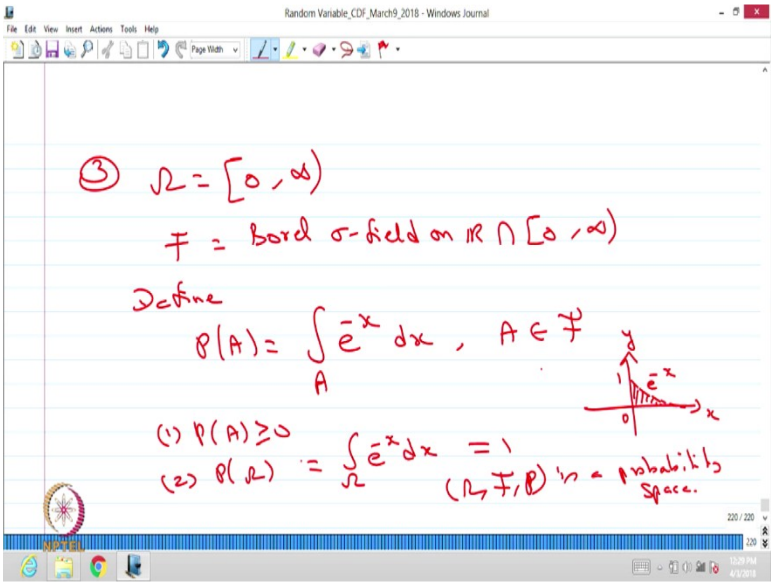
So, the first condition is going to be ≥ 0. So, whatever be the element of w belonging to Ω; this value $\frac{1}{2^{w+1}}$ that is going to be ≥ 0. And P(Ω) you can verify the second condition, P(Ω) = $\sum_{w=0}^{\infty} P\{w\}$. That is same as P(Ω) because P(Ω) is nothing, but the union of all the elements that is same as $\sum_{w=0}^{\infty} P\{w\}$, $\frac{1}{2^{w+1}}$.

So, the first element is $\frac{1}{2}$, second element is $\frac{1}{4}$, third element is $\frac{1}{8}$ and so, on; if you add all the elements it is going to be 1. If you take any disjoint, mutually disjoint events nothing but disjoint samples and find out the P of union that is same as a summation of P of singleton elements. That is same concept I used it in finding the P(Ω) also, I have used the P(Ω) is nothing but the summation of P of individual elements I am using the third property; so, third property also satisfied.

So, all the 3 properties are satisfied all the 3 conditions or all the axioms are satisfied therefore; the P is going to be probabilities. Therefore, (Ω, F, P) is probability space; again the way I define the P(Ω) = $\frac{1}{2^{w+1}}$, someone can think why I has to be this number can I chose some other number so, that I can have all the 3 conditions are satisfied? Yes, you can always create another real valued function so, that all the 3 conditions satisfied; we can have probability.

So, I have just created easiest one that is $\frac{1}{2^{w+1}}$ because I know that the summation of this is going to be convergent series that summation is going to be 1 therefore, I have created probability over this. We will go for the third example that is Ω consisting of uncountably many elements that is 0 to ∞.

(Refer Slide Time: 28:14)



So, the σ- field that is Borel; Borel σ- field on the real line which intersect 0 to ∞ correct. Creating Borel σ- field on real line intersecting with 0 to ∞ that is going to be the largest σ- field on Ω, where $\Omega = [0, \infty]$.

I am going to define a set function P such that this is going to be the probability. The easiest one that is $\int_A e^{-x} dx$, where A is belonging to F; A is belonging to F. The F consisting of singleton element; sorry empty set, singleton element and many intervals between 0 to ∞ which is in the same form of Borel σ- field on real line.

Therefore, the integration over A that is Riemann integral; you know that if A is going to be a singleton element that is going to be 0. If A is going to be closed interval or open interval because it is a Riemann integral; the answer is going to be $\int_A e^{-x} dx$. So, always the first condition is going to be satisfied that is $P(A) \geq 0$. The second condition we can verify $P(\Omega)$ that is nothing, but $\int_\Omega e^{-x} dx$ that is nothing, but 1.

So, it start from 1, it start from 1; so this is $\int_{\Omega} e^{-x}dx$, Ω is from 0 to ∞. So, $\int_{0}^{\infty} e^{-x}dx$is nothing, but area below the curve e^{-x} between the interval 0 to ∞, it is asymptotically 0 at ∞. So, if you find the area that is going to be 1; the third condition if you take a mutually disjoint events in F that is nothing, but if you take non overlapping intervals and find the area in those intervals and find out the intervals separately and sum it up; both are going to be same.

Whether you go for union of non overlapping intervals, find out the P or finding out individually and sum it up both are going to be same; for any mutually disjoint events that is nothing, but the intervals. So, the third condition also satisfied therefore, we can conclude this P is going to be a probability. We can have many more P for the same Ω and F. So, this is the easiest one (Ω, F, P) that is probability space.

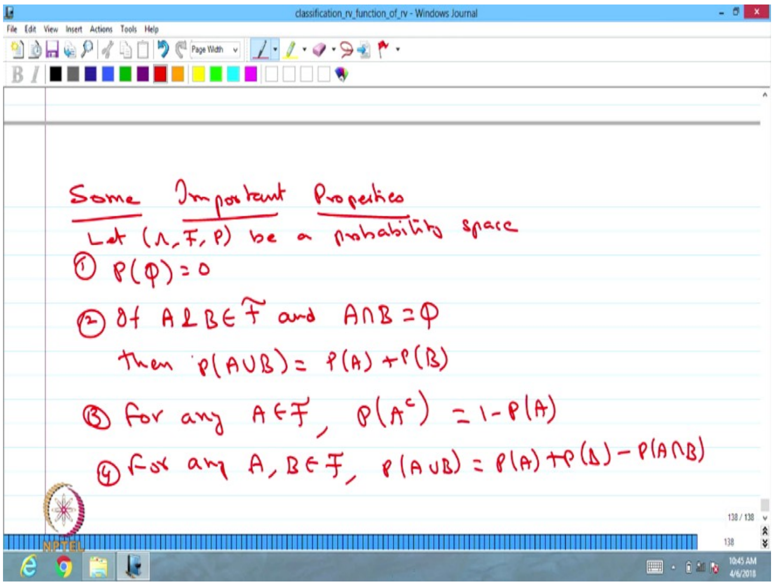
With these three examples one is with finite, other one with countably ∞ elements and the third one with uncountably many elements of Ω; we have created different function P and we got the probability space.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 03
Important properties of a Probability Space

Now, we are going to discuss some important properties of the probability.

(Refer Slide Time: 00:07)



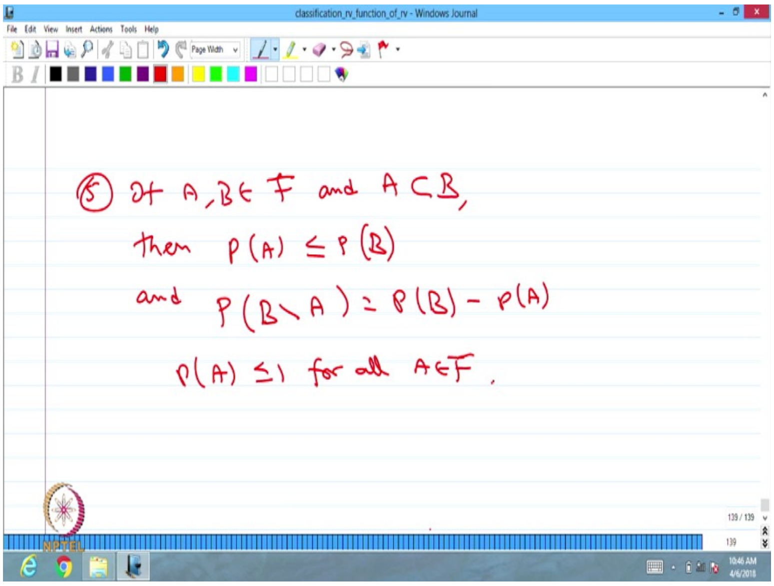
Some important properties: the first property. So, before that let me start with let (Ω, F, P) be a probability space, the first properties P(φ) = 0; in the probability is space definition we have a P(Ω) = 1. So, one can prove P(φ) = 0, because φ is the negation of the whole set the whole set probability is 1. Therefore, this probability is going to be 0.

Second property if a two events A and B belonging to F and both are mutually disjoint events, intersection of events is empty. Then if you want to find out the P(A∪B) = P(A) + P(B). If 2 events are mutually exclusive mutually disjoint events, then the P(A∪B) = P(A) + P(B) otherwise in general for any 2 events that is P(A∪B) = P(A) + P(B) - P(A∩B). Since A ∩B = φ and the P(φ) = 0.

Therefore, P(A∪B) = P(A) + P(B). Third result for any event A belonging to F, $P(A^c) = 1 - P(A)$ if you know the probability of A, if you want find out the $P(A^c) = 1 - P(A)$, the same concept is easy to find out the P(φ) = 0 also.

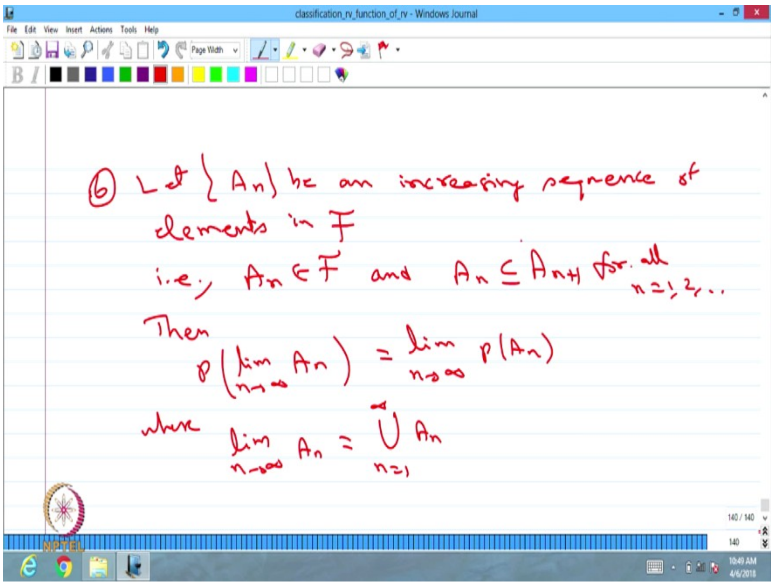
The fourth result, for any 2 events A, B belonging to F, you can always have a P(A∪B) = P(A) + P(B) - P(A∩B). As in the result 2 when A∩B = φ, then it is going to be just P(A) + P(B) there is no - P(A∩B).

(Refer Slide Time: 03:19)



The next result, site number 5 if A, B belonging to the F and A *contained* B, then one can conclude P(A) ≤ P(B). Not only that P(B/A) = P(B) – P(A), in particular P(A) ≤ 1 for all A belonging to F.

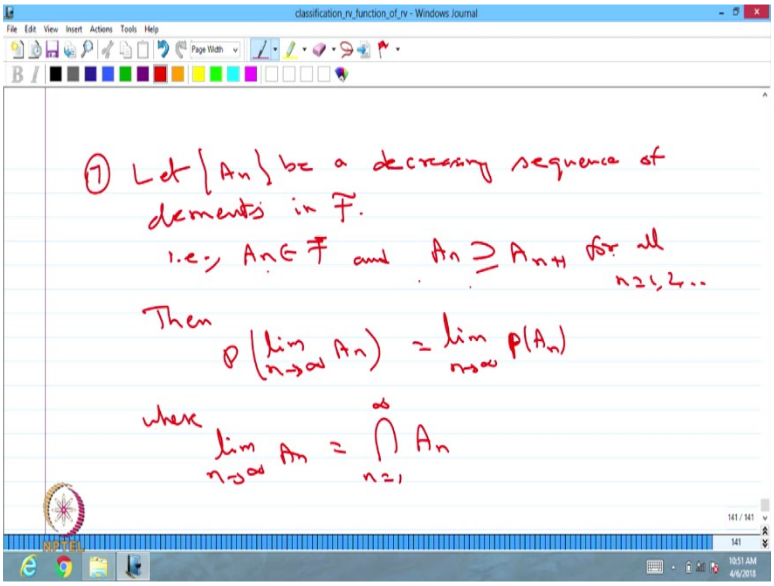
(Refer Slide Time: 04:11)



Then next result let A_n be an increasing sequence of events or elements in F, the element in the F is nothing but the events. So, let A and B an increasing sequence of a elements in F.

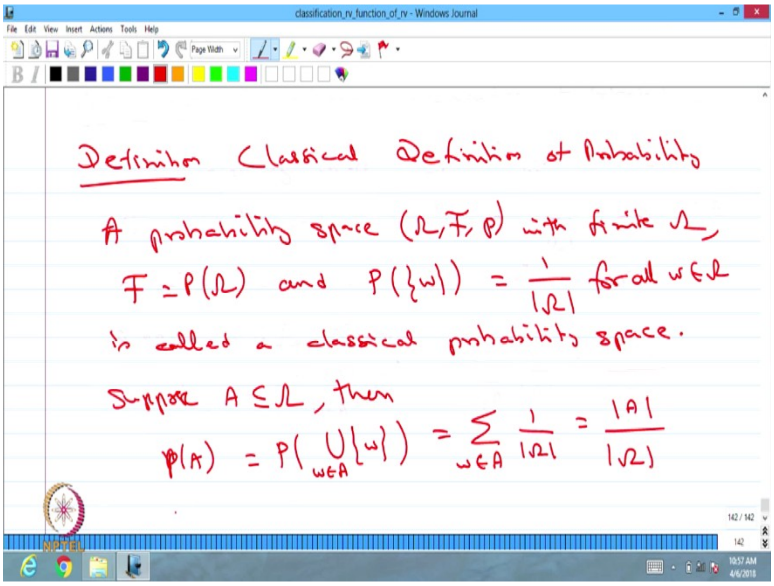
That is if you take A_n belonging to F and A_n is contained in A_{n+1} for all n =1, 2, 3 and so on that is the meaning of A_n be a increasing sequences of elements F. Then what the result says then $P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A¿¿n)¿$ where the $\lim_{n\to\infty} A_n$ is nothing but; since $A_n$are the increasing sequence of elements $\lim_{n\to\infty} A_n = ¿n=1¿\infty A_n$. Similarly one can go for sequence of a elements which are a decreasing.

(Refer Slide Time: 06:02)



So, the next result or next property that is let A_n be decreasing, a decreasing sequence of elements in F that is A_n belonging to F and it satisfies for all n is equal to 1, 2 and so on. In this case P($\lim_{n\to\infty} A_n$) = $\lim_{n\to\infty} P(A¿¿n)¿$, where the $\lim_{n\to\infty} A_n = ¿n=1¿\infty A_n$, because $A_n^{'}$sare the decreasing sequences of elements in F and not giving the proof of this results, but it can be proved easily.

(Refer Slide Time: 07:28)



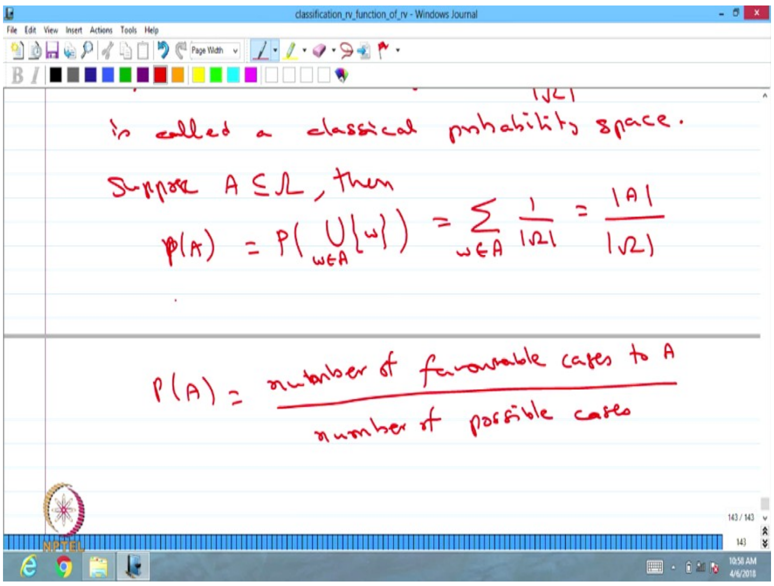
Now, we are moving into the next concept or next definition that is called a classic definition of probability. A probability space that is (Ω, F, P) with finite Ω, the F is equal to power set of a Ω the P(Ω) means power of Ω and the probability measure of a singleton element w, that is equal to 1 divided by the number of elements in the Ω for all w belonging to Ω, that is called a classical probability ,a classical probability space.

Because we are defining a probability space here probability space (Ω, F, P), with the conditions on Ω, F as well as P, which basically as special case of a probability space by satisfying Ω is finite and F is a largest σ- of field which is a power set of Ω and the probability measured P is defined on each possible outcome as a 1 divided by total number of elements in the Ω, which is finite for all w belonging to Ω. If these three conditions are satisfied by a probability space, then that probability space is called a classical probability space, there is the another name for these that is called Laplace probability space.

In this situation the probability P is called a classical probability, the probability measure P which is defined in this probability space with Ω is finite and F is a power set of Ω and P on each possible outcome is equally likely that is 1 divided by number of elements the Ω. Then this probability measure is called a classical probability measure and the probability space is called a classical probability space. So, the classical probability space is a special case of the probability space, which we have defined it earlier that is called axiomatic definition of probability the special case is the classical probability space.

Since, P on each possible outcome is going to be a 1 divided by total number of elements in the Ω. So, the probability of any event is nothing but number of favorable cases on event A divided by the total possible outcomes that is a whatever the finite number. Suppose, A is contained in Ω where A is event then P(A) is nothing but a P of union of a singleton elements samples, where all the samples belonging to A the event that is nothing but summation of w belonging to A, $\frac{|A|}{|\Omega|}$

(Refer Slide Time: 11:52)



In other words, P(A) is nothing but number of favorable cases to the event A divided by number of a possible cases. So, this is what we have done it in the school days when we are computing the probability of event that is nothing but the number of favorable cases to the event A divided by the total number of all possible outcomes.

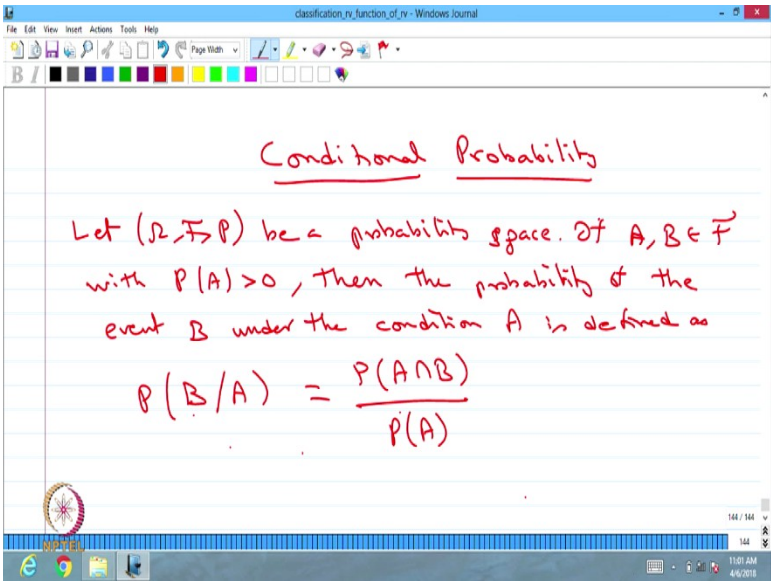
So, this is going to be the classical probability, with the assumption that Ω is finite and σ-field is the largest σ- field which is nothing but the power set and equally probable outcomes.

That means, P of each possible outcome that is same as the $\frac{1}{|\Omega|}$. So, this is called the classical probability space.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 04
Conditional probability

(Refer Slide Time: 00:08)

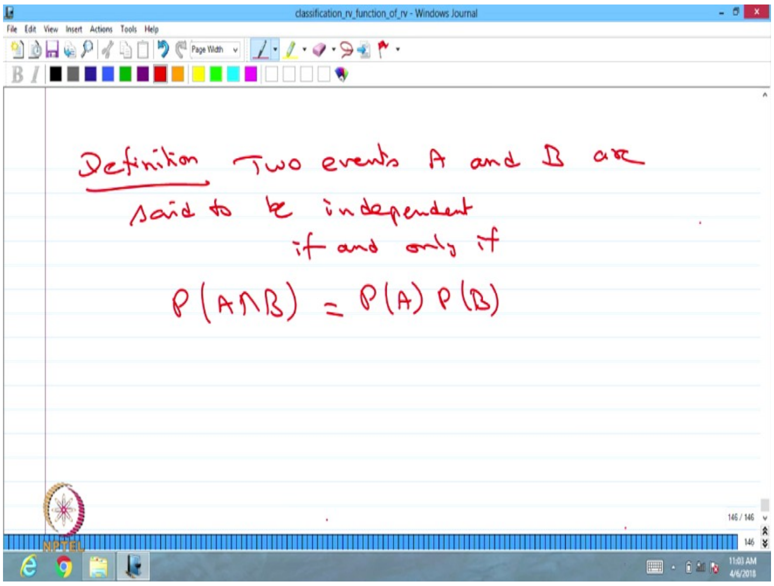


Now, we will move into the next topic that is called conditional probability. Let (Ω, F, P) be a probability space, we are not making any assumption over this probability space. If A, B belonging to F both are events with the P(A) > 0; that means, it is not impossible events impossible events means the probability with that event is equal to 0 and, sure event means the probability of that event is equal to 1, rare event means the probability of a that event is open interval between 0 to 1 so, here it is a non impossible event.

Then the probability of the event B, under the condition the event A, that is defined as P(B/A) $= \frac{P(A \cap B)}{P(A)}$, this is well defined, because the P(A) > 0.

So, this condition probability is going to be P(B/A) = $\frac{P(A \cap B)}{P(A)}$, this is called the conditional probability of event B given event A. Few properties on the conditional probability, but before that let me give the definition of independent events, then I will come to the results on the conditional probability.

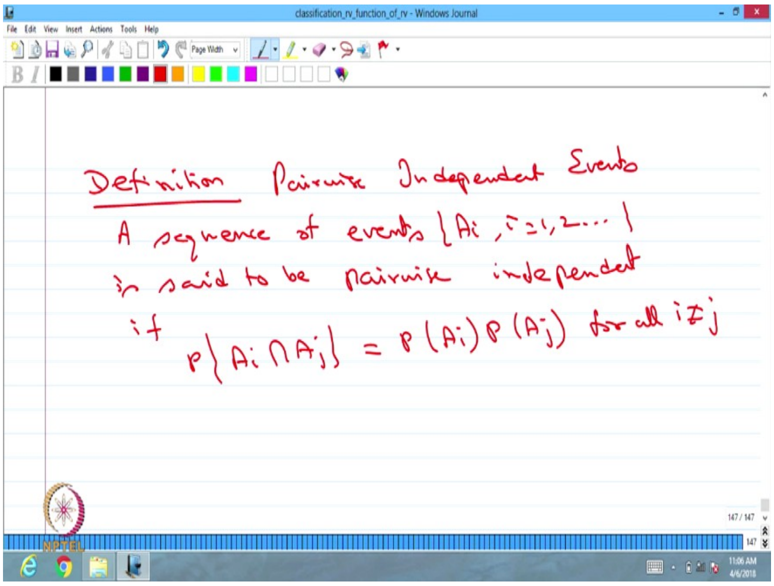
(Refer Slide Time: 02:43)



The first definition two events A and B are said to be independent if, and only if the intersection probability is same as the product of the individual probabilities.

We say if and only if condition, whenever this condition is satisfied by any two events, then we call it as both the events are independent. If two events are independent, then this condition will be satisfied, based on this I am going to give a two definition, we always called pairwise independent and other one is called mutual independent.

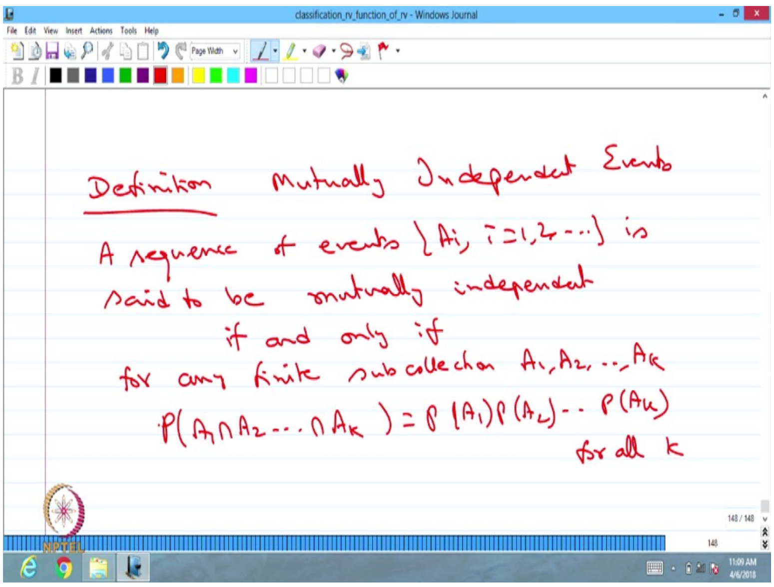
(Refer Slide Time: 03:45)



Then its definition that is a pairwise independent events A sequence of events A_i, where i is equal to 1, 2, so on is said to be pair wise independent, if you take any two events that satisfies the independent property, that is a $P(A_i \cap A_j) = P(A_i)P(A_j)$, for all i $\neq$ j.

So, if you take any two events that satisfies the independent condition, then we conclude this sequence of events or this collection of events are called pairwise independent, if it satisfies the independent condition.

(Refer Slide Time: 05:08)

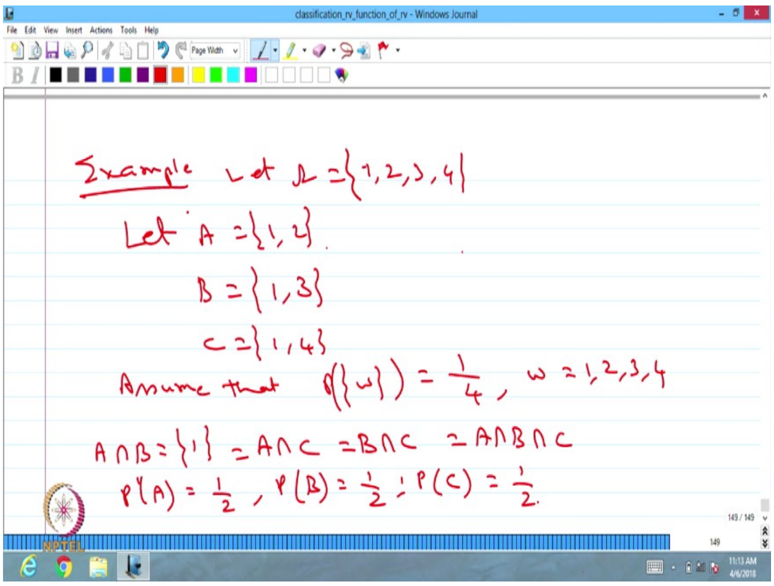


The next definition, that is mutually independent events, A sequence of events, A_i's is said to be said to be mutually independent, mutually independent if and only if for any finite, for any finite sub collection that is a A_1, A_2 so on A_k.

The $P(A_1 \cap A_2 \cap \ldots \cap A_k) = P(A_1)P(A_2)\ldots P(A_k)$for all k, that is very important that means, if you take k equal to 2, so if you take any 2 events in this collection that satisfies independent property, or independent condition. Or, if you take any 3 events that also satisfies the independent condition, like that if you take all the possible number of events, that is the sub collection from that collection of events, satisfies the independent property. Then we conclude they are mutually independent events sometimes, we would not use the word mutually, if you say that a few events are independent; that means, by default they are mutually independent events.

You see the previous definition that is the pairwise independent events, but that is satisfied only for any two events not 3 or 4 or 5, independent properties not going to be satisfied; that means, the mutually independent events implies their pairwise independent, but the pairwise independent need not imply the mutually independent events, for the collection of events.

(Refer Slide Time: 08:00)



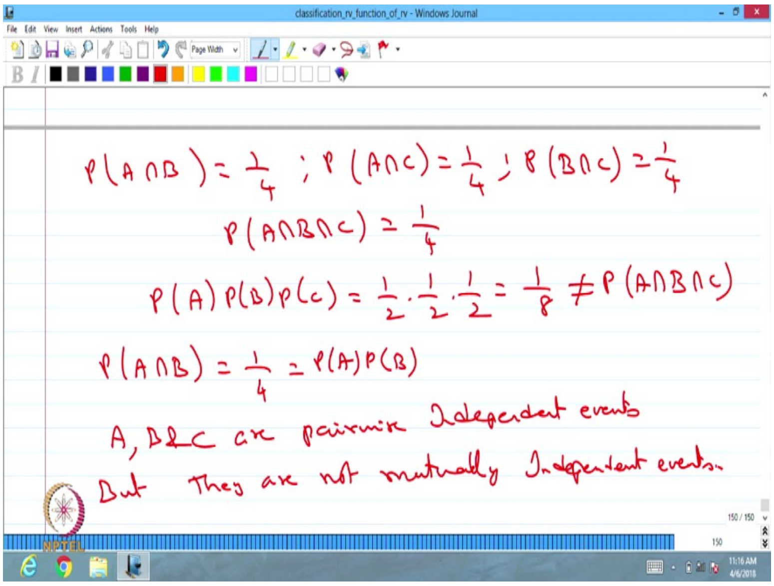
So, let us go for one simple example for pairwise independent and mutually independent, then we will move into the properties on the conditional probability, so the example as follows. Let Ω consisting of 1, 2, 3, 4. Let A be the event consisting of only two elements 1 and 2.

Similarly, B is the element consisting of 1 and 3 and C is the element consisting of 1 and 4. And we assume that the probability of singleton element w that is equal to 1 divided by 4, for w is equal to either 1 or 2 or 3 or 4.

Now, we will verify whether this event satisfies the pairwise independent as well as mutually independent, or only satisfies pairwise independent not the mutually independent. Let us go for finding $A \cap B = \{1\}$. And similarly, if you go for $A \cap C = \{1\}$.

If you go for $A \cap B \cap C = \{1\}$., where as you compute the P(A)¿$\frac{2}{4}, that\, is \frac{1}{2}$ and P(B) ¿$\frac{2}{4}\, that\, is \frac{1}{2}$ and P(C), that is also again $\frac{2}{4}, that\, is \frac{1}{2}$.

(Refer Slide Time: 10:11)



Now, if you compute $P(A \cap B) = \frac{1}{4}$. Similarly, if you compute $P(A \cap C) = \frac{1}{4}$; similarly, $P(B \cap C) = \frac{1}{4}$ whereas, $P(A \cap B \cap C) = \frac{1}{4}$.

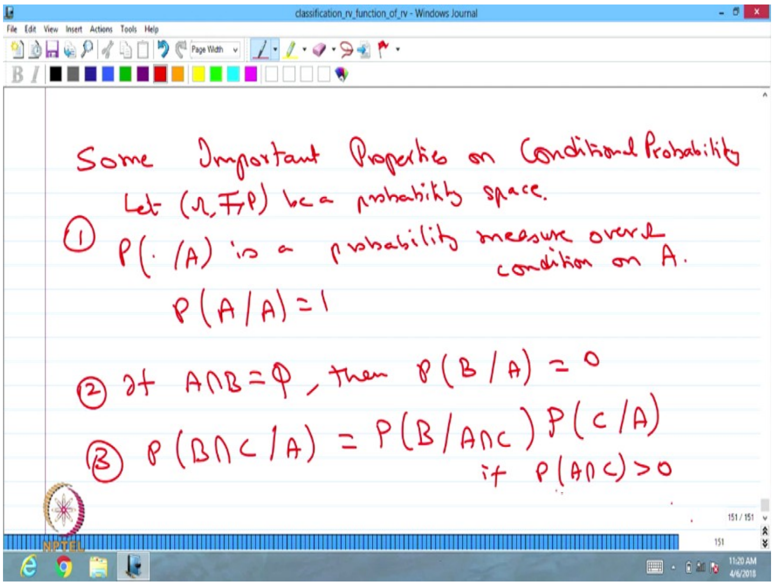
For example if you test P(A)P(B)P(C) ¿$\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$, which is not same as P$(A \cap B \cap C)$.

Whereas A if you go for P$(A \cap B) = \frac{1}{4} =$¿ P(A)P(B).

Similarly, P$(A \cap C)=¿$ P(A)P(C). Similarly, P$(B \cap C)=¿$ P(B)P(C) that means, it satisfies the independent condition for any two events not for the all 3 events that means, in this example A, B and C, are pair wise independent events.

But they are not mutually independent events, because it does not satisfies independent property for three events. So, this very simple example conclude pairwise does not imply the mutually independent events in general.

(Refer Slide Time: 12:32)



Now, let us go for some properties on conditional probability, some important properties on conditional probability. The first one so, before that we have probability space, let (Ω, F, P) probability space, in that probability space we going to give the properties. The first one, P of the conditional probability that is probability measure.

The conditional probability that is also a probability measure over Ω condition on the event A that means, it satisfies the axiomatic properties of the probability, that is a three properties P(A) ≥ 0 for all A belonging to F. And P of whole set is 1 and, mutually disjoint events $P(¿ i A_i)=\sum_i P(A_i)$.

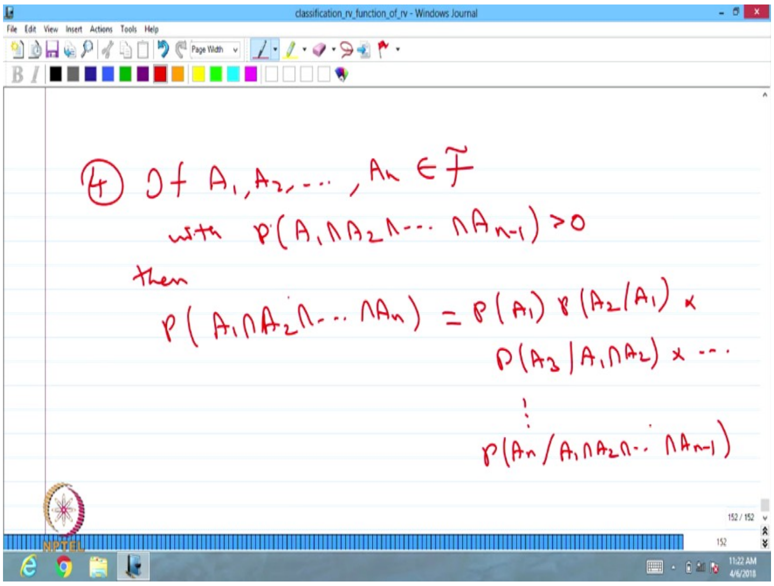
The same thing here also satisfied here, the P(A/A) = 1, where as in the axiomatic definition probability the P(Ω) = 1 here, the conditional probability that is also probability space here

P(A/A) = 1. Second important property that is if $A\cap B=\varphi$, then the conditional probability P(B/A) that is doing to be because same as $\frac{P(A\cap B)}{P(A)}$.

Since, $A\cap B=\varphi$, P(φ) = 0 therefore, P(B/A) = 0, third result or third property $P(B\cap C/A)=P(B/A\cap C)P(C/A)$. If $P(A\cap C)>¿\,0$.

Whenever you want to compute a $P(B\cap C/A)$, that can be computed in the form of $P(B/A\cap C)P(C/A)$, whenever $P(A\cap C)>¿\,0$.

(Refer Slide Time: 16:16)

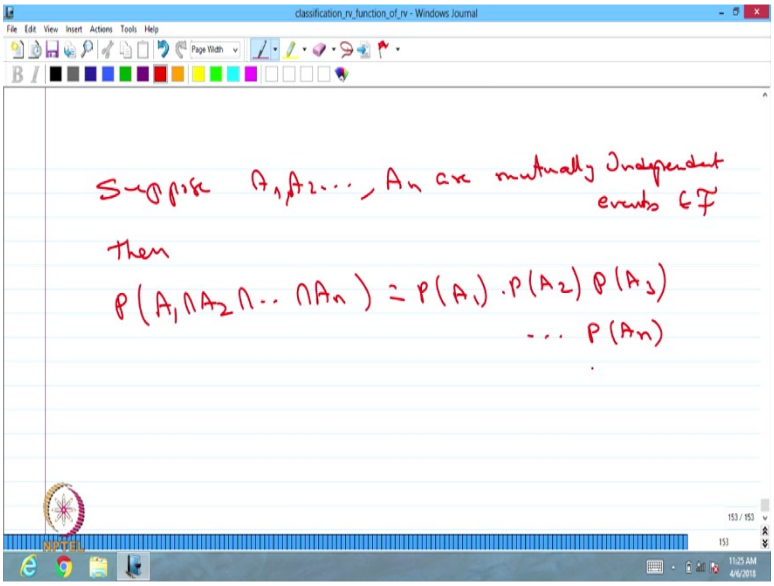


The next result related to the conditional probability. If A_1, A_2 and so on, A_n, n events belonging to F with $P(A_1\cap A_2\cap\ldots\cap A_{n-1})$, if that probability is larger than 0.

Then one can define probability of intersection of n events the assumption is A intersection of n -1 events probability greater than 0, we have n events belong F. Then one can find the probability of intersection of n event is same as $P(A_1)P(A_2¿A_1)\ P(A_3/A_1\cap A_2)\ldots$. The last element is $P(A_n/A_1\cap A_2\cap\ldots\cap A_{n-1})$, This is possible whenever you have A_n events and, the probability of intersection of n events provided probability of intersection of n -1 events is greater than 0, one can always find the product of condition probability with the interaction of the events. This is valid for any countable number of events satisfying this condition, there no assumption over the events to apply this result.

Now, suppose these events are mutually independent, or mutually independent events, then the results are going to be simplified as follows.

(Refer Slide Time: 18:34)



That means, suppose $A_1, A_2, \ldots, A_n$are mutually independent events, all are belonging to F. Suppose, these events are mutually independent event belonging to F.

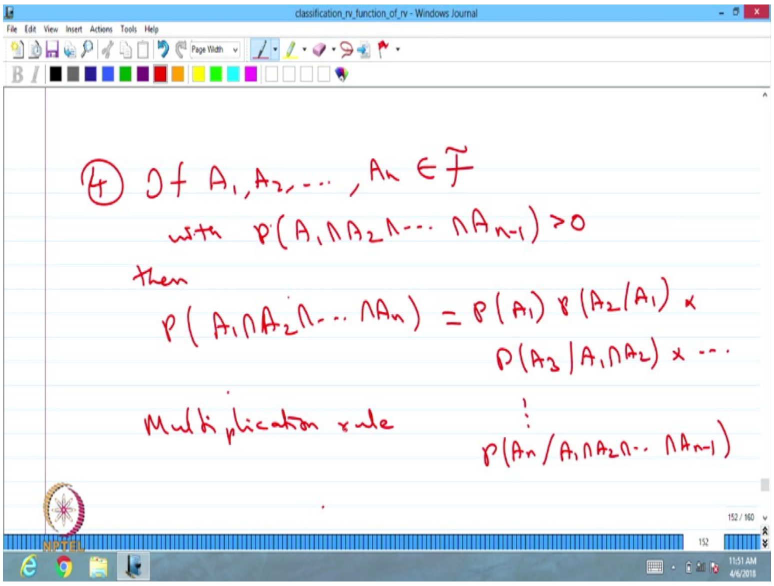
Then the P $(A_1 \cap A_2 \cap \ldots \cap A_n)$, that is same as, we are retaining the other condition intersection of P $(A_1 \cap A_2 \cap \ldots \cap A_{n-1})$, n - 1 events that is probability is greater than 0, we are keeping that in P$(A_1 \cap A_2 \cap \ldots \cap A_n)$,. That is same as you see the previous one, first one is $P(A_1)$, the second 1 $P(A_2 | A_1)$. Since $A_1, \wedge A_2 \wedge A_n{}'s$ are mutually independent events the intersection.

The condition $P(A_2 | A_1)$ = $P(A_2)$ therefore, this is going to be $P(A_2)$. Similarly, $P(A_3 / A_1 \cap A_2)$ = $P(A_3)$ and so on; the last one it is $P(A_n)$. That means, if events are mutually independent, then the intersection probability same as the product of individual probabilities. So, this can be proved also. So I am just using that result.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

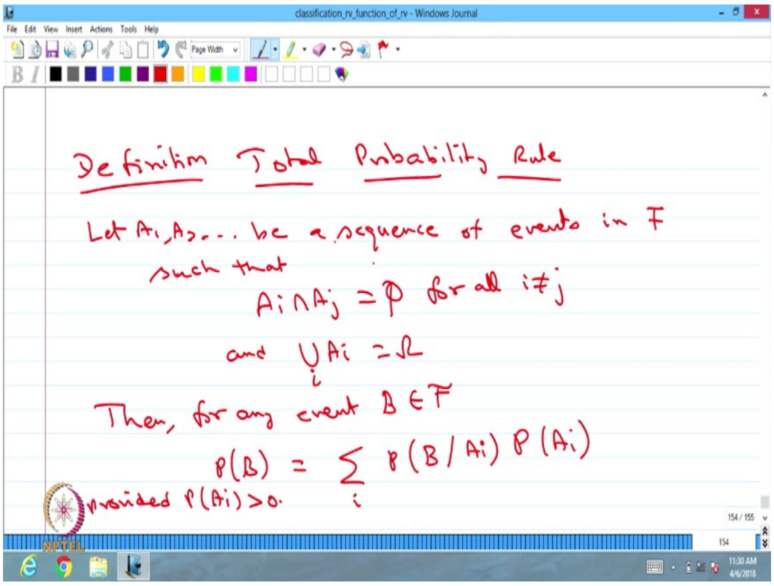
Lecture – 05
Conditional Probability and Bayes Rule

(Refer Slide Time: 00:01)



This result, this probability of intersection of events is same as the product of this probability. This rule is called multiplication rule. Later I am going to introduce the total probability rule, whereas, this one is called multiplication rule that is for the probability of intersection of events is same as the probability of individual events.

(Refer Slide Time: 00:33)

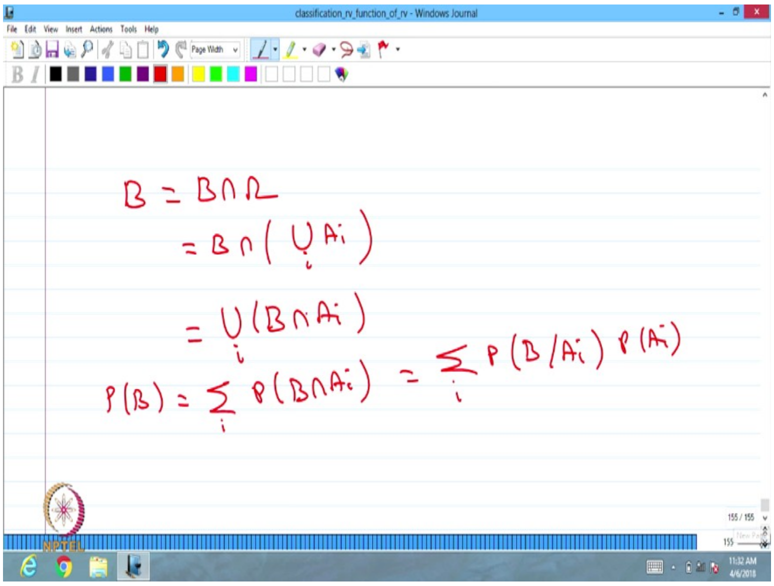


Now we are moving into the next important definition before going to the Bayes theorem; that is called total probability rule or we can give it as a theorem also. So, this called total probability rule, what it says, let A_1, A_2, so on, be a sequence of events in F; such that $A_i \cap A_j = \varphi$,; that means, their mutually disjoint for all i is not equal to j not only that if a make a union of A i's that is going to be the whole set; that means, indirectly this is partition of Ω. This sequence of events could be finite or countably infinite.

Therefore, I making union of i's union over i it could be finite or countably infinite such that they are mutually disjoint union is going to be the whole set, that is basically partition events. Indirectly what you are saying is their mutually exhaustive also, these events are mutually exhaustive satisfying this condition, then for any event capital B belonging to F $P(B) = \sum_i P(B/A_i)P(A_i)$. This summation over i provided $P(A_i)$'s > 0.

So, the total probability rule says if you have partition events on Ω, the probability can be computed for any event with the conditional probability on partition events multiplied by the probability of those partition events for all i's. This is going to be call it as total probability rule. You can give the proof of this, even though I made it as a definition, we can prove it.

(Refer Slide Time: 03:46)

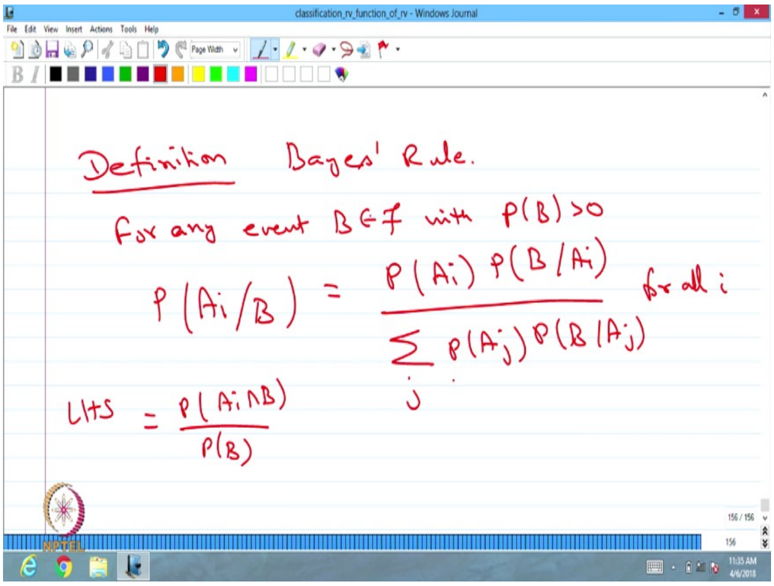


Then any event B can be written has $B \cap \Omega$ that is same as we can write rewrite event B intersection, the Ω ¿¿$i\, A_i$. Then this is same as ¿$i(B \cap A$¿¿$i)$¿.

Therefore, the P(B) is nothing but, since B$(B \cap A$¿¿$i)$¿'s are mutually disjoints. Therefore, this is same as $\sum_i (B \cap A$¿¿$i)$¿, that is same as you can apply the conditional probability, therefore, it is going to be $P(B/A_i)P(A_i)$. This is possible when $P(A_i)' s > 0$.

So, whenever the $P(A_i)$'s > 0, then $P(B) = \sum_i P(B/A_i)P(A_i)$. Using this total probability rule, we are going to define next rule that is called Bayes rule, as a next definition, that is called Bayes rule.

(Refer Slide Time: 05:17)

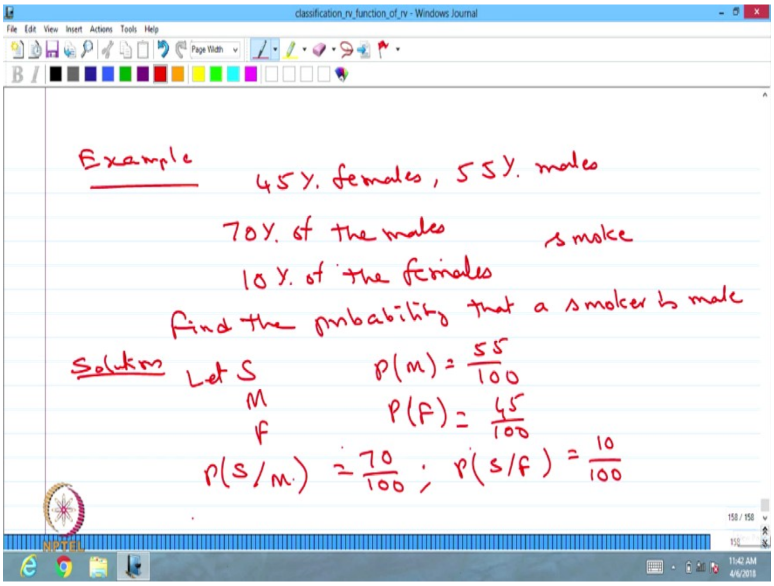


What this rule says, the continuation of the total probability rule that is for any event B belonging to F with; the P(B) > 0, already we made the assumption $P(A_i)' s >$ ¿ 0, now we are making P(B) > 0. In that case, one can find the $P(A_i/B) = \frac{P(A_i) P(B/A_i)}{\sum_j P(B/A_j) P(A_j)}$. This is valid for all i. Because we have already defined A_i's are the partition events of Ω.

So, you can go for finding the condition P$\left(\frac{A_i}{B}\right)$provided P(B) > 0. That is same as the numerator is $P(A_i) P\left(\frac{B}{A_i}\right)$, whereas the denominator is all possible values of the same thing in the numerator. Therefore, this ratio is going to be P$\left(\frac{A_i}{B}\right)$you can prove it easily, the left hand side is same as $\frac{P(A_i \cap B)}{P(B)}$.

So, in the numerator you can apply the conditional probability and the denominated you can use total probability rule therefore, you can get the answer. So, these are very important result on whenever you know the partition events, you can find the probability of any event using total probability rule as long as you know the conditional probability on partition events, and the individual probability of A_i's partition events. Then using the same result, you can find out the condition probability of partition event given any event B provided P(B) > 0.

(Refer Slide Time: 08:18)

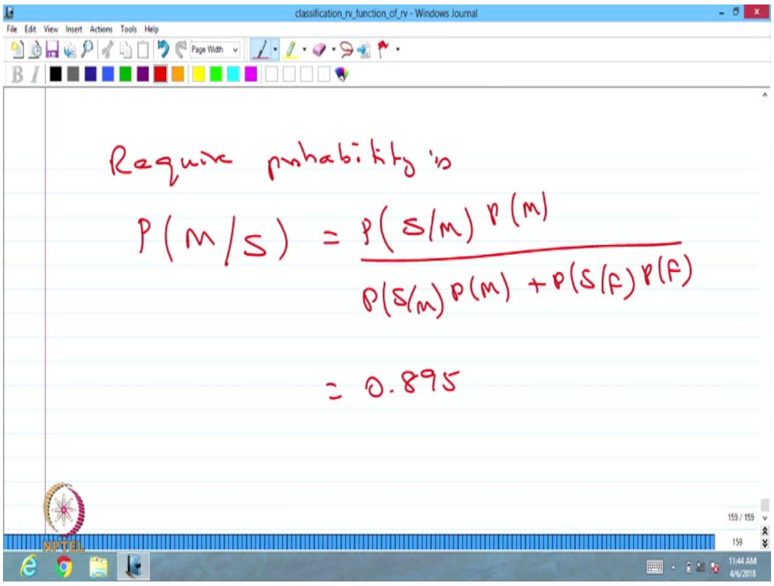


We will go for the simple example, it is known that the population of certain city consisting of 45 percentage females, and 55 percentage males.

Suppose that 70 percentage of the males, and the 10 percentage of the females smoke, suppose that 70 percentage of the males and 10 percentage of the female's smoke, the question is find the probability that find the probability that a smoker is a male. Let me read the question again. It is known that the population of the certain city consisting of 45 percentage females and 55 percentage males. Suppose that 70 percentage of the males and 10 percentage of the female's smoke, find the probability that smoker is male. Let us solve this problem by treating, let S be the event S be the event that the person is smoker.

Let M be the event that the person is male, F be the event that person is female. Let S be the event that person is smoker, M be the event that person is male, F be the event the person is female. Therefore, you can find the P(M), that is 55 percentage therefore, 55 by 100, P(F) that is 45 percentage. So, 45 by 100, and you can find the P(S/M) that is 70 divided by 100. Similarly, you can find P(S/F) that is also given, that is 10 percentage therefore, 10 by 100 these are all the given information.

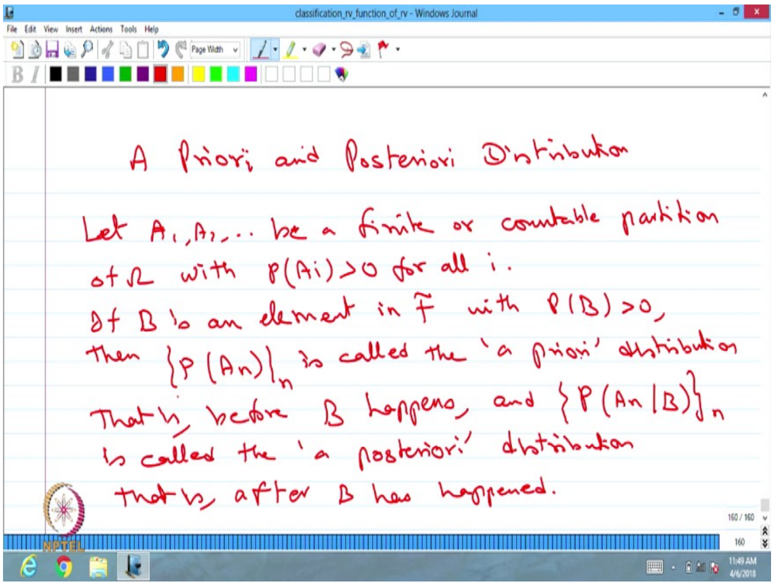
(Refer Slide Time: 11:05)



Now, the required probability is P(M/S), that is same as $\frac{P(S/M)P(M)}{P(S/M)P(M)+P(S/F)P(F)}$.

So, you can substitute all the values then you can get the answer that is 0.895 by substituting all the probability values you can get the P(M/S). So, by using the total probability rule as well as the Bayes rule we are getting the result of P(M/S).

(Refer Slide Time: 12:14)



So, there is very important result on using the Bayes rule, that is called A priori and posteriori distribution, there is very important concept on A priori and posteriori distribution. What it says? Let A_1, A_2 and so on be a finite or countable partition of Ω; that means, intersection is empty and ¿ i A_i = Ω with P(A_i) > 0 for all i , it is basically partition events of Ω.

If B is an event if B is an element or event in F with the P(B) > 0, then the distribution or the P(A_n) for all n, that is called the priori distribution. Because this information is known to you earlier, therefore, this is called priori distribution knowing the probability of partition events for all n.

That is before B happens and the P(A_n/B) this collection for all n, that is called the posteriori distribution. Because this distribution is known after B happens, whereas P(A_n)'s is a before B happens, therefore, it is called priori distribution and P(A_n/B) that is after B happens therefore, it is called posteriori distribution, that is after B has happened.

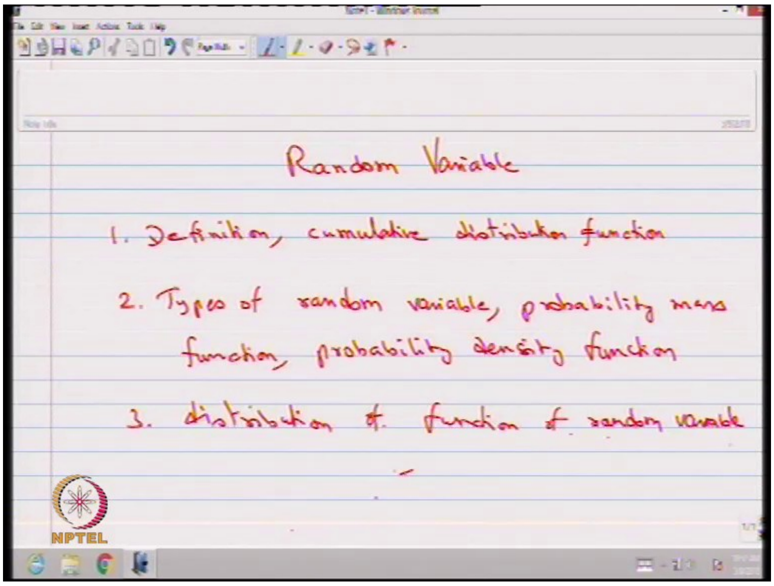
So, this is very important result on probability, the priori and posteriori distribution based on Bayes rule, the one is call the probability of partition events these are called the priori distribution, and after the event B happens, P(A_n/B) for all n, that is going to be called as posteriori distribution; This as the wide application in statistics.

So, with these we are completing basic, basics of probability starting from random experiment, sample space probability space probability measure. Then some results on the probability space, then independent events, then mutually independent and the pair wise independent events, then after that we have introduce conditional probability, then we have introduced two important result one is total probability rule, and the multiplication rule. Then finally, we have discussed the Bayes rule.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 02
Random Variable
Lecture - 06

(Refer Slide Time: 00:17)

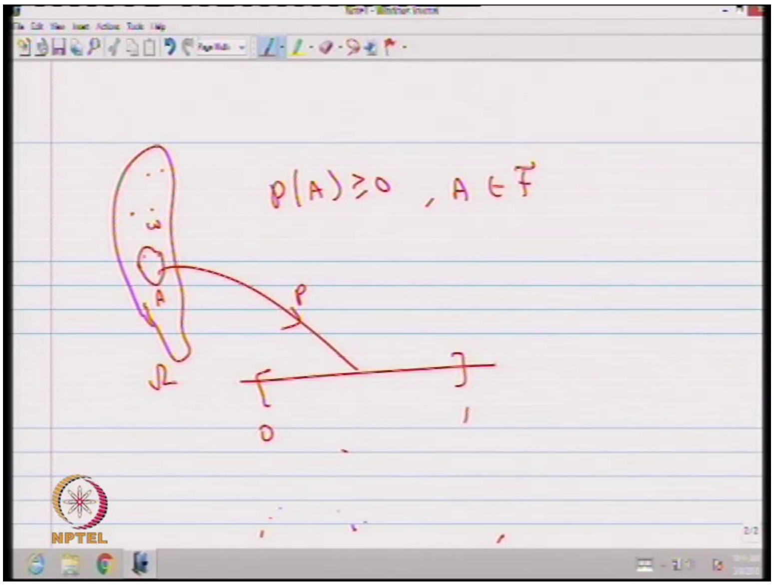


In this topic we are going to discuss three things in three different lectures. In the first lecture which we are going to discuss the definition of a random variable and then we are going to discuss cumulative distribution function of a random variable. So, this is going to be the lecture 1 of week 2, and the second lecture which we are going to discuss i.e., types of a random variable, followed by we are going to discuss probability mass function, we are going to discuss probability mass function. Then we are going to discuss probability density function.

In the third lecture which we are going to discuss, i.e., distribution of a function of random variable. So, this is going to be the week 2, 3 lectures. The first lecture is definition of a random variable and then a cumulative distribution function. And the second lecture is types of random variable, probability mass function and probability density function and the third lecture is a distribution of a function of random variable.

Let us start with the first lecture, definition of random variable. This plays an important role of describing a random data therefore, the random variable is very important concept in the probability theory.

(Refer Slide Time: 02:38)



As you know in the random experiment you have a collection of a possible outcomes which you denoted by w's, and the collection of possible outcomes is a sample space **Ω**, which consist of all possible random sample or sample which you denoted by w's. Whenever we defined the probability of an event, the event is a collection of an elements of **Ω** that is nothing but the event.

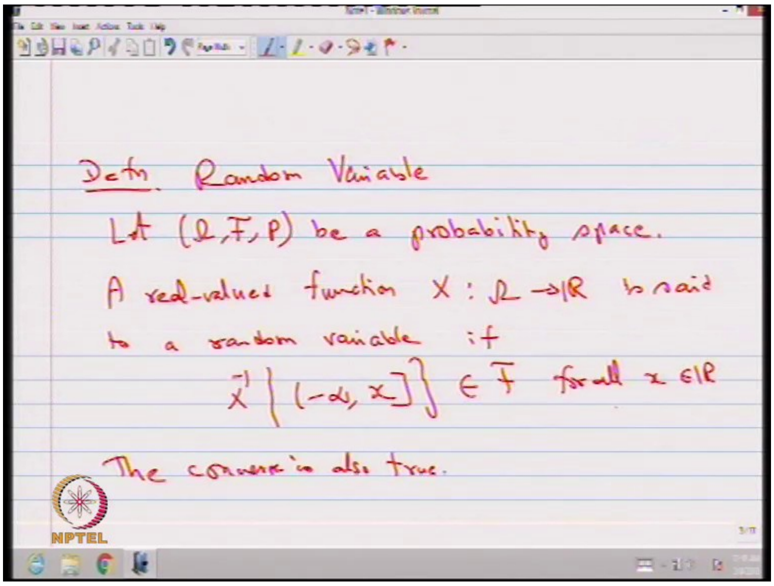
So, the probability of event A is greater than or equal to 0 where A is belonging to the sigma field, which we discuss in the first week. Suppose, you know the probability of event you can find out the probability of a compliment of event, probability of a collection of a few events as a union of or so many form of a probability of event you can find, but the problem with event is, event with the probability definition you need all the set operations. The set operation means union, compliment, intersection. But this is going to be very tedious job when you have a very complicated example in which the collection of possible outcomes is huge. Therefore, the sigma field is going to be very big, it is a tedious to find out the probability of any event where event is belonging to F where F is sigma field.

So, to avoid the set operation, we transform by real valued function the probability of event can be computed through the set operations by nicely defining a real valued function. Now, we are going to define the random variable that is nothing but the real valued function, which will be useful to find out the probability of some events, not only probability many more information you can get it for the random experiment instead of the set operation, which we have discussed in the first week.

So, let me just recall how the earlier definition goes. So, you take few possible outcomes that you make it as an event A. So, you define a probability over event A; therefore, you are getting the values from 0 to 1. So, this is the way one can define the probability of an event A, where event is nothing but the collection of fewer elements of Ω. So, this value is going to be 0 to 1.

Now, what we are going to do in a little different way that is, I will go to the next slide. Let me give the definition of a random variable through that definition I am going to explain how it goes through the definition of a random variable.

(Refer Slide Time: 06:00)



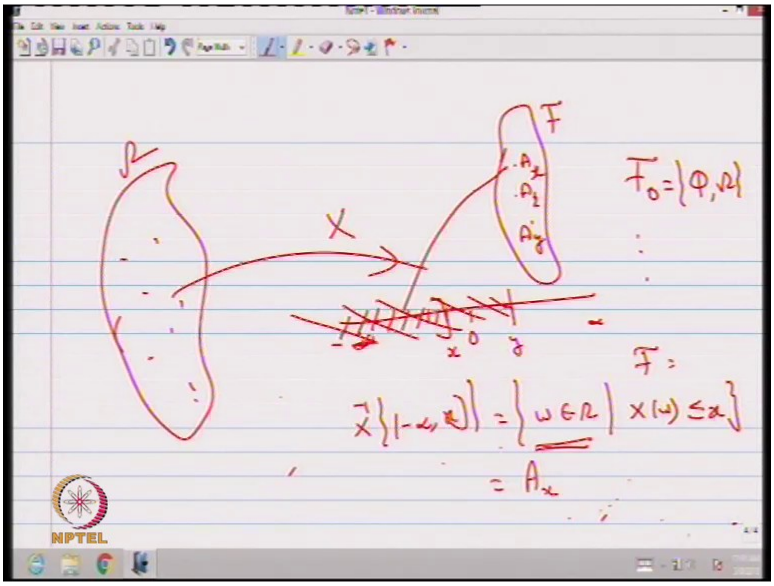
Let (Ω, F, P) be a probability space, Ω is a collection of possible outcomes, F is the sigma field on Ω and P is the set function defined on F such that satisfying the 3 conditions that is Kolmogorov axiomatic conditions of probability that is P(A) ≥ 0, P(Ω) = 1, $P(\cup A_i)$ when Ai's are mutually exclusive events. That is a must $\sum P(A_i)$,then P is going to be call it as a probability function. Therefore, (Ω, F, P) be a probability space.

Now, we are defining here real valued function X defined on Ω is said to be a random variable, a real valued function X define on Ω is said to be a random variable, if

$X^{-1}(-\infty, x] \in F$, x belonging to real line. Whenever a real valued function defined on Ω satisfying the $X^{-1}(-\infty, x]$. That is belonging to F for all x belonging to R then that real valued function is set to be a random variable.

Let me give the pictorial representation or pictorial way of explaining the same thing.

(Refer Slide Time: 09:06)



That means, you have an Ω having many elements in it, it could be finite or it could be countable infinite or it could be uncountably many elements in it. You make a mapping it is a real valued function, from Ω to R. So, it is a real line, if you go for inverse image from - ∞ to any point x. Suppose I take this point as x, $X^{-1}(-\infty, x]$collect all the possible outcomes which is going to give the value from - ∞ to x that is nothing but the few elements of Ω satisfying the condition that .

Under the mapping X it gives the values from - ∞ to x, if I collect those possible outcomes then that should be one of the elements in the F. So, the F may have many elements A_1, A_2 and so on. So, there is a possibility this - ∞ to x will give one of the elements here.

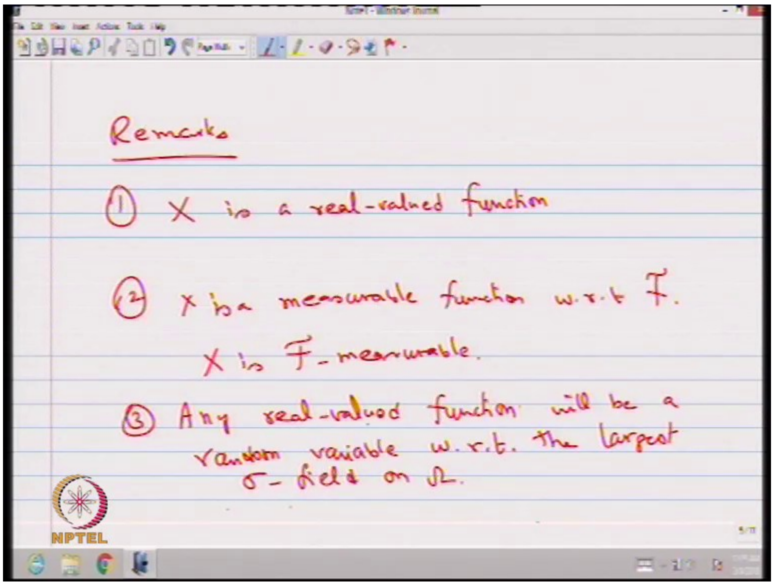
So, if when you change the x from - ∞ to ∞, whatever the inverse image of under the mapping X - ∞ to that x if that is belonging to F or that is one of the element in the capital F, we call this real valued function as a random variable. That means, there is a possibility some real valued function may not be a random variable for that particular F you recall the sigma field. The smallest sigma field is empty set, and the whole set that denoted by F_0; empty set and the whole set like that you can create the many more sigma field the largest sigma field is nothing but the power set collection of all subsets of Ω, and the number of elements in the largest sigma field is going to be 2^n where n is a number of elements in the Ω.

So, some real valued function may not be random variable with respect to the given F. So, I can expand the $X^{-1}(-\infty, x]$ that is nothing but a collection of w's belonging to Ω, such that under the mapping X it is a real valued function that is going to give the value less than or equal to x that is equivalent of the X(w) lies between - ∞ to till x.

So, since you are getting a collection of possible outcomes, this is going to be the event this is denoted by A_x. If I define A_x, x as an event; that means, it is nothing but collection of possible outcomes such that under the operation X, real valued function it gives the values from - ∞ to x; that means, this I can rewrite A_1 as the A_x. Suppose I change the value instead of x into some other y, then $X^{-1}(-\infty, y]$ will be belonging to A_y like that if. Whatever be the X between - ∞ to ∞, the inverse image from - ∞ to till that point is belonging to one of the elements in the capital F then this X is said to be a random variable.

Now, I will go for few remarks, then I will move into the next topic.

(Refer Slide Time: 13:57)



Remark number 1, X is a real valued function that is X is a real valued function, then why it is called a random variable? Because it is defined from Ω to R and the Ω comes from the all possible outcomes of the random experiment therefore, this real valued function is called it as a random variable. So, it is a real valued function define it from Ω to R, Ω is the collection of all possible outcomes of the random experiment therefore, it is called a random variable. By default, random variable means it is a real valued function.

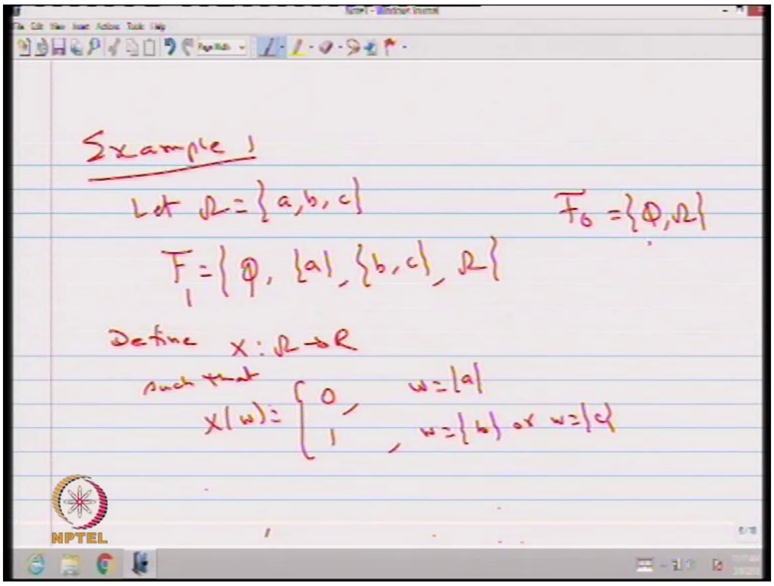
Someone may ask can I have a complex valued function. It is possible to create complex valued random variable, in the form of a one real valued function plus $\sqrt{-1}$ times another real valued function. That means, with the help of two real valued functions, you can make a complex valued random variable that is possible. But as far as this course is concern, we keep X as a real valued function and the random variable is formed with the real valued function that is it.

Second one, X is a measurable function, with respect to the sigma field F what is the meaning of measurable function? The inverse image of a semi closed interval that is belonging to one of the elements in the F. Therefore, X is going to be call it as a measurable function with respect to sigma field F. So, there is another name for the random variable, that is a measurable function.

So, if you do the advanced level probability theory course, so, we call a random variable as the measurable function with respect to F. So, there are some books they use called a X is a F measurable; that means, X is a random variable with respect to the probability space (Ω, F, P). Kindly note that, there is notion of probability is not coming to the picture when you define the random, variable to define a random variable you need Ω and F there is no need of P; P is nothing but the probability.

So, through probability one can study the random variable X that is called studying the distribution of the random variable X. So, there are few more remarks I want to make it so that I will do some more remarks after giving some examples. So, we will go to the example of random variable.

(Refer Slide Time: 17:56)

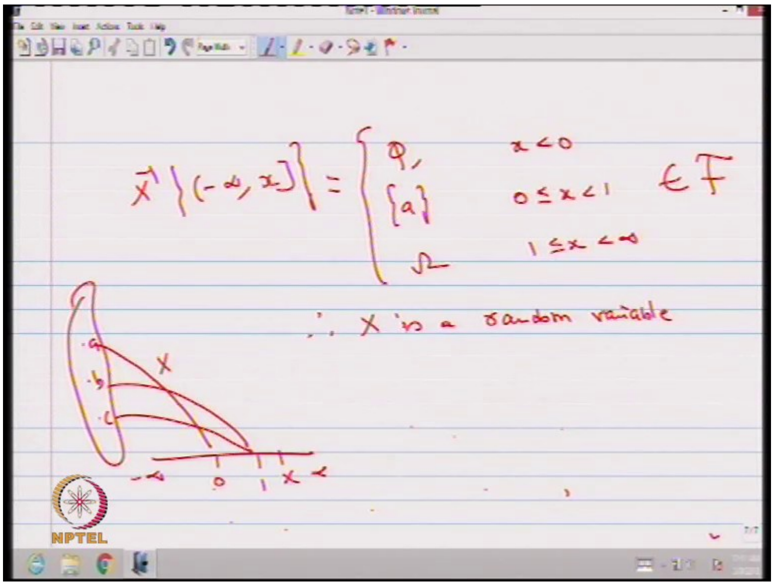


So, the first easiest example, example 1: let **Ω** consist of three elements with the three elements, you know that we can frame the sigma field the smallest one which consist of empty set and the whole set, then you can go for the largest one, but since I want to explain the random variable. So, I am going to introduce, I am going to define the sigma field, which is in between the smallest and the largest.

So, the F which I am going to consider here that has four elements; empty set, singleton element a and both the elements b and c and the whole set. So, this is not the smallest and this is also not largest sigma field. Now, I am going to define a real valued function from **Ω** to R such that X(w) that is going to take the value 0, when w is going to be singleton element a. It is going to take value 1 if the w is going to be b or w is going to be c. So, this a real valued function mapping from **Ω** to R, **Ω** consist of three elements a, b, c.

Now, we will check whether this real valued function is a random variable or not. Let us go for it.

(Refer Slide Time: 20:04)



We will go for finding $X^{-1}(-\infty, x]$, that is going to be empty set, if x < 0, you recall X is the mapping from Ω to R it takes a value. So, a is connected with the 0 and b and c both are connected with 1. Therefore, I can redraw the diagram. So, there is no this line.

So, X is a mapping from Ω to R, the $X^{-1}(-\infty, x]$. So, when small x is less than 0 suppose you treat x is going to be somewhere here, between - ∞ to 0 then you are looking for what is the inverse image of - ∞ to x under the operation X. You are not getting any possible outcomes because one possible outcome a is mapped with 0, b and c is mapped with 1. Therefore, when x is less than 0 you are not getting any possible outcomes. Therefore, it is going to be empty set.

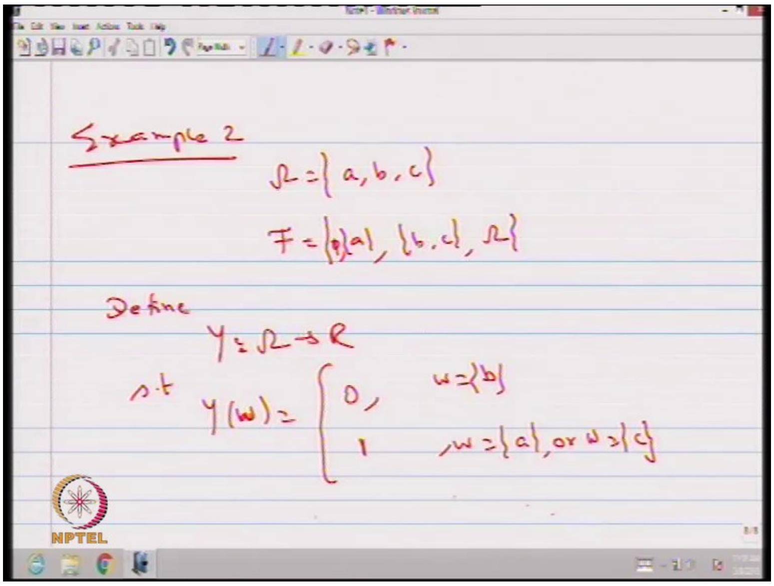
Now, we will go for when x is lies between 0 to 1 including 0 excluding 1; that means, somewhere now the treatment is x is not there, x lies between 0 to 1, it can include 0 also. Now, the inverse image of - ∞ to x, where x is lies between 0 to 1. Now you find out what is the inverse image; that means, you look for what is the possible outcomes which gives the values from - ∞ to x where x is lies between 0 to 1. Since 0 is included, therefore, the inverse image is going to be the singleton element a.

Now, we will go for when x is lies between 1 to ∞; that means, now the x lies between 1 to ∞ somewhere here what is the inverse of each? That means, whatever the possible outcomes which is going to give the value from - ∞ to x where x is one to ∞, you can include all the possible outcomes.

So, since a is mapped with 0, b and c are mapped with 1, therefore, it is going to be singleton element a singleton element b singleton element c. So, when you include all these three elements, then it is nothing but **Ω**. Now we will check whether this inverse image of the values are belonging to F or not. So, for us the F = {φ, {a}, {b, c}, **Ω**}, whereas here we land up empty set, singleton element a and **Ω** all are belonging to F. There is a possibility few elements of the F may not come into the picture here that does not matter, but whatever the element comes here that should be belonging to F or not that is the question.

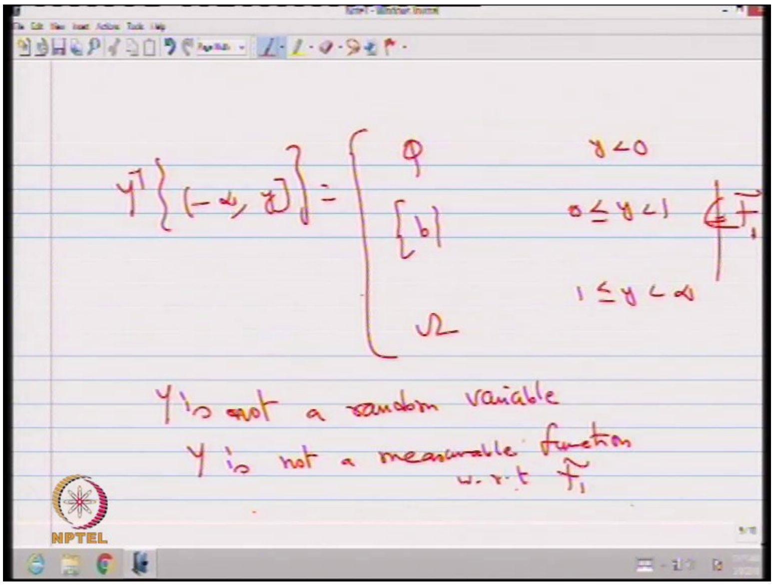
So, when x moves from - ∞ to ∞ whatever be the possible x all possible x the inverse image of semi closed interval that is belonging to the F. Hence, X is a random variable, it is a real valued function in a probability space (**Ω**, F, P), this real valued function satisfies this condition. Therefore, this real valued function is a random variable, it is a measurable function with respect to F. In the hidden, thing is random variable with respect to F that is hidden, but we usually won't write we just say that X is a random variable, but actually it is a random variable with respect to F. Why I am saying this now we are going to give another example, in which that real valued function is not going to be a random variable for the same F. So, then we will go for concluding some result then we will make the remark or the other remarks.

(Refer Slide Time: 25:40)



We will go for the second example, for the second example we take the same **Ω**, we take the same F empty set singleton element b and c and the whole set. Now, we define the new real valued function Y from Ω to R such that Y(w) is going to give the value 0, when w is going to be a singleton element b or it is going to give the value 1, when w is going to be a or w is going to be c. So, this is a change. So, this is also real valued function Ω to R and Y(w) takes a value 0, when w is equal to b it takes a value 1 for w is equal to a or w is equal to c. Let check whether this is going to be a random variable.

(Refer Slide Time: 27:05)

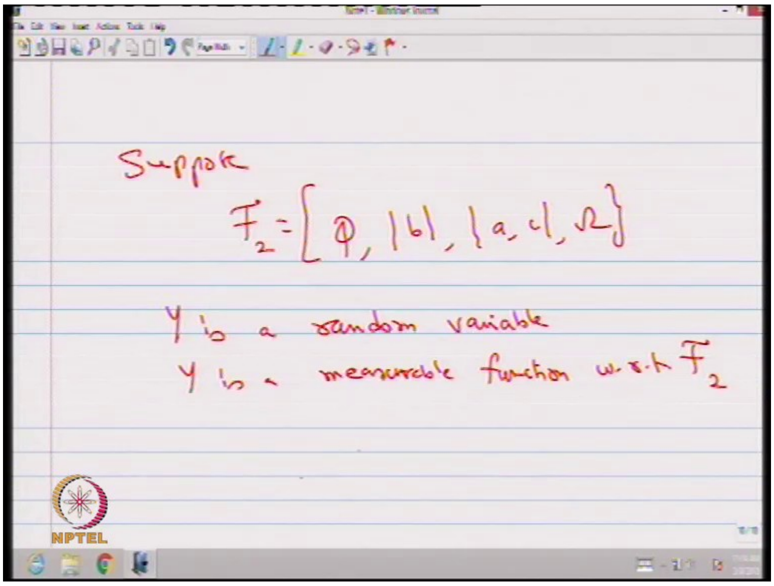


$Y^{-1}(-\infty, y]$, that is going to give the value, when y is between - ∞ till 0 you are not going to get anything any possible outcomes therefore, it is an empty set. When y is from 0 to excluding 1, the inverse image is going to give only the singleton element b, because Y(w) is equal to 0 when w is equal to singleton element b. When y is lies between 1 to ∞ similar to the earlier exercise example, I will end up with collecting all possible outcomes therefore, it is a or b or c union everything therefore, it is going to be the whole set.

Now, let me check whether these three elements are belonging to F or not. So, the here the F = {φ, {a}, {b, c}, Ω}. Empty set is belonging to F, whole set is belonging to F whereas, the {b} is not belonging to F therefore, all these elements are not belonging to F. That means, it is a real valued function, but it is not going to be a random variable with respect to the F, which we have discussed; that means, Y is not a random variable with respect to this F. You can raise the question when this real valued function is going to be the random variable. That means, by changing or by having a new probability space, same Ω with the different F may give this real valued function going to be a random variable. Can you guess what will be the

F in which this real valued function is going to be the random variable? Yes, you would have guessed it basically.

(Refer Slide Time: 29:52)

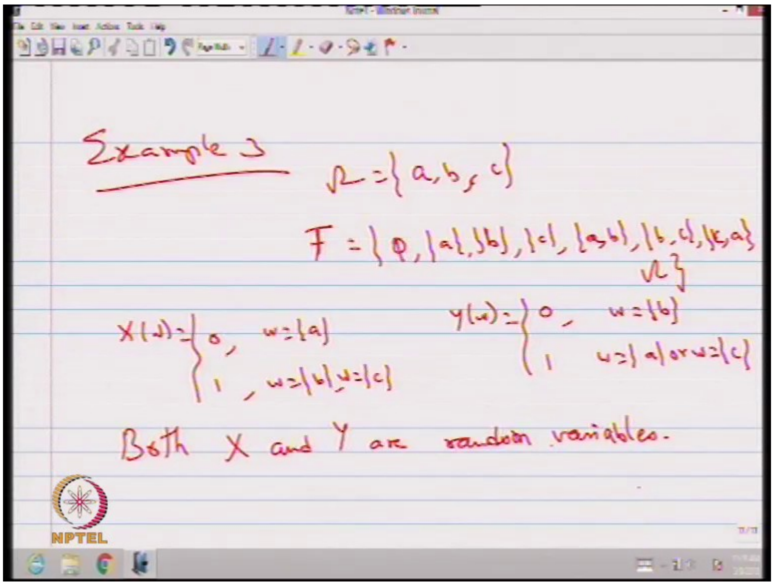


Suppose Ω sorry suppose F = {φ, {b}, {a, c}, Ω}, in this F the Y is a random variable or Y is a measurable function with respect to this F. Since, we have more than one sigma field, I am going to note down the notation as a this I treat it as I will go to the example 1, this I treat it as the F_1. F_0, I will keep it as empty set and the whole set. So, I denote this as the F_1, therefore this is belonging to F_1, therefore X is a random variable. Come to the second example, same F we are using F_1. So, Y is real valued function it is not belonging to F_1. Therefore, Y is not a random variable Y is not a measurable function, Y is not a measurable function with respect to F_1.

Suppose we keep another sigma field that I denote it as a F_2. Y is a random variable or Y is measurable function with respect to that is very important, with respect to F_2not F_1. Now, we can make another question based on the 2 examples what is the F in which both the real valued function going to be a random variable. I will repeat the question, in the example 1, X is a real valued function that is going to be a random variable with respect to F_1.

Second example Y is a real valued function, but Y is not a random variable with respect to F_1 whereas, Y is a random variable with respect to F_2 another sigma field. Now I am going for the third example, same real valued function that is X and Y what is the F in which both the real valued functions are going to be random variable that is a third example.

(Refer Slide Time: 33:12)



That is I have a three elements, what is the F in which the real valued function Ω is equal to a, it takes a value 1, when Ω is equal w = b or w = c and another real valued function that is also takes a value 0, when w is equal to b and takes a value 1,w = a or w = c. These two real valued functions are going to be a random variable under one sigma field.

So, what is the sigma field? Suppose I keep the largest sigma field suppose I keep the largest sigma field that is nothing but the power set. You can count 1 2 3 4 5 6 7 8 elements because Ω has three elements. So, $2^3$is 8. So, the number of elements in the largest sigma field that has the 8 elements that is nothing but the power set. Under this sigma field F, both the real valued function going to be the random variable you can prove it the same way whatever I have done the derivation, you can go for it its already there, only you have to check whether that is belonging to F. But if you go for X inverse of semi closed interval, similarly, Y inverse of semi closed interval both are belonging to F. Therefore, one can conclude both X and Y are random variables, that is both X and Y are measurable functions with respect to this F that is a largest sigma field.

From these three examples one can conclude, whenever you have a largest sigma field any real valued function is going to be the random variable. So, that is the next remark, which I have given earlier. Yes, we have given two remarks now I am going to make a third remark, that is any real valued function will be a random variable with respect to the largest sigma field or sigma algebra on Ω.

Therefore, many of the probability course if it is a elementary, they would not discuss the random variable in the form of a through the measurable function concept, that is X inverse of semi closed interval is belonging to F. Whenever you have a largest sigma field or you keep the probability space with the largest sigma field, then any real valued function will be a random variable. That means, it is a mapping from Ω to R, that real valued function will be the random variable.

The way I have given two examples, you can think of many more examples instead of three elements you can go for easy example of tossing a coin which has a 2 possible outcomes head and tail, then you can have a F, since it has two elements in the Ω. Therefore, the you will end up with the F is going to be the largest sigma field unless otherwise you take the smallest one.

So, you will have a F = {φ, h, t, Ω}. So, it has four elements. So, whatever the way you define the real valued function on the random experiment of tossing a coin, any real valued function is going to be a random variable because the F is largest sigma field. So, having an example with the two elements, it would not serve the purpose of defining a random variable in a nice way that's what I have taken three elements.

Now, the next question can I go for 4 and 5 elements or finite number of elements. So, if you have a finite number of elements for example, you can think of throwing a dice, which has Ω has 6 elements 1, 2, 3, 4, 5, 6 here it has 3 elements. So, when you throw a dice you have consist of 6 elements. Therefore, other than the smallest sigma field other than the largest sigma field which is the power set you will have many sigma fields in between. In that case again you can have any real valued function for example, getting even number that real valued function is 0 getting an odd number that values is going to be 1. So, you can map X is getting even number becomes 0 and odd number becomes 1. Need not be 0 and 1 also you can go for any values as long as it is a finite value, you can have any value mapping from Ω consisting of 6 elements to the any real numbers that is a real valued function.

Then the F which you have sigma field accordingly some real valued function, may be a random variable, some real valued function may not be a random variable. But in that example also if you keep the power set as the sigma field the largest one, then any real valued function you define it over the Ω consisting of six elements also that is going to be a random variable. You can go for some books they use the indicator function, whether that is going to be a random variable or not. So, you can go for the inverse image whether that is belonging to the F or not; if you have a F has all the elements including these elements then it is going to be a random variable if not it is not going to be a random variable.

So, I have given three remarks. So, I want to introduce one more remark, that is, the one which I have given as a, if condition, the real valued function is said to be a random variable if inverse image of this is belonging to F, this condition is if and only if condition; that means, the converse also true, that is the fourth remark.

Basically, the - ∞ to x, a semi closed interval that is nothing but the Borel set. So, since I made the Borel set of the form - ∞ to x therefore, many of the other events can be written in terms of - ∞ to x, this semi closed set form. Therefore, this Borel set will make if and only if condition; that means, if this condition is satisfied then that is going to be a random variable, if you have a random variable always this condition will be satisfied that is the meaning of if and only if, that is a meaning of converse.

So, since I am going for a very particular type of (-h,x], this makes this condition is going to be an if and only if condition. That means, any real valued function satisfying this condition will be a random variable as well as any random variable having this condition sorry; any random variable will always have this condition for all x belonging to R, x inverse of - ∞ to x that is belonging to F.

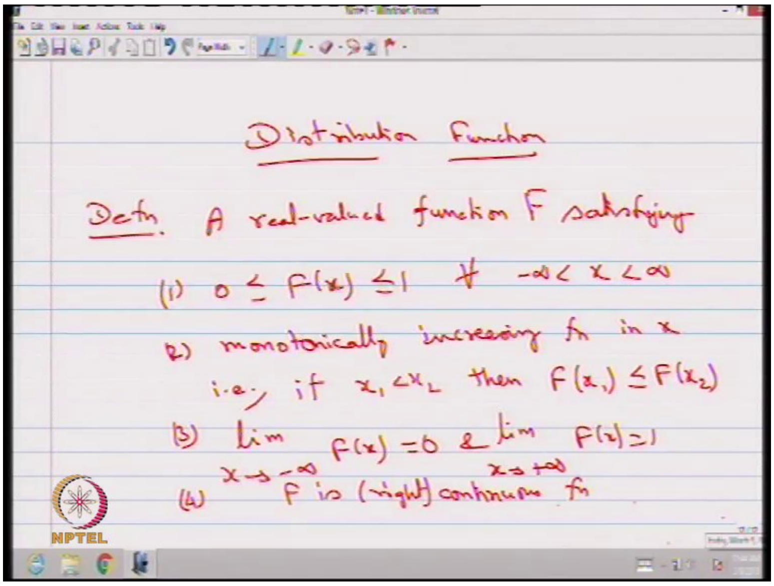
So, this is also one of the very important remarks whatever I have said it in earlier three remarks, this is going to be the fourth remark.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 07

Now, we will move into the next topic that is distribution function.

(Refer Slide Time: 00:18)



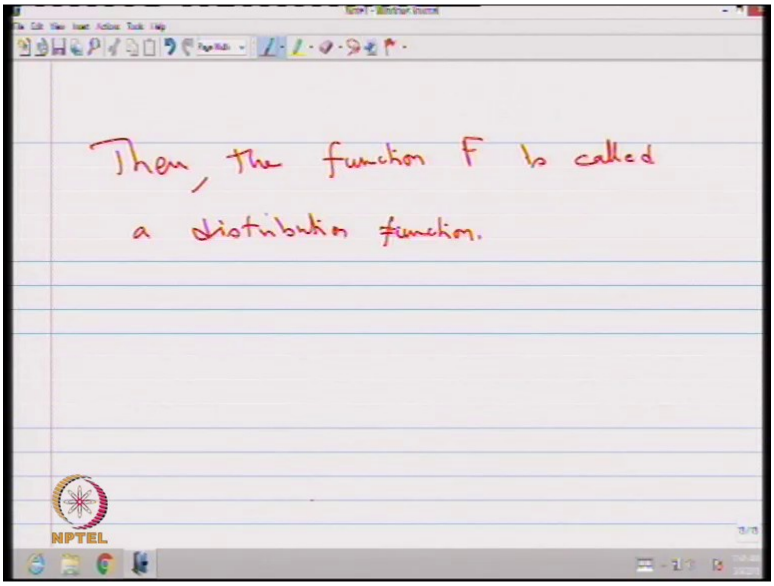
That is distribution function, this is nothing to do with probability. So, first let me define the distribution function, then we will connect the distribution function with the probability. Therefore, the same distribution function will be call it as a cumulative distribution function of a random variable X. So, when I define now the distribution function, there is no probability at all.

Let me define distribution function. Definition, a real valued function F, because later we are going to use this F as a CDF. Therefore, I am going to use now itself F. A real valued function F satisfying 4 conditions, the first condition; it is always lies between 0 to 1 for all x lies between - ∞ to ∞; this is the first condition. Second condition, it is a monotonically increasing function in x; that means, if two values you take $x_1, x_2; x_1 < x_2$, then the F(x_1) $\le F(x_2)$. It is same as a non-decreasing function whether you say it is a monotonically increasing function for two different values of $x_1 < x_2$, then your F(x_1) $\le F(x_2)$ or non-decreasing function both are one of the same.

Third condition, $\lim_{x \to -\infty} F(x) = 0$ as well as $\lim_{x \to \infty} F(x) = 1$. It is not substituting the F at - ∞ is 0, F at +∞ is 1, it may holds in some examples, but the definition is $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$ it is not the substitution, it is a limit.

The fourth condition, that is the F is right continuous function, I can use the word I can make right in the bracket. That means, the function can be continuous, if not it is a right continuous. The function is continuous function if not it has to be a right continuous function. If these 4 conditions are satisfied then you can conclude the F is distribution function.

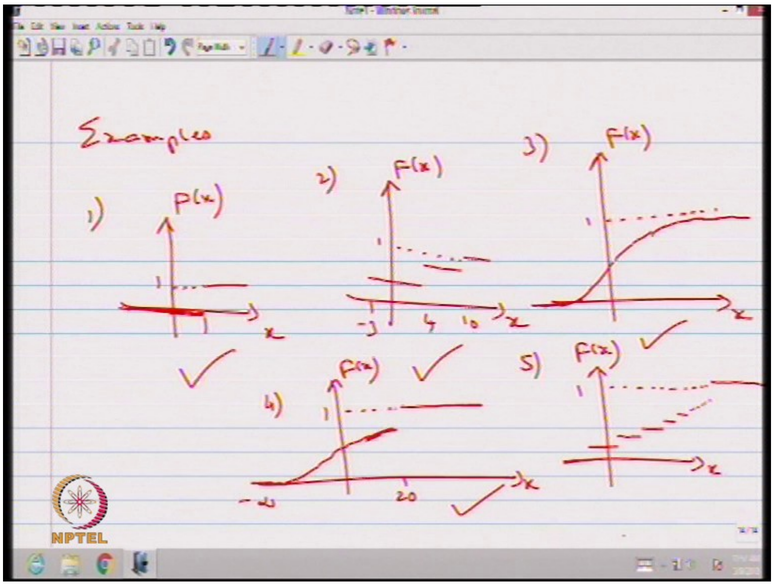
(Refer Slide Time: 04:31)



Then the function, the real valued function F that is called a distribution function. As I said this is nothing to do with the probability, any real valued function satisfying these 4 conditions is going to be call it as a distribution function if given any one of the properties are not satisfied then you cannot conclude. It has to be real valued function value lies between 0 to 1 it should be a monotonically increasing limit value has to start from 0 at - ∞ and it had end up 1 as a limit x tends + ∞ and either it should be continuous or right continuous.

You can draw some diagram for the function F and you can get the feel. After that we can introduce the probability, then we land up cumulative distribution function of the random variable X. So, let me start what are all the graph of F so, that that is going to be the distribution function. You can also try you can also try to draw the graph x axis x, y axis is capital F, so that satisfying these 4 conditions. Therefore, that is a distribution function you try I will also going to give some two three diagrams so that you can conclude which one is going to be distribution function which is not a distribution function and so.

(Refer Slide Time: 06:43)



Examples: first example when I asked the students to draw the diagram, they can make it different type whether it is going to satisfy the 4 condition or not that is important. Since, it values from 0 to 1 the easiest diagram the values is 0 till this point, and this point it increases becomes 1. Therefore, this satisfies all the four conditions. Therefore, it is distribution function I can put the tick mark.

I draw another diagram, at some point it has an increment. So, this point you treat it as -3, at some other point it is not scaled at some point it increased, at some other point again it increased and this value suppose that is going to be 1, suppose this value is going to be 1; that means, till - 3 it is 0, -3 to 4 it has some value constant. There is an increment at the point 4 and between interval 4 to 10 it has a same value, at the point 10 it becomes 1 and it is 1 till the end. Therefore, this also satisfies.

And go for another diagram as a third example, it is 0 and it is increasing and it touches it becomes 1 at ∞; that means, asymptotically it goes and it becomes 1 at ∞. So, you see the difference of the three different distribution function, the first one has the only one jump at the point 1, the second one has three jumps and the fourth third example which does not have any discontinuity or no jumps it starts from 0 and it goes to 1 and it is a continuous function, whereas the first and second example are the right continuous function. That means, the way

I define the value at the point wherever there is a jump and where what the value at the right limit are same, which is different from the left limit in the first two examples, whereas in the third example it is a continuous function or I can go for one more diagram.

It takes a value 0 and it is increased to some value, at some point there is a jump, then it becomes 1. That means, it is continuous between the interval - ∞, suppose this point is a 20. This function is a continuous between the interval - ∞ to 20, 20 there is a jump and it becomes 1 from 20 to ∞. Therefore, it has a continuous in some interval jumps at one point or there is a possibility it may have jumps at many points, it may have a combination of continuous in some interval as well as jumps.

Whereas, if you see the first example and second example, the distribution function increases the values only by jumps, I am repeating the word the first example and the second example the distribution function start from 0, it incremented at some value and it becomes a same value till the next jump, then the next jump it increases till the next jump then it again increase then it goes to the same value.

Therefore, the value increases only by jumps in the first example and the second example whereas in the third example it increases continuously from one point till the end whereas, in the fourth example it increases in some interval as well as it retains the same value with some jumps. Therefore, all four types are going to be the distribution function, I have not given what is the function which is not a distribution function.

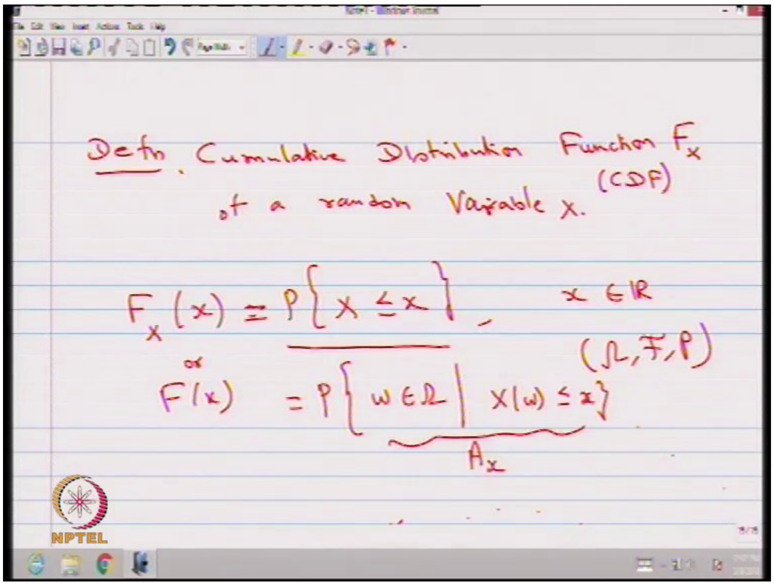
So, you can think of a wrong example of decreasing or not values are lies between 0 to 1; it is more than one or less than 0 or it is discontinuous. So, violating these four conditions you can always create n number of examples, which are going to be a non distribution function. But what we want is a real valued function satisfying the four conditions. therefore, this is going to be a distribution function.

So, these four example makes the first example has a only one jump, the second example has a 3 jumps like that you can have a finite number of jumps. The third example does not have a jump at all, it is increasing in the whole interval and the fourth example is the combination of increasing as well as having jumps. We can go for one more example which has the countably infinite jumps, that makes a completeness of the distribution function.

So, the fifth example is having countably infinite jumps. So, it starts from 0 and it has a jump, again it has a jump, again it has a jump, again it has a jump, like this it is keep going, it reaches one at some point.

The first example has only one jump, second example has three jumps, fifth example has the countably infinite number of jumps, third example does not have a jump, fifth example has a one jump as well as increasing between some interval; this shows the distribution function can be of many forms. So, this will lead us to create types of random variable in the later stage, but before that let me explain what is cumulative distribution function, through the distribution function. That is going to be the next definition is the cumulative.

(Refer Slide Time: 16:15)



Next definition is cumulative distribution function of a random variable X. So, this is a CDF is denoted by F_X, that is denoted by F_X, that is a cumulative distribution function and how it is defined for all x that is same as $P\{X \leq x\}$, where x is in the real line.

The F which we define that is a distribution function, extra we are using a suffix X that is to show that the same distribution function F is call it as the cumulative distribution function of a random variable X. So, whether I write capital F_X or capital F(x) both are on in the same.

So, I can use a word or here to distant gives or to tell that both are distribution function and the way you define the distribution function through the P, that P comes from the probability space (Ω, F, P) the same capital P is used. In the probability space you have a P that is a probability measure set function. So, if you find out $P\{X \leq x\}$ for all x belonging to R, this way if you define the distribution function and that distribution function is going to be call it as a cumulative distribution function of a random variable X.

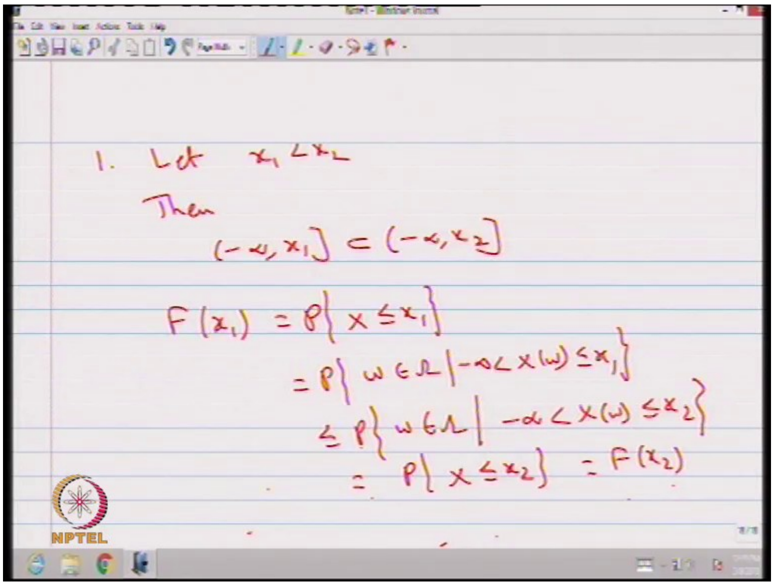
Now, the way we write $P\{X \leq x\}$, the way you define this way the distribution function is going to be call it as a cumulative distribution function of a random variable X. In short, they use CDF; CDF means cumulative distribution function. I can expand what is the meaning of $P\{X \leq x\}$; that means, that is same as $P\{X \leq x\}$ is nothing but for possible outcomes w belonging to Ω such that, under the operation X, the w gives the values, X(w) gives the values less than or equal to x, that is a meaning of a $P\{X \leq x\}$.

Means you are collecting few possible outcomes which satisfies the condition $X(w) \leq x$. This is nothing but a set this is nothing but event, the way I explained in the earlier this can be denoted by the later called A_x; that means, it is a $P(A_x)$. A_x is an event, this event is belonging to the F by using a Kolmogorov axiomatic definition, the $P(A_x) \geq 0$ for all events and the $P(\Omega) = 1$ and P of union of mutual it is joint events that is same as sum of $P(A_x)$.

The using the same Kolmogorov axiomatic condition, the capital F is defined for all x belonging to real line; that means, when x between - ∞ to plus ∞ you will get the values of $P\{X \leq x\}$, that will satisfies the four condition which we said it earlier, that is it is a real valued function lies between 0 to 1, monotonically increasing, limit at - ∞ is 0, + ∞ is 1 and F is a right continuous function. So, all these four conditions will be satisfied. Therefore, this distribution function the F that is going to be call it as a cumulative distribution function of the random variable X, that is denoted by F or F_X. When we have more than one random variable the suffix is important, now we have only one random variable. Therefore, whether you right F(x) or F_X both are one and the same

Now, let us prove how these function F(x) is going to be the distribution function. Once you prove that it is a distribution function through the $P\{X \leq x\}$, then you can conclude the same distribution function is going to be call it as a cumulative distribution function of the random variable X. Even though there are four points to be proved the first one is lies between 0 to 1, and second one is monotonically increasing or non-decreasing function and the limit at - ∞ is 0 and + ∞ is 1 and right continuous. So, we will prove only few things the remaining proof can be done it in the similar way.

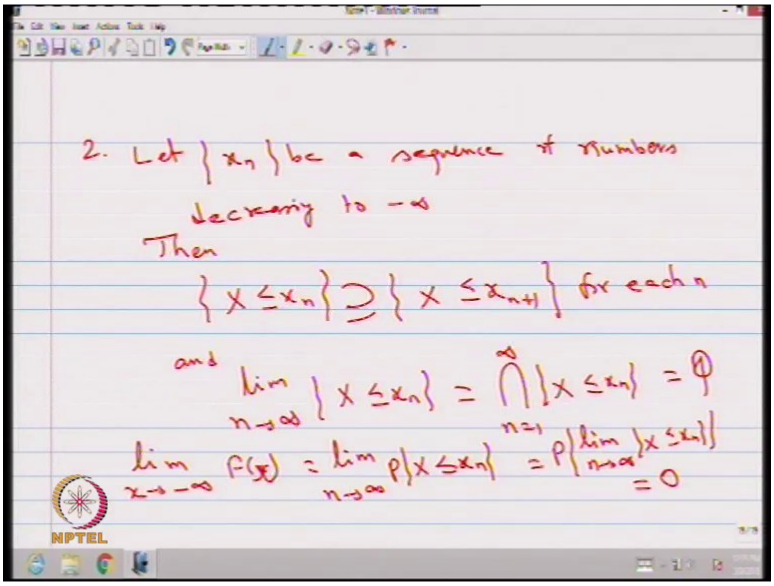
(Refer Slide Time: 23:30)



The first one to prove that it is non decreasing or monotonically increasing, you take two points with $x_1 < x_2$, and you can always conclude - ∞ to x_1 which is contained in - ∞ to x_2 and this is a one set and is another set and once it is contained in the other set.

Suppose you go for finding what is the F(x_1), that is nothing but the P{X ≤ x_1}, that is same as P of set of all w's belonging to Ω such that X(w) lies between - ∞ to x_1. Since, once it is contained in the other set or equivalent of one event is contained in the other event, then the probability is going to be less than or equal to, therefore, that is less than or equal to the P of collection of w's such that - ∞ to X(w) ≤ x_2 this is nothing but P{X ≤ x_2}, and that is same as F(x_2). Therefore, we can conclude F is the non-decreasing in X or monotonically increasing.

(Refer Slide Time: 25:25)



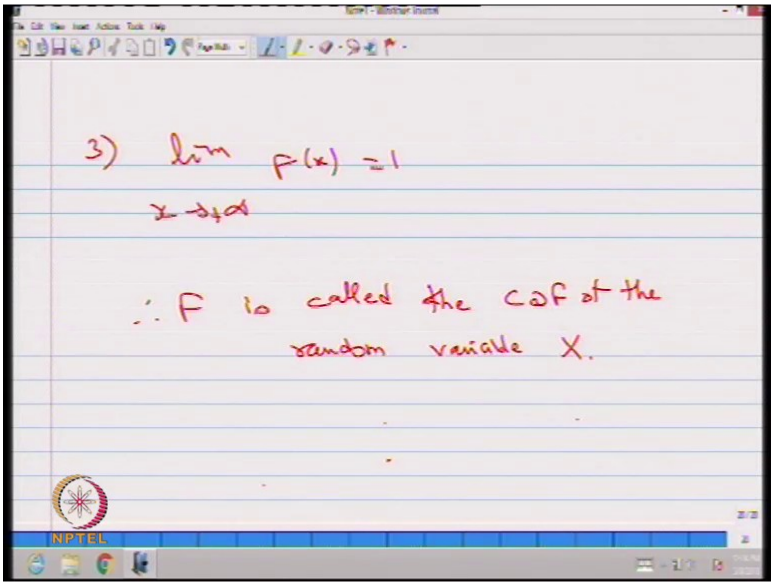
The second one we can prove $\lim_{x \to -\infty} F(x) = 0$, for that we take sequence x_n be a sequence of real numbers such that it is decreasing to - ∞, then you can conclude X ≤ $x_n$that satisfies this condition. For each n not only that the limit n tends to ∞ the event X ≤ x_n that is nothing but $\bigcap_{n=1}^{\infty} \{X \le x_n\}$.

Since x_n's are the sequence of a numbers decreasing to - ∞ satisfying this condition. Therefore, $\lim_{n \to \infty} \{X \le x_n\} = \bigcap_{n=1}^{\infty} \{X \le x_n\}$, where n running from 1 to ∞ that is nothing but the empty set.

Therefore, $\lim_{x \to -\infty} F(x) = \lim_{n \to \infty} P\{X \le x_n\}$. Just now we concluded $\lim_{n \to \infty} \{X \le x_n\}$ is empty set. Therefore, this is nothing but P of, I can interchange the role of a limit and probability, because P is a continuous function P{$\lim_{n \to \infty} \{X \le x_n\}$} and that is empty set. So, the P(φ) = 0.

So, we have proved $\lim_{x \to -\infty} F(x) = 0$.

(Refer Slide Time: 28:41)



Similarly, one can prove $\lim_{x \to \infty} F(x) = 1$ and forth one, one can prove the F(x) is the right continuous in x. Since, all the four conditions are satisfied for the distribution function. Therefore, this distribution function F is called the CDF of the random variable X. So, the distribution function satisfying $P\{X \le x\}$ form that distribution function will be the cumulative distribution function for the random variable X.

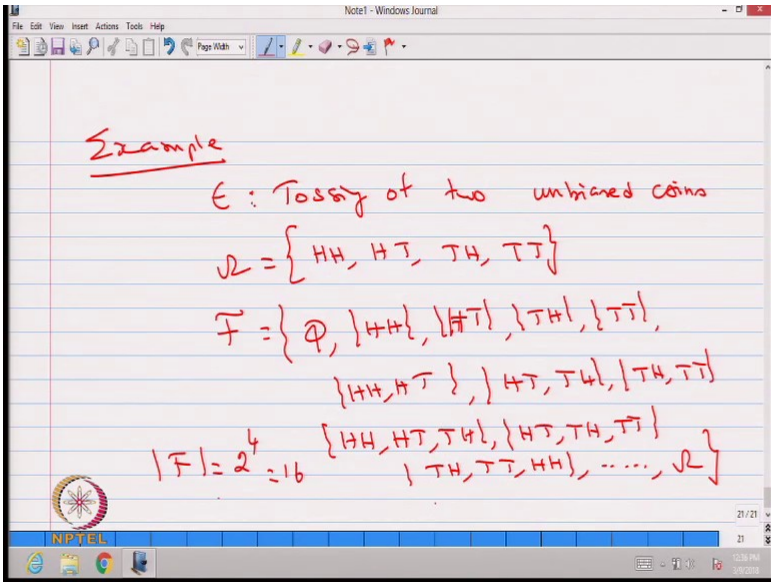
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 08

So, now we have explained how to get the cumulative distribution function of a random variable, through the concept of distribution function. That is what we explained the distribution function first, that is a real valued function satisfying the four conditions. Then we are connecting through the $P\{X \leq x\}$ therefore, it is going to be the CDF of the random variable.

First, we will go for the very easy example, in which how to get the cumulative distribution function for the random variable. So, for that we will take a very easy example, start from the scratch.

(Refer Slide Time: 00:50)



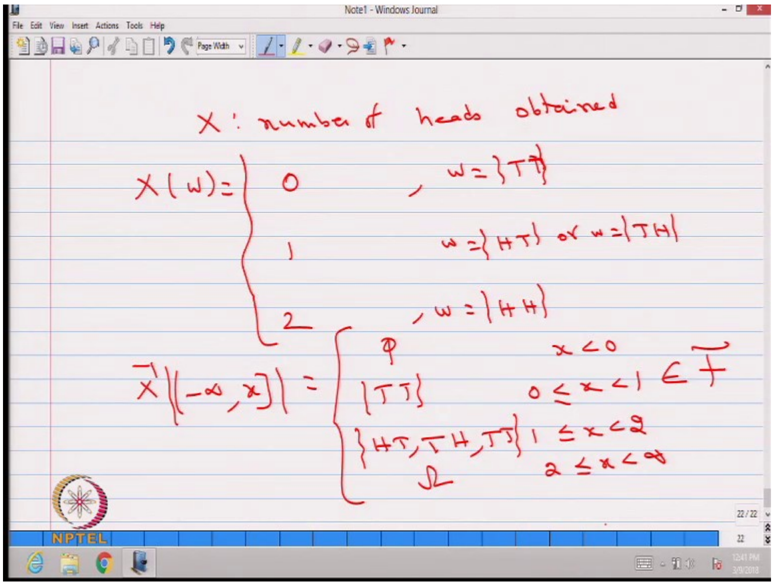
Start from the scratch means, you start with the random experiment. So, the random experiment is denoted as a tossing. The random experiment is nothing but tossing of tossing of two unbiased coins. Random experiment is nothing but tossing of two unbiased coins; either you can use unbiased coins or fair coins; that means, there is equiprobable of getting tail or head that is the meaning of unbiased coin or fair coin. So, the random experiment is the tossing of two unbiased coins.

Therefore, you will get collection of all possible outcomes is going to be I use a notation H for getting head, T for tail. So, HH means both we got head and head, head and tail, tail head, or tail tail. When you tossing two unbiased coin therefore, you will get either head head or head tail or tail head or tail tail. Therefore, these are all the four possibilities.

We have taken the largest sigma field F, further consideration for this problem that is empty set then singleton element, any two elements, oh I have written wrongly that is HH, HT any two elements, then any three elements then; so 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, so the largest sigma field consisting of a 16 elements. So, instead of writing other elements I am just putting …, you can fill up all the other elements so, that this is going to be the largest sigma field.

So, for this problem the F is taken as a largest sigma field, which has number of elements is 2^4 that is 16 elements. So, one can fill it up all the elements. Now we are going to define the random variable that is nothing but the real valued function where X denotes number of heads obtained.

(Refer Slide Time: 04:46)



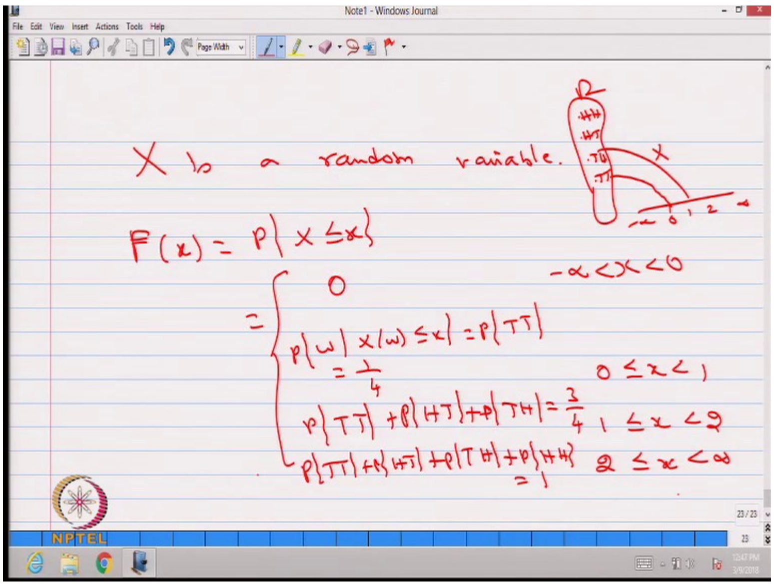
X is a real valued function which denotes number of heads obtained, when you toss two unbiased coins.

So, X(w), the possible values are going to be since it is a number of heads. therefore, there is a possibility no heads or there is a possibility you may land up with only one head or 2 heads therefore, the possible values are 0, 1 or 2. You can list out what are all the w's will give 0 as well as 1 and 2. So, for 0 it is both are going to be tail then it is going to be the both tail will give X(w) = 0 when w is equal to HT or w is equal to tail head, you are going to get X(w) = 1.

When w becomes both head you are going to get the value 2. So, the possible values of the random variable X are 0, 1 and 2. So, this is a real valued function. One can verify whether this real valued function is going to be a random variable or not. So, if you go for finding X^{-1}¿, since the possible values are 0, 1 and 2. Therefore, when x < 0 you do not have any possible outcomes. Therefore, it is an empty set. When x lies between 0 to 1 when x takes a value from 0 to 1 excluding 1, the X inverse of - ∞ to x becomes tail T when x takes the value from one to 2 excluding 2, it is going to be HT, TH, TT when x lies between 2 to ∞, all possible outcomes will be included therefore, this is going to be the **Ω**.

Now, we have to verify whether these four elements are belonging to F or not. Since we have taken F as the largest sigma field, therefore, empty set, TT, HT, TH, TT and the **Ω** all are belonging to F. Since for all x, X^{-1}¿ belonging to F, X is a random variable, because it satisfies the condition.

(Refer Slide Time: 08:40)



Now, we will go for finding the cumulative distribution function of the random variable X, that is defined as capital F(x), that is $P\{X \leq x\}$. For all possible values of x, $F(x) = P\{X \leq x\}$. Since, the possible values of X is going to be 0, 1 and 2 therefore, we can define the F of x when it is before 0 between 0 to 1, 1 to 2 then 2 to ∞.

So, based on the problem you can always split the ranges. So, since here the X is the number of heads obtained and the possible values of heads are going to be 0, 1 and 2 therefore, we are going to define F(x) is a value, when $x < 0$, then 0 to 1, then 1 to 2, then 2 to ∞. So, this is $-\infty$ to 0. So, what is the value when x lies between $-\infty$ to 0, excluding 0, when x is lies between 0 to 1, then x lies between 1 to 2, then 2 to ∞. When x lies between $-\infty$ to 0 the $P\{X \leq x\}$ is nothing but collection of possible outcomes, which is going to give the value from $-\infty$ to till x where x lies between 0 to ∞.

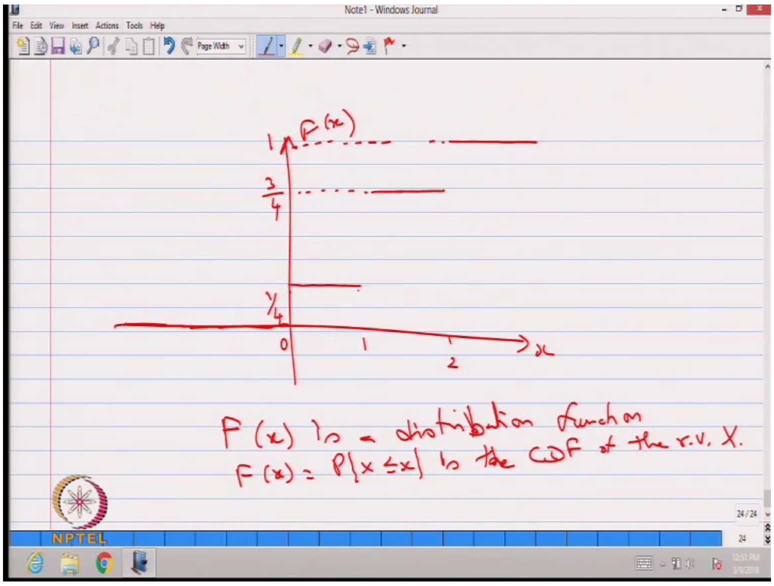
Since the possible values of x, I can just draw one small diagram also. Ω consist of HH, HT, TH, TT and this is mapped with 0, 1, 2 under the operation X, when $P\{X \leq x\}$ when x is lies between $-\infty$ to 0 it is nothing no possible outcomes, therefore, $P(\varphi)$. So, $P(\varphi) = 0$.

When x lies between 0 to 1, since the TT is mapped with 0 collection of possible outcomes this is nothing but w such that $X(w) \leq x$, here it is P of, the w here is a TT and since it is the fair coin or unbiased coin, it is equiprobable of getting tail as well as head and the probability is 1. Therefore, it is 1/2 probability of heading head 1/2 probability of heading tail, and probability of getting HT in both the tosses is going to be 1/2 times 1/2. Therefore, it is going to be 1/4. The probability of obtaining head in one toss is 1/2 and probability of getting another toss is 1/2. Therefore, it is a probability of a 1/4 getting head HT.

When x is lies between 1 to 2 that is nothing but collection of possible outcomes gives the values $X(w) \leq x$, here it is going to be P of TT + probability of a HT + probability of TH. So, all these possibility gives the value of a $P\{X \leq x\}$, when x is lies between 1 to 2. So, we know that the probability of TT is 1/4 and a HT is another 1/4, TH is another 1/4, therefore, the probability is going to be 3/4. When x lies between 2 to ∞ this is going to be probability of TT + probability of HT + probability of TH + probability of HH because that gives the value 2.

So, 1/ 4 + 1/4 + 1/4 + 1/4 therefore, this is going to be 1. So, the CDF is 0 between $-\infty$ to 0, the CDF value is 1/4 between 0 to 1 and 3/4 between the interval 1 to 2, and 1 between the interval 2 to ∞.

(Refer Slide Time: 14:46)



we can draw the diagram for this nicely, when x takes the value - ∞ to 0 it is 0, 0 there is a jump of height 1/4, I am not scaling the graph. At the point 1 it has a next jump of height 3/4. So, suppose this is 1/4, another 1/4, another 1/4, so this much height at the point 2. So, this value is basically 3/4. At the point 2 it has a next jump. So, that jump value is 1. That means, it is right continuous function because the left limit at the value at 0 that is 0, value at the 0 is 1/4 right limit at the point 0 is 1/4 therefore, it is right continuous at the point 0.

Similarly, the value of CDF at the left limit of one that is 1/4, the value at the point one that is 3/4 and the right limit at the point 1 that is again 3/4 therefore, it is a right continuous at the point 1 also. Similarly, the function value at the point 2 that is a right continuous. Therefore, this is the example in which the CDF is satisfying all the four conditions of the distribution function starting from values lies between 0 to 1, monotonically increasing and left limit is 0, right limit is 1 the fourth condition here it is the right continuous function.

Therefore, this function F, that is the first it is distribution function, the way we defined the distribution function through $P\{X \leq x\}$. Therefore, this is a CDF of the random variable X. It satisfies all the four conditions including it is a right continuous function. Therefore, this distribution function is the cumulative distribution function of the random variable X.

Similarly, we can create examples for CDF, which is continuous function, similarly one can give the example for CDF which is continuous in some interval, it has jumps in some other interval we can create. So, the one example which I have created which has only jumps; we can create example of no jumps; we can create an example of jumps as well as increasing in some interval, so this is the easiest example.

So, in the next class we will classify the random variables based on the CDF of the random variable.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

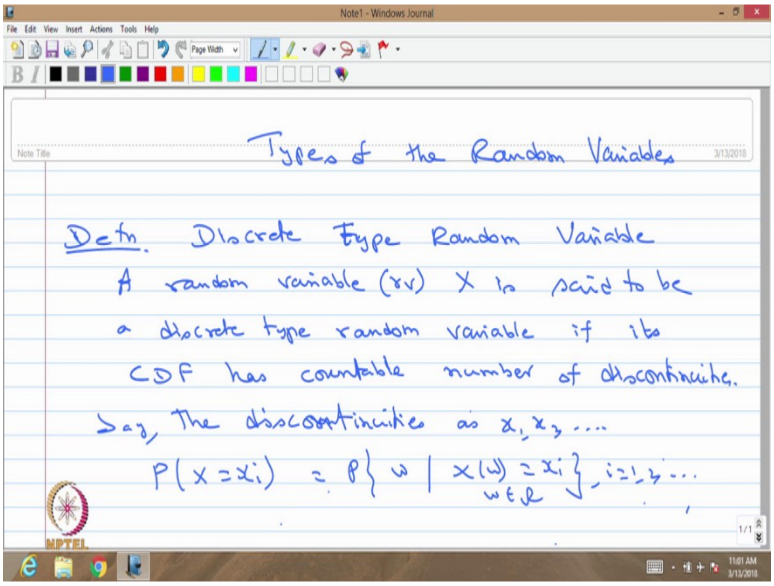
Lecture – 09

We are in week 2 and in the last class we have discussed the definition of a random Variable and then we have discussed the cumulative distribution function of the random variable. And we have given the one example for the CDF of the random variable also we have given five different CDFs in graphical way.

Now, we are going to discuss the types of random variable based on the CDF of the random variable. If you recall that five different diagrams which we have made it for the CDF out of those five different examples; we can classify those five examples into three types of random variable. First, we discuss the discrete type random variable and second, we discuss continuous type random variable and third, we discuss mixed type random variable. So, any random variable can be classified into any one of these three types of random variable; namely discrete type random variable, continuous type random variable or mixed type random variable.

So, in this class I am going to explain the types of random variable with the form of, first the definition of the discrete type random variable, then continuous type random variable, then mixed type random variable. And followed by the definition I am going to give one or two examples for each type; so that is a plan for today's lecture. So, let me start with the first type. So, it is called types of the random variable.

(Refer Slide Time: 01:53)



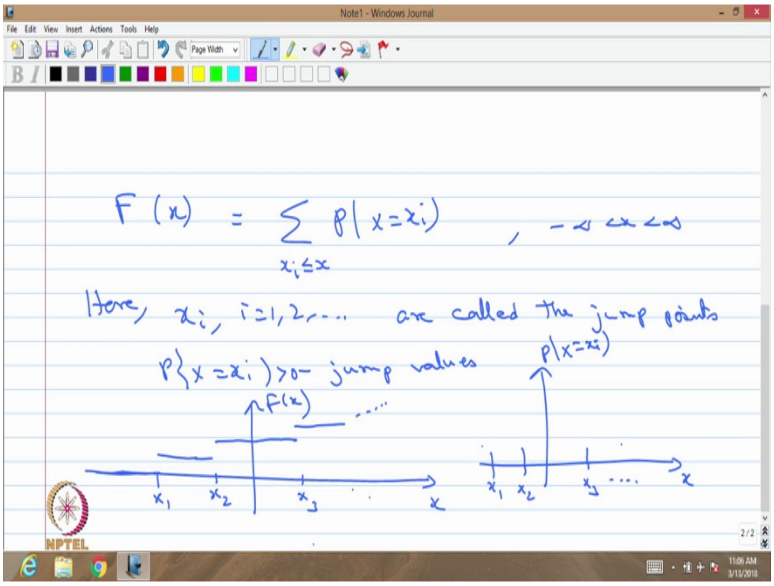
Types of the random variables; the first definition is discrete type random variable; when we say the random variable is discrete type random variable. A random variable, in short, we will be keep writing rv; a random variable X is said to be a discrete type random variable if its CDF, cumulative distribution function has countable number of discontinuities.

The countable number of discontinuities means, the discontinuities are jumps, can be finite or countably infinite; that means, a random variable X is said be a discrete type random variable; if its CDF has a countable number of a discontinuities or it has a countable number of jumps. Say, the discontinuities are jumps as x_1, x_2 and so on, initially I am writing countable infinite it could be finite also.

Then one can define the $P\{X = x_i\}$ that is nothing, but $P\{w: X(w) = x_i\}$ where w is belonging to **Ω**. So, this is for i is equal to 1, 2 and so, on; you have a probability space (**Ω**, F, P); P is a probability measure. So, the $P\{X = x_i\}$ that is nothing, but collection of possible outcomes gives the probability in which the X(w) gives the value x_i. So, you collect those possible outcomes, which gives the values x_i under the operation capital X; so, that is nothing, but the event.

So, find out the probability of event that event is going nothing, but the $P\{X = x\}$.

(Refer Slide Time: 05:30)



Therefore when you write the CDF of the random variable X; I can go for suffix X or I can omit the suffix x; that means, I am talking about the CDF of the random variable X that is nothing, but collection of all the $P\{X = x_i\}$ such that all the x_i's $\leq$ x.

So, here x is a between - ∞ to ∞. So, whenever you have a discrete type random variable the CDF is nothing, but collection of adding a $P\{X = x_i's\}$ where x_i's $\leq$ x. Here, the x_i's where i is equal to 1, 2 and so, on; it could be finite or it could be countably infinite or called the jump points why it is called the jump points? The CDF of the, this random variable has the jump only at these points and the $P\{X = x_i\}$ that is nothing, but the jump values.

So, this jump values are strictly greater than 0. So, wherever the X = x_i which the P is equal to greater than 0, $P\{X = x_i\} >$¿ 0. So, these values are called a jump values at those jump values the CDF has the jump. And wherever there is a jump and those points are called the jump points. that means, the $P\{X = x_i\}$ for different x it could be either 0 or greater than 0; if it is greater than 0 then those xi's are called jump points. And the $P\{X = x_i\}$ in which it is greater than 0 and those are called the jump values.

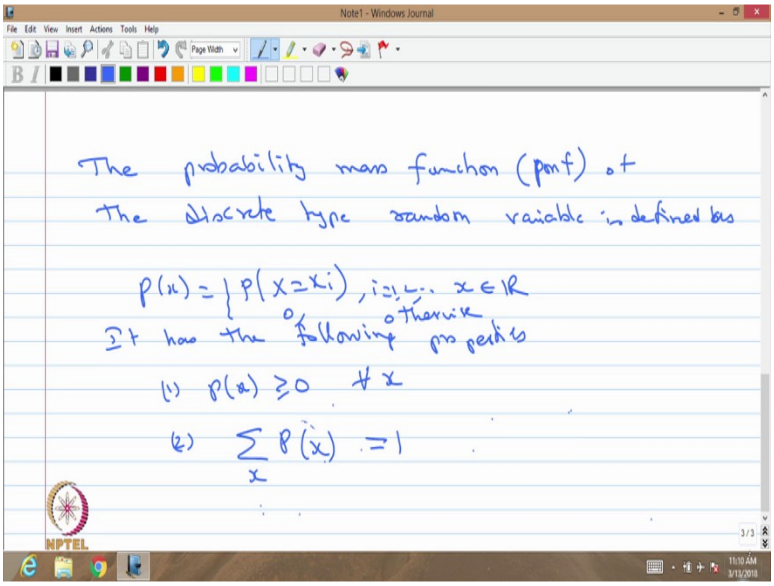
Therefore, I can draw the CDF of, in general CDF of a discrete type random variable as suppose x_1 is here, then the CDF is 0 till x_1, at the point x_1 there is a jump. So, that height is $P\{X = x_1\}$ that is a jump value and this x_1 is jump point, till the next point x_2 the value same and the CDF has a jump at the point x_2 suppose at x_2; the next jump is some value. So, $P\{X = x_2\}$ there is a jump.

So, this difference height that is a jump value till x_3 it has the same value then there is a next jump, like that it may keep going. And you know that it is a CDF therefore, it always starts from 0 it will land up 1. And this CDF has only jumps that is very important then only it is going to be call it as a discrete type random variable and other points the values is 0.

I can equivalently draw the another diagram for $P\{X = x_i\}$; that means, at point x_1 it has some height and x_2 it has another height and x_3 it has another height; suppose this height need not be same, it can be a different values. So, these heights are nothing, but the jump values and these are all the jump points like that I may have many more. So, whenever you have a CDF of some random variable; which has only jumps, the jumps could be finite or countably infinite then that random variable is called it as a discrete type random variable.

The way I have written $P\{X = x_i\}$ which has the jump values and jump points; I can define a another function that is called probability mass function.

(Refer Slide Time: 10:28)



The probability mass function, in short, it is pmf, the probability mass function of the discrete type random variable; as the probability mass function of a discrete type random variable is defined as p of x is nothing, but the $P\{X = x\}$ this p(x) is called the probability mass function. So, this is defined for all x belonging to R real line it has it has the following properties.

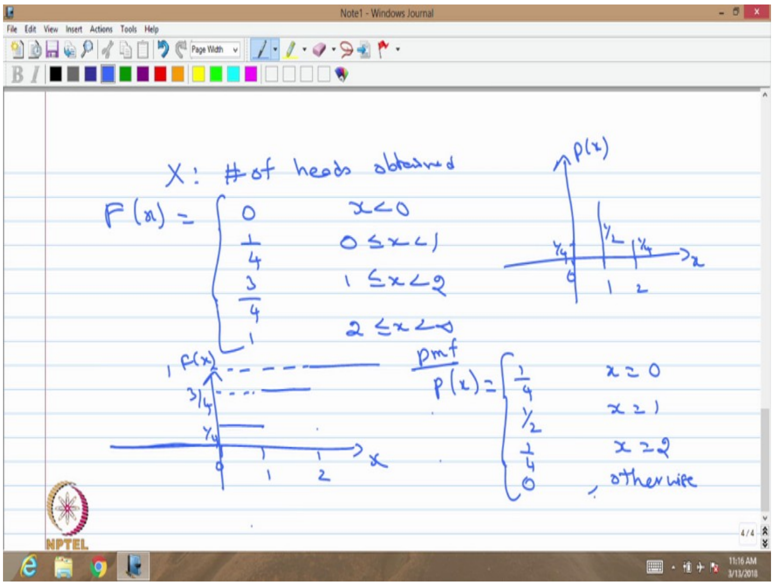
The first property; since we made p(x) is probability of X takes the value x, probability of X equal to x. So, wherever there is a probability of X equal to x it is this much otherwise it is 0;

this is the way the of X is defined it has the property the $p(x) \geq 0$ for all x; that means, wherever there is a jump value that value is going to be strictly greater than 0 wherever there is no jump; that means, the $p(x) = 0$.

The second property if you add all the values of a $P\{X = x\}$, you will get $p(x)$ summation will be 1. Always the probability mass function satisfies these two properties, it is greater than or equal to 0 and the summation over x its going to be 1. So, if you have a discrete type random variable; it has a probability mass function satisfies these two conditions or these two properties. If any real valued function satisfying these two properties; then one can say this is the probability mass function of some discrete type random variable. So, any probability mass any discrete type random variable has probability mass function satisfying these two properties or any real valued function satisfying these two properties is the probability mass function of some discrete type random variable.

I will go for one easy example which we have discussed earlier in the form of CDF.

(Refer Slide Time: 13:52)



Recall, there is a one problem in which the random variable X is number of heads obtained for the random experiment of tossing two unbiased coins, when you toss a two unbiased coin. And if you define the random variable X is a number of heads obtained then the $F(x)$ is 0 till $x < 0$, it is 1/4 from 0 to 1; 3/4 it is from 1 to 2. And the value is going to be 1 from 2 to ∞; so, this is the example we have discussed in the CDF for the random variable.

Now, we are discussing the same example if you draw the CDF of this random variable CDF F(x). So, till 0; it is 0, at 0 there is a jump of 1/4, at the point 1 there is a next jump it is 3/4; the value at the point 1 is 3/4. So, the jump values 3/4 – 1/4, so, 1/2. At the point 2 there is another jump of 1/4; so, it will be 1 from 2 onwards. Therefore, the p(x), at x is equal to 0; it is 1/4 at x is equal to 1; 1/2 at x is equal to 2; it is 1/4 and all other point it is 0 all other points. So, I write otherwise say when we write otherwise; that means, the probability mass function value is 0 at sorry it is 1/4 at 0; 1/2 at 1 and 1/4 at 2 and all other points it is 0. So, I can draw the probability mass function also.

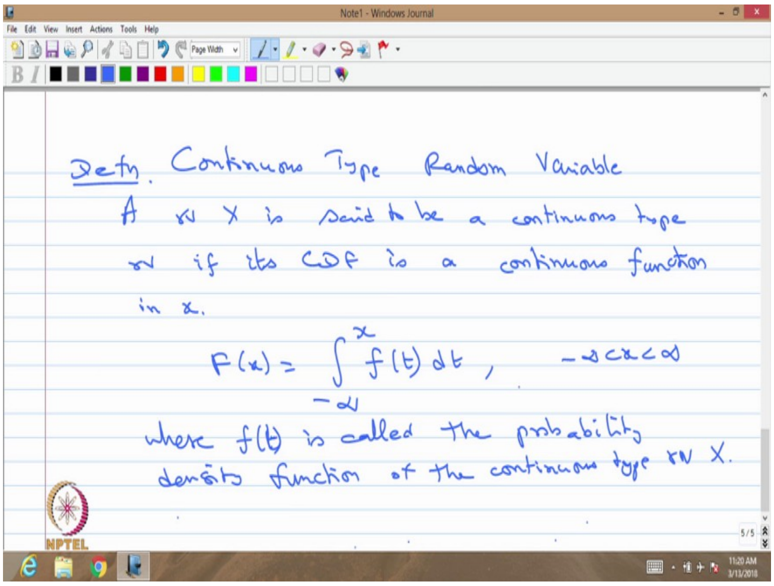
So, at 0 there is a height of 1/4 and at x is equal to 1 at height of a 1/2 and at x is equal to 2 there is another height 1/4 and all other points it is 0. And if you add all the jumps the jump values 1/4 + 1/2 + 1/4 that is 1; that is the second property and all the values are greater than or equal to 0; that means, at only these three jump points it has the jump values and all other place the values are 0. Therefore, this is the probability mass function of the discrete type random variable; why it is called a discrete? Because you see the CDF the CDF has only jumps and the jump points are 0, 1 and 2 and the jump values are 1/4, 1/2 and 1/4.

Since the CDF has discontinuities, therefore, this random variable that is number of heads obtained, this random variable X that is a discrete type random variable. And the probability mass function of this random variable X is p(x) that is 1/4 at x equal to 0; 1/2 at x equal to 1 and x equal to 2 is 1/4; otherwise it is 0. Since it has the three jump points and three jump values; therefore, I am just writing the probability mass function this way and 0 otherwise. There is a possibility in general you may have a CDF as a finite number of jumps or countable infinite number of jumps. And one can find out what is a probability mass function of that discrete type random variable also.

So, this is the very easiest example in which we can represent the CDF first; by seeing the CDF you can conclude it is a discrete type random variable. And from the CDF you can get the probability mass function by subtracting because the CDF is nothing, but the summation of probability mass till that point. So, from the CDF one can get the probability mass function or from the probability mass function one can get the CDF of a discrete type random variable.

Now, we will move in to the second type of a random variable that is a continuous type random variable.

(Refer Slide Time: 19:45)



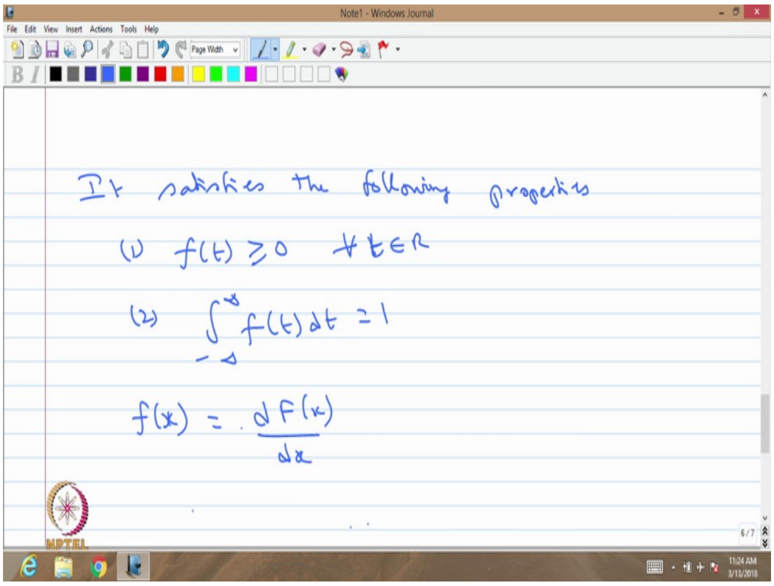
Continuous type random variable when we say the given random variable is of continuous type. A random variable X is said to be a continuous type continuous type random variable; if its CDF is the continuous function in x.

That means, whenever the CDF of random variable is a continuous function in x in the whole range from - ∞ to + ∞. That means, there is no discontinuity in the CDF of the random variable; then you can conclude it is a continuous type. For a discrete type random variable; it has a finite or countably infinite discontinuities, but for continuous type random variable, the CDF is a continuous function in the whole range from - ∞ to ∞. That means, if you draw the CDF without lifting pen or pencil, you draw the CDF then that is a continuous type random variable CDF.

Since it is a continuous function one can write the continuous function, left hand side is the continuous function in the form of $\int_{-\infty}^{x} f(t)\,dt$, this is possible whenever the function is a continuous function in the whole interval - ∞ to ∞. Here, x is from - ∞ to + ∞ because of the CDF is the continuous function where the f(t) till now we are using the F; now I am start using small f, where f(t) that is called the probability density function of the continuous type random variable X. This is possible whenever the CDF is a continuous function. So, whenever the CDF is a continuous function, we call that random variable as a continuous type random variable.

The integrant in this equation, that integrant is called probability density function. The way the probability mass function satisfies a few properties, the two properties. Similarly, one can identify what are all the properties going to be satisfied by this probability density function of the continuous type random variable.

(Refer Slide Time: 23:48)



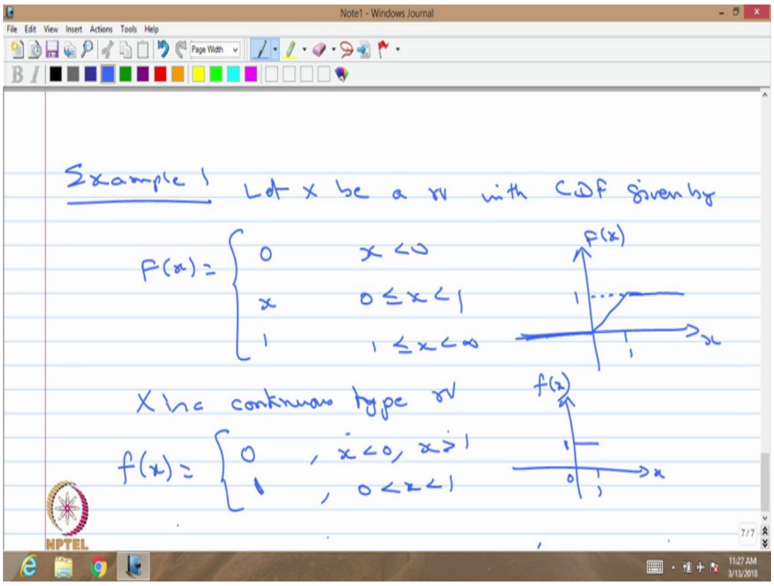
So, it satisfies the following properties; that means, the probability density function satisfies the following properties you know that the CDF values lies between 0 to 1; it starts from 0 it will end up 1 and monotonically increasing continuous or right continuous function.

Therefore, the integrant is always greater than or equal to 0 for all t; not only that since you are writing the F(x) = $\int_{-\infty}^{x} f(t)\,dt$. So, if you go for limit x tends to ∞ you know that limit x tends to ∞ of F(x) is 1 from there you can get the second property that is $\int_{-\infty}^{\infty} f(t)\,dt = 1$; that means, a real valued function satisfying these 2 properties will be the probability density function of some continuous type random variable.

If you have a continuous type random variable whose CDF is the continuous function in the whole real line - ∞ to ∞. Therefore, you will get the probability density function; so, we can relate, suppose you know the CDF you can get the probability density function I am writing f (x) either f(t) or f(x) does not matter by differentiating the CDF with respect to x, you will get the probability density function. From the CDF you can get the probability density function, from the probability density function you can get the CDF.

We are going to do some example through that we will explain in detail also; I am going for two examples through that I am going to explain this continuous type random variable.

(Refer Slide Time: 26:27)

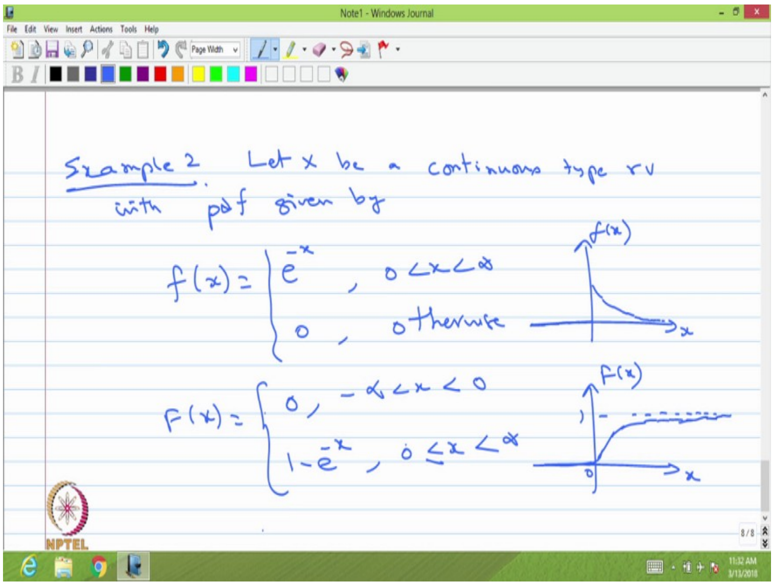


The first example, let us consider the example 1; let X be a random variable with the CDF given by F(x) that takes a value 0 till 0 and it takes a value x between 0 to 1 and it takes a value 1 from 1 onwards. You can draw the CDF it is a rough diagram. CDF is 0 till 0 from 0 onwards it is x and at the point 1 onwards it becomes 1. By seeing the CDF, you can conclude this is a continuous function from - ∞ to + ∞ continuous function in x; therefore, this is a continuous type random variable therefore, X is a continuous type random variable.

You can find out what is the probability density function of this continuous type random variable by differentiating the CDF. So, if you do the differentiation of CDF you will get 0 when $x < 0$. Similarly, $x > 1$ and if you differentiate x you will get 1 between the interval 0 to 1; that means, the probability density function of this continuous type random variable; the value is 1 between the interval 0 to 1.

And otherwise it is 0; that means, you can draw, this is the rough diagram the probability density function this is f(x); the F(x) is the CDF, f(x) is the pdf. So, in short it is the pdf; so, the value is between 0 to 1; the height is 1 otherwise it is 0. So, the horizontal line 0 to 1, the height 1 that is the probability density function of these are continuous type random variable.

(Refer Slide Time: 29:33)



I will go for one more example; example 2, let X be a continuous type random variable with pdf given by f(x) takes a value e^{-x} between 0 to ∞; 0 otherwise. Whenever we say 0 otherwise or elsewhere from books they use a word elsewhere. That means, the probability density function which is positive value between the interval 0 to ∞ is e^{-x} and remaining intervals it is 0. So, here the remaining interval is - ∞ to 0.

So, you can draw the probability density function. So, e^{-x}; so, it takes a value, it is going down asymptotically touches 0 at ∞ and at 0. So, the f(x) value is 0 till 0. From the probability density function, you can get the CDF by integrating - ∞ to till that point of probability density function.

So, since the probability density function is 0 from - ∞ to 0; you can say it is 0 from - ∞ to 0. From 0 to ∞ you have to integrate and find out; so, if you do the simple integration you will get 1 - e^{-x} between the interval 0 to ∞; you can include 0 also.

So, - ∞ till 0 and 0 onwards; it is 1 - e^{-x}. You can draw the CDF of this continuous type random variable. So, till 0 it is 0 and 0 to ∞ it increases, it touches 1 at ∞ and this is the CDF of discrete sorry; this is the CDF of a continuous type random variable. So, from the CDF you can always get the probability density function of a continuous type random variable; from the probability density function one can get the CDF of the continuous type random variable.

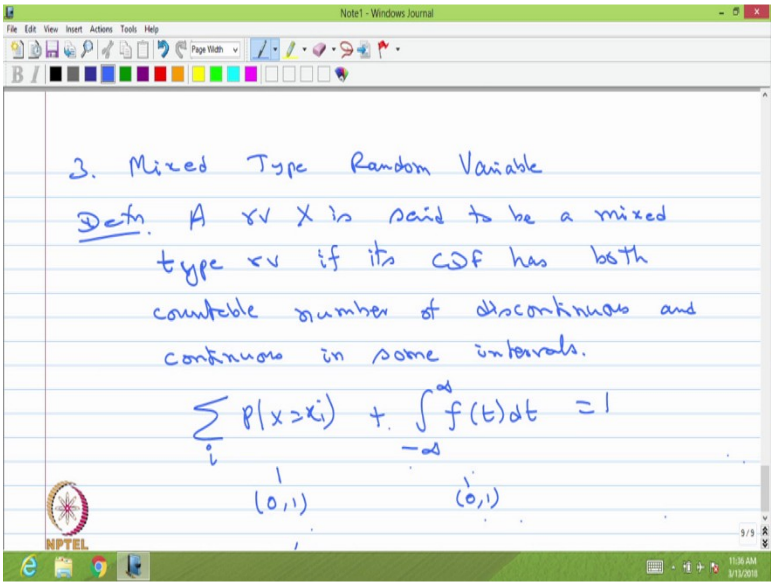
Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 10

Now, we will move into the third type of a random variable.

(Refer Slide Time: 00:06)



That is mixed type random variable; the definition, A random variable X is said to be a mixed type random variable, if its CDF has both countable number of a discontinuities and continuous in some intervals.

You see the definition very carefully for a discrete type random variable, it has countable number of discontinuities, for a continuous type random variable it is a continuous function in X that means, in the whole real line. Whereas, mixed type random variable is the combination of both; that means, the CDF has discontinuities, as well as a continuous function in some interval. In that case that random variable is going to be call it as a mixed type random variable.

In this case, if you add the masses at countable number of discontinuities points and if you integrate the probability density; both will give the value 1, that is very important I am writing integration from $\int_{-\infty}^{\infty} f(t)\,dt$; here the f(t) value will be greater than 0 in some interval.

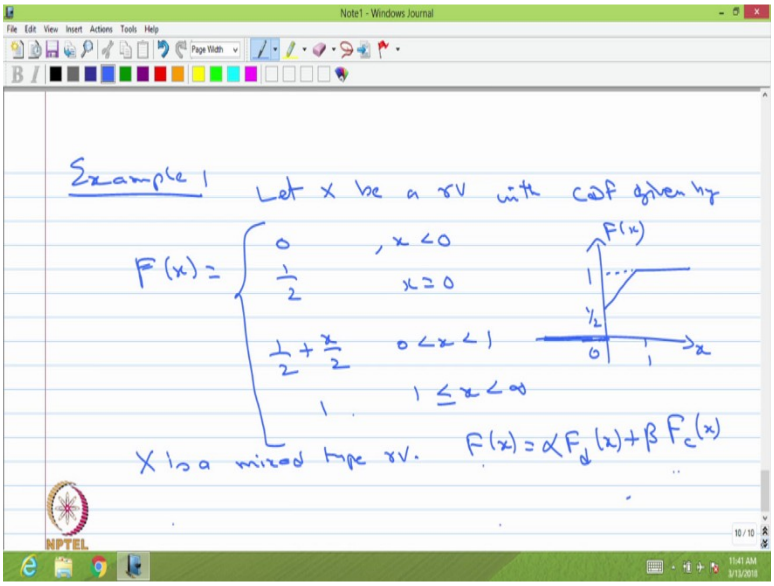
Definitely the integration alone will not give the value 1, as well as the $\sum_i P(X=x_i)$ alone may not give value 1.

If you add both it is going to be 1; that means, this value lies between open interval 0 to 1 and this value also lies between open interval 0 to 1 and put together that is going to be 1. That means, the unit mass is distributed over the countable number of a points in the real line as well; as in few intervals or some intervals; the whole unit mass is distributed both in countable number of a points those are jump points and the values are jump values and the density between some interval.

Whereas for discrete type random variable the unit mass is distributed over countable number of points; for a continuous type random variable the unit mass is distributed over some intervals. Whereas, for mixed type random variable it has both jumps as well as density in the interval.

So, I am going to give one simple example through that example you will understand how the mixed type random variable look like. The way I started today's class I have given five different CDFs in the previous class. So, out of those five in different CDF; few CDFs are going to be a discrete type, few CDFs are going to be continuous type and few CDFs are going to be mixed type. So, I am going to give the example for the mixed type random variable.

(Refer Slide Time: 04:47)



Example, only one example I am going to introduce for this mixed type random variable let X be a random variable with the CDF is given by F(x); it takes a value 0, it is less than 0, it takes a value 1/2 when x takes the value 0. Between the interval 0 to 1 it is 1/2 + x/2, from 1 onwards the value will be 1.

The CDF of this particular random variable is 0 till 0 at 0 there is a jump from 0 onwards till 1 it is 1/2 +x/2, from 1 onwards it the value is 1. So, I can draw the CDF of this random variable till 0; it is 0 then at the point 0 there is a jump and from 0 to 1, 1/2 + x/2. So, it is landing line till 1, then from 1 onwards it becomes 1.

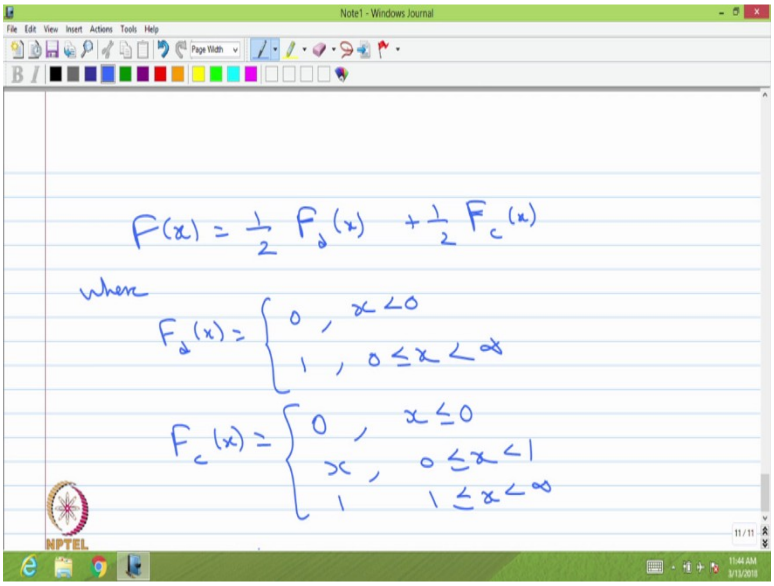
You see that this particular CDF is 0 till 0 at 0 there is a jump then it is a continuous between 0 to 1; then from one onwards it is 1. So, basically it is a continuous from 0 to ∞ whereas, 0 there is a jump and from - ∞ to 0 it is 0. You cannot conclude that this random variable is a discrete type random variable because it has jump as well as there is a continuous function between 0 to ∞.

You cannot consider this as a continuous type random variable because it has a jump also whereas; this is going to be a mixed type random variable because it has the one jump and the continuous between the interval 0 to ∞ by seeing the CDF. The CDF has the jump as well as the continuous in the interval 0 to ∞ therefore, it is a mixed type.

In general, you may have a countable number of jumps in the CDF and the continuous function in many intervals in this case it is only 1 interval. Therefore, this is the mixed type random variable in general; one can always right the CDF of any random variable of discrete part and the continuous part. So, the suffix means the CDF with the discrete part discrete type; the CDF with the continuous type.

So, the mixed type random variable has a CDF in the form of sum of α times the discrete part as well as β times the continuous. So, one can write for this example this is going to be F(x).

(Refer Slide Time: 09:09)



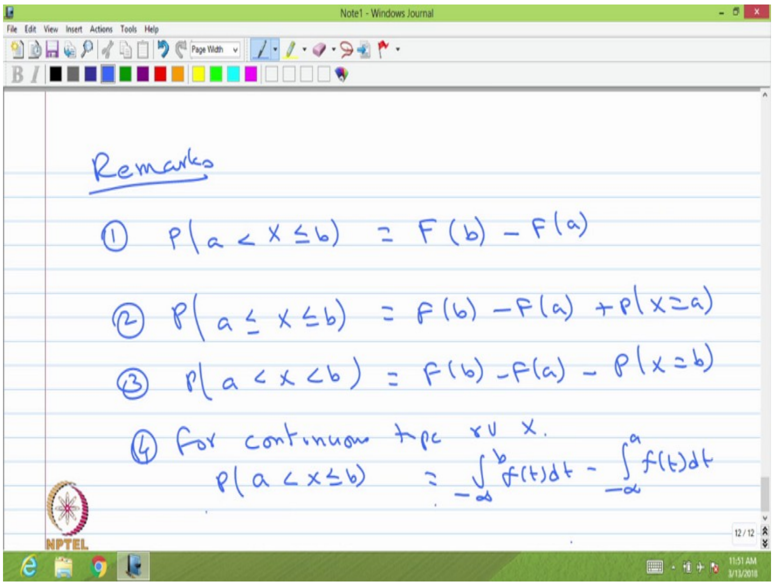
I can find out the α is going to be 1/2 of a discrete part times discrete part + 1/2 times the continuous part where the discrete part that itself CDF less than 0 till 0. And 1 from 0 onwards whereas, the continuous part of CDF that is again a CDF till 0 it is 0 and it take a value x between 0 to 1 and it becomes 1 from 1 onwards.

So, the discrete part $F_d(x)$ that alone it is a CDF of a discrete type random variable which has the mass at 0 and the jump value is 1. And $F_c(x)$ is it is a continuous part of CDF and that itself a CDF which is 0 till 0 and x between 0 to 1 and 1 onwards it is 1, but that is a continuous type random variable CDF for mixed type random variable it is both that is $\frac{1}{2}F_d(x)+\frac{1}{2}F_c(x)$.

So, if you have any continuous type random variable you do not have a discrete part, if you have a discrete type random variable; there is no continuous part, if you have a mixed type random variable you will have a CDF in the some constant times discrete part + constant times continuous part in this constant one can identify. So, this call it as a canonical form of a CDF; the way we have written F(x) is α times discrete type + β times the continuous type that we call it as a canonical form of a CDF. So, this is very important for immaterial of the random variable is a discrete type or continuous type or mixed type for every type of random variable; it will be simplified.

So, with this definition of a discrete and continuous and mixed type random variables and one or two examples I am concluding any random variable can be classified into the discrete or continuous or mixed type by seeing its CDF. If it is discontinuous only then it is a discrete if it is a continuous function then it is call it as a continuous type; if it has both then it is going to be call it as a mixed type random variable.

(Refer Slide Time: 12:37)



I have to give some more remarks over the random variable with respect to CDF. So, I am going to list out those remarks. The first remark you can always find the probability of X lies between any interval. You can always find probability of X lies between a and b for a < b, a can be a real as well as b can be real where a < b. One can find the probability of X lies between a to b that is nothing, but the CDF of the random variable at the point b - the CDF at the point a; one can always find out the probability of X lies between a to b by using the CDF by substituting the value at x equal to b and x equal to a.

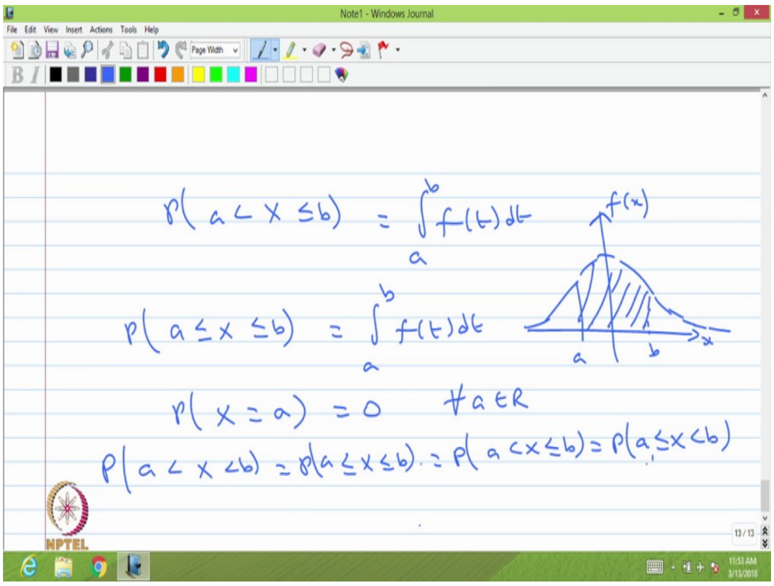
Whereas, if you want to find out the probability of X lies between a ≤ X ≤ b; then it is same as F(b) – F(a) and you have to include the P{X = a}. Suppose you want to find out the P{a < X < b} this also can be computed in the same way it is F(b) – F(a) – P{X = b}.

Now, the question is when you need to add P{X = a} or when you have to subtract P{X = b}; based on which type of random variable you are in the discussion. If it is discrete type random variable and X is equal to a; where a is a jump point such that a probability of X = a > 0, then you can add. If X is discrete type random variable, X is a discrete type random variable where a is not a jump point; that means, P{X = a} = 0 then you do not need to add.

So, based on the discrete type random variable in which it is a jump point or not that is one discussion. The second discussion when X is a continuous type random variable when X is a continuous type random variable the P{X = a} that is nothing, but the integration and it is a Reimann integration. So, the probability of for continuous type random variable X, for a continuous type random variable; the P{a < X ≤ b} that is nothing, but the integration from

$$\int_{-\infty}^{b} f(t)\,dt - \int_{-\infty}^{a} f(t)\,dt.$$

(Refer Slide Time: 16:26)



That is same as the P{a < X ≤ b} that is nothing, but integration from $\int_{a}^{b} f(t)\,dt$ when X is a continuous type random variable.

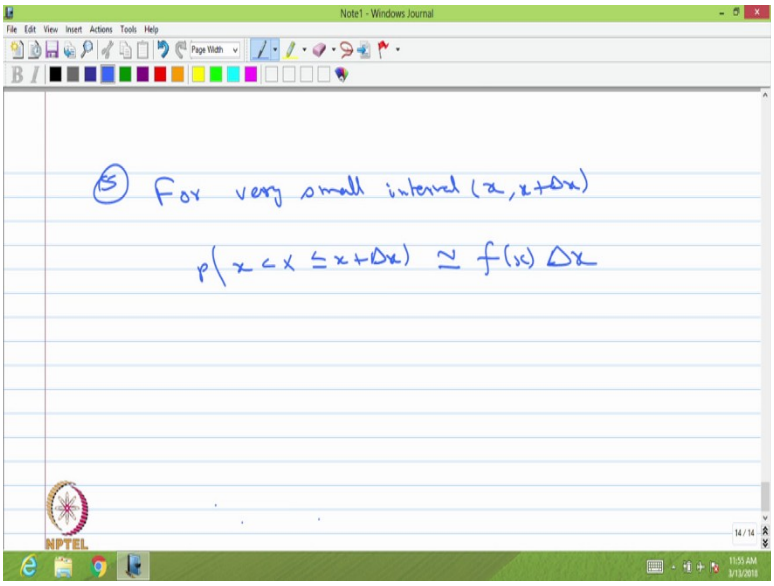
Since it is $\int_a^b f(t)dt$ that is nothing, but suppose you have a probability density function like this, then a to b is nothing but, suppose b somewhere here this is the. So, this is basically a Reimann integration of f(x) between the interval a to b. Therefore, the P{a ≤ X ≤ b} that is also again it is $\int_a^b f(t)dt$ when X is a continuous type random variable; whether you include the point or exclude the point the integration is going to be again $\int_a^b f(t)dt$.

Hence, the probability of X takes any point that is going to be 0 for all a belonging to R; hence for a continuous type random variable the probability of X lies between a to b open interval probability of X lies between closed interval or one side closed, all the values are one and the same. Because the probability of X takes any point in the real line that value 0 for a continuous type random; that means, there is no mass at any point whereas, for a continuous type random variable there is a density between some interval therefore, that is going to be greater than or equal to 0; so, this is going to be the next remark.

Therefore, the previous result the result of 2 and 3 whether you have to add or subtract the P{X = a} or P{X = b} this is going to be the issue for a discrete type random variable, for continuous type random variable it is going to be 0. For a discrete type random variable if it is a jump point then there will be addition, there will be a subtraction if that is not the jump point again it is going to be 0.

The next remark the fifth one; the probability mass function is the probability at that point X takes the value x_i for a discrete type random variable. Whereas, there is no probability of X takes some value for a continuous type random variable the probability is 0.

(Refer Slide Time: 19:40)



But still for a smaller for very small interval x to x + Δx; Δx is very small, then you can make out the P{x < X ≤ x + Δx} it whenever a Δx is very small that can be approximately say the f(x) Δx; as it is f(x) is not a probability, f(x) is a probability density function. If you integrate between the interval you will get the probability in that interval, but if the interval is very small then you can make approximately it is going to be f(x) Δx, that is for when Δx is very small.

So, if these are very important, these five remarks about the random variable; whenever you know the CDF you can get the probability of X lies between the interval immaterial of it is a discrete or continuous. If it is a continuous type random variable then whether it is open interval or say closed interval or semi closed and so on everything is one and the same because you are integrating the probability density between the interval.

Therefore, some books they write the probability density function in the open interval; otherwise it is 0, some books they use probability density function in the closed interval. So, whether you write the open interval or closed interval.

It is immaterial because you are going to do the Reimann integration to find out the probability of X lies between any interval; for a continuous type random variable. For a discrete type random variable; it is sum of probability masses at the jump points. If it is a mixed type then it is combination of both; with these remarks along with the definition and the examples I am concluding the types of random variable.

Introduction to Probability Theory and Stochastic Processes

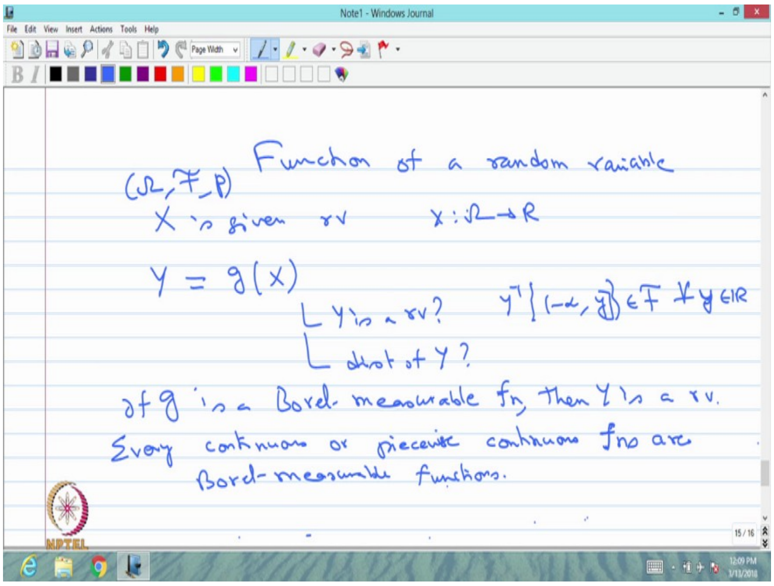
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 11

In week 2, we are discussing the random variable and we started with the definition of the random variable. Then, we discussed the distribution function, then we have related the distribution function with the random variable with the probability in the form of, we get the CDF; CDF of the random variable.

And by seeing the CDF, we classify the random variable or we get the types of random variable that is a discrete type or continuous type or mixed type random variable.

(Refer Slide Time: 00:36)



So, now we are moving into the function of random variable. Function of a random variable. Sometimes, we started creating one random variable and then we come to the conclusion oh we need another random variable, what to do?

So, we can always create a new random variable from the scratch that is our real valued function from Ω to R satisfying the condition, then we can conclude that is a random variable.

But the easy thing is, suppose you are able to relate the already found one random variable with the new random variable in some form, then you can find the new random variable and its distribution through the already existing random variable.

So, the existing random variable, we use the letter X is given and we are interested to create a new random variable that is Y. Instead of creating a new real valued function and find out we make a relation in the form of earlier random variable as g(X); where g is the function from R to R.

I will repeat X is the given random variable that is defined from Ω to R. Already we know it is a random variable and we know the CDF of the random variable. From the CDF, we know it is a discrete type or continuous type or mixed type random variable. We are interested to find the distribution of another random variable that is Y; for that we identify the relation that is g(X).

Now, the first question is whether the way we make a relation g(X), whether that is going to be a random variable? That means, whether Y^{-1}¿ that is belonging to the same probability space? So that means, we have a given probability space (Ω, F, P) and in this probability space X is a given random variable.

And now, the first question is whether Y is a random variable. That means, whether it satisfies Y^{-1}¿ that is belonging to F for all y belonging to real. If it is a random variable, then the second question is what is the distribution of the random variable Y?

I repeat the issue; we have a probability space; we have a one random variable and we are interested to find the distribution of the other random variable for that we make the relation g(X). Therefore, first question is whether Y is a random variable and then, if Y is a random variable what is the distribution of Y?

First, we can answer the first question whether Y is a random variable. Whenever X is a random variable and g is Borel-measurable function; then if g is a Borel-measurable function, then Y is a random variable. Now, what is the Borel-measurable function, in the measure theory course, one can study if you have a real valued function; the inverse image of a Borel set is belonging to the Borel set.

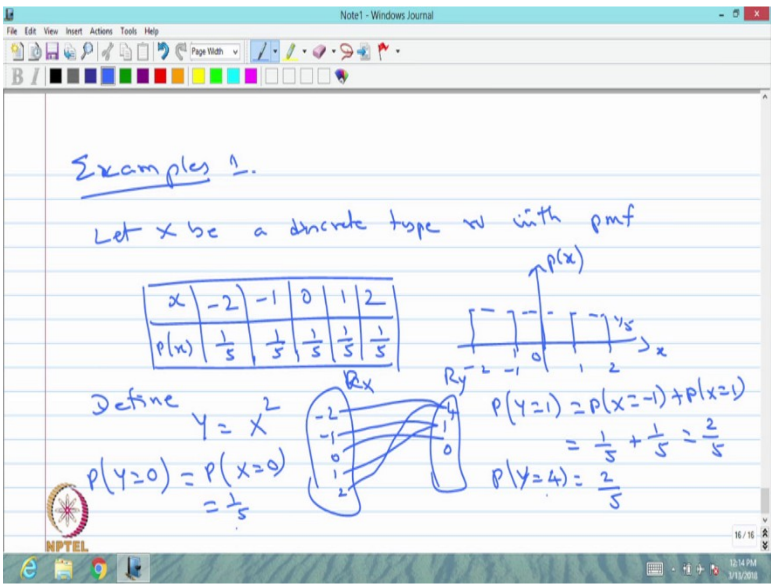
Then, we can conclude the given real valued function is a Borel-measurable function. As far as this course is concerned, we do not need to worry about the Borel-measurable function and so on. We can take it every continuous or piecewise continuous functions are Borel-measurable function. So, we can make sure whether the g is going to be a continuous or piecewise continuous function; therefore, it is going to be a Borel-measurable function. Therefore, Y is a random variable.

So, we can use this concept every continuous or piecewise continuous; one should know the definition of piecewise continuous. So, every continuous or piecewise continuous functions are Borel-measurable functions. Therefore, the first question is answered. So, as far as this course is concerned, we always give g such a way that Y is going to be a random variable.

Now, the question is how to find the distribution of Y. Distribution means in general, it is CDF of the random variable. If you know that it is a discrete type random variable, if you find the probability mass function of that, that is also called distribution function. If you know that it is a continuous type random variable, then if you find the probability density function of the random variable; then that is also called the distribution function.

So, in general, distribution means cumulative distribution function of the random variable. Otherwise, it could be a probability mass function or probability density function based on the random variable is discrete type random variable or continuous type random variable. Now, we will start to find out the distribution of function of random variable with the few cases.

(Refer Slide Time: 07:05)



So, let us start with the examples, through that we will study the distribution of function of random variable. The first example, let X be a discrete type random variable with the probability mass function given for different values of x, what is a p(x). So, this way also one can give.

Suppose, x takes the value -2, -1, 0, 1 and 2; what is the probability mass at those points. Then, we are defining the probability mass function; that means, other than this points the values are 0 and if you add this is going to be 1. Therefore, this is probability mass function.

So, let us give some values such a way that it is going to be probability mass function. So, we are going for the example to discuss function of random variable. Therefore, I am just going for the easy example in which the probability mass function at the point. So, this is a probability mass function at the point -2, -1, 0, 1 and 2; all the values are 1 by 5.

Now, I am going to define new random variable or new real valued function which is $Y = X^2$ is an easiest function. So, the possible x values are -2, -1, 0, 1 and 2. The way we defined Y ¿ X^2, now the random variable Y is going to be a discrete type; because for different values of X, the Y values are going to be either 0 or 1 or 4 because Y ¿ X^2. Therefore, the way I have made the relation Y ¿ X^2 and X is a discrete type random variable.

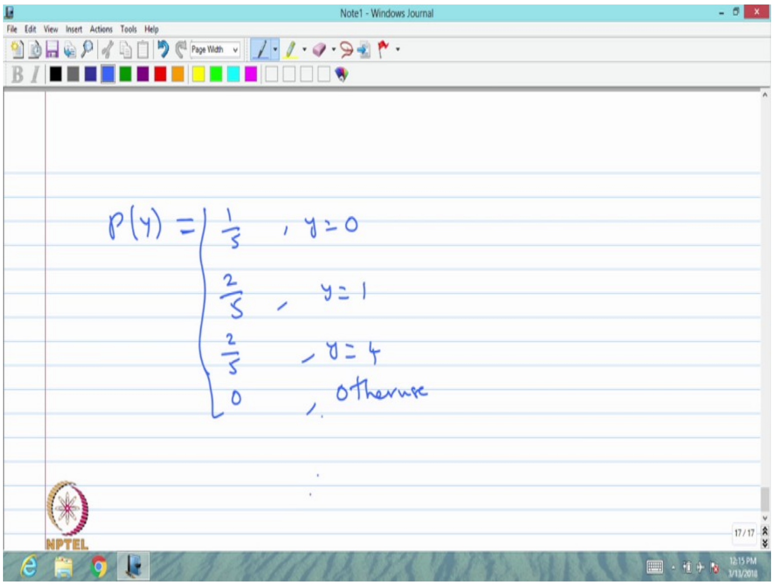
I can very well say the random variable Y is of the discrete type. Therefore, I can go for either CDF of the random variable Y or I can go for the probability mass function of Y. We can make a nice the possible values of X that is -2, -1, 0, 1 and 2 and this is mapped with the possible values of Y that is 0 is mapped with 0 and -1 and +1 is mapped with 1 and -2 and +2 is mapped with 4. So, this is the mapping from X to Y.

That means, if I want to find out the probability of Y takes a value 0 that is same as I have to go for what is the possible outcome or what is the possible values of X, which gives the value Y is equal to 0. So, X = 0 will give the value Y = 0. Therefore, P(Y = 0) =P(X = 0) and I know that P(X = 0) = 1/5; therefore, this is 1/5.

Similarly, I can find P(Y = 1) that is same as either P(X = -1) or P(X = 1); both will give the value Y = 1. Therefore, P(Y = 1) is nothing but it is a 1/5 + 1/5; therefore, it is 2/5.

Similarly, you can find P(Y = 4) that is X takes a value -2 as well as X takes the value +2; those probabilities we will put together, give the P(Y = 4). Therefore, that is again 2/5.

(Refer Slide Time: 11:55)

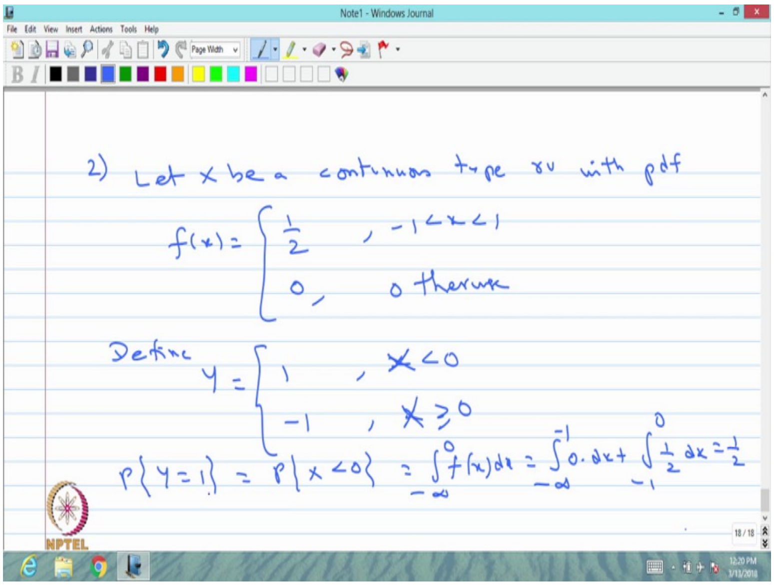


Therefore, I can make probability mass function for Y that is takes a value 1/5, when Y is equal to 0 and 2/5, when Y takes a value 1 and 2/5, when Y takes a value 4 and 0 otherwise. That means, from discrete type random variable and Y = X^2 gives the again discrete type random variable.

Because it has the mass at the 0, 1 and 4; if you add all the probability masses, it is going to be 1. So, this is the probability mass function and if draw the CDF, then it is 0; till 0, at 0 it is 1/5 height. Then, it is 1/5 till 1, at the point 1 it becomes 3/5. Then, 3/5 till 4 and at the point 4 onwards, it becomes 1.

Therefore, this CDF has 3 jumps. So, this discrete type random variable whose probability mass function is probability of Y takes the value 1/5, 2/5, 2/5 at those points otherwise it is 0. So, this is the easiest example in which we are getting discrete random variable into discrete random variable.

(Refer Slide Time: 13:21)



Then second example is let X be a continuous type random variable with since it is a continuous type random variable, I am going to give the probability density function of the random variable X that is takes a value 1/2 between the interval -1 to 1; otherwise it is 0. Whenever the problem is given, you can verify whether it is a correct. Probability density function, it is always greater than or equal to 0; that is a first property.

Second property, if you integrate in the whole interval - ∞ to ∞; it has to be 1. So, if we check - ∞ to - 1, the probability density function is 0. From -1 to 1, it is 1/2. So, if you integrate - 1 to 1, 1/2 you will get 1 and integration from 1 to ∞, again the probability density function is 0; therefore, it is 0. So, the whole interval - ∞ to ∞, the probability density function, that integration is 1; therefore, it is a probability density function.

Now, I am going to define new random variable. Why I am saying the random variable? That means, we are saying it is a Borel-measurable function; therefore, it is a random variable. So, we do not need a question about the whether it is a random variable or not we started with the random variable.

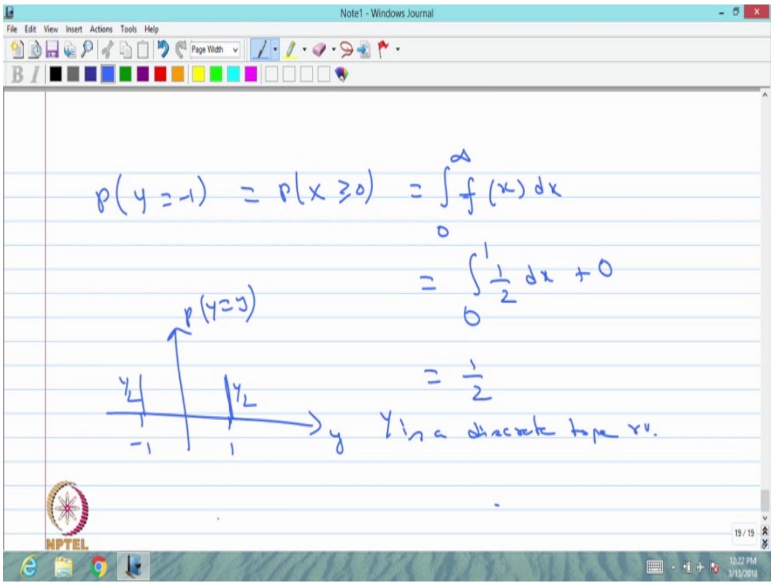
So, define a random variable Y which takes the value 1 when X is less than 0. It takes a value -1, when X takes a value greater than 0. I am defining a new random variable which takes the value 1, when $X < 0$; -1 when $X \geq 0$. I can write capital letter also. So, capital X is a random variable. So, when the random variable takes a value less than 0, it is a 1; when it is greater than 0, it is a -1. I can use greater than 0 also.

So, now you can find P{Y = 1}, that is same as the P{X < 0}, that is same as since X is a continuous type random variable X is less than 0 or less than or equal to 0; both are one and the same and it is same as - ∞ to 0 and the probability density function..

And the probability density function is greater than 0 between the interval - 1 to 0; therefore, it is $\int_{-\infty}^{-1} 0\,dx + \int_{-1}^{0} f(x)\,dx$ here it is $\int_{-1}^{0} \frac{1}{2}dx$. Therefore, you can integrate and you can get the value and this value is going to be 1/2.

The way we define Y takes a value 1 for X is less than 0. So, P{Y = 1} that is same as P{X < 0}.

(Refer Slide Time: 17:01)

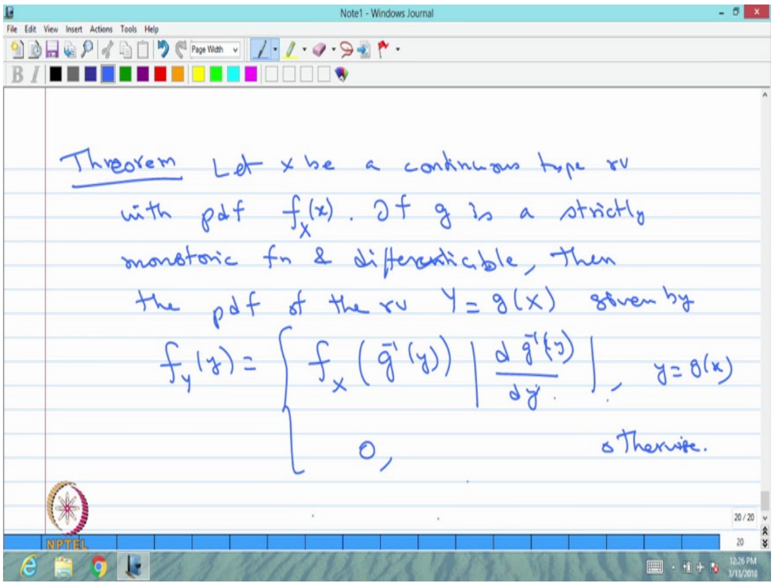


Similarly, P{Y = -1} that is same as P{X ≥ 0}; that is same as integration from $\int_0^\infty f(x)\,dx$ that is same as the again the f(x) > 0 between the interval 0 to 1; whereas, 1 to ∞ the density function is a 0. Therefore, it is a 0 to 1, f(x) is 1/2 the other quantity is 0. So, if you simplify you will get the value 1/2.

Therefore, it has the mass the probability mass at the point -1, 1/2 and at the point 1, it has another 1/2 and if you add both the masses it becomes 1. Therefore, this is a discrete type random variable. Y is a discrete type random variable. The first example X is a discrete type, Y is also discrete.

Now the X is a continuous type random variable. The way we defined the function Y, the Y is a discrete type random variable whose probability mass function is 1/2 at the point 1 and -1 otherwise it is 0.

(Refer Slide Time: 19:07)



Now, we are going for one general result as a theorem and this theorem will be useful whenever you want to find out the probability density function of a continuous type random variable, whenever X is also continuous type. So, I am going to give the theorem first; then we will give the proof of the theorem followed by one example, then I will conclude.

Let X be a continuous type random variable with probability density function is f(x). If g is a strictly monotonic function and differentiable, then the probability density function of the random variable Y = g(X) is given by, see the theorem, you can directly write the probability density function of the random variable Y in terms of the probability density function of X.

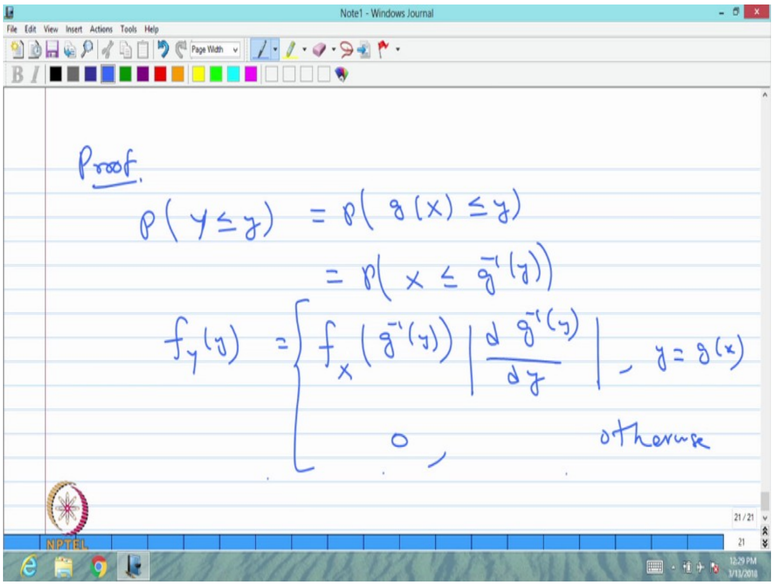
I can rewrite this as the suffix X. I am using capital small f for all the probability density function by writing suffix X or suffix Y, we know that we conclude it is a probability density function of X; it is a probability density function of Y.

So, the probability density function of Y, we can write it in the form of probability density function of X by replacing x by $g^{-1}(y)$; not only that by multiplying the ¿$\frac{d g^{-1}(y)}{dy}$∨¿.

So, this is going to be greater than 0 whenever y takes a value g(x); whenever y is not going to take the value g of x, it is going to be 0. That means, by just substituting x by $g^{-1}(y)$ in the probability density function of X, you will not get the probability density function of Y unless otherwise you multiply the absolute of $\frac{d g^{-1}(y)}{dy}$.

That means, you recall the probability density function has two properties; it is going to be greater than or equal to 0 and the integration is going to be 1. By multiplying this absolute quantity, the integration is going to be 1. Therefore, this multiplication absolute of ¿ $\frac{d g^{-1}(y)}{dy}$, that is called normalizing constant.

(Refer Slide Time: 23:01)



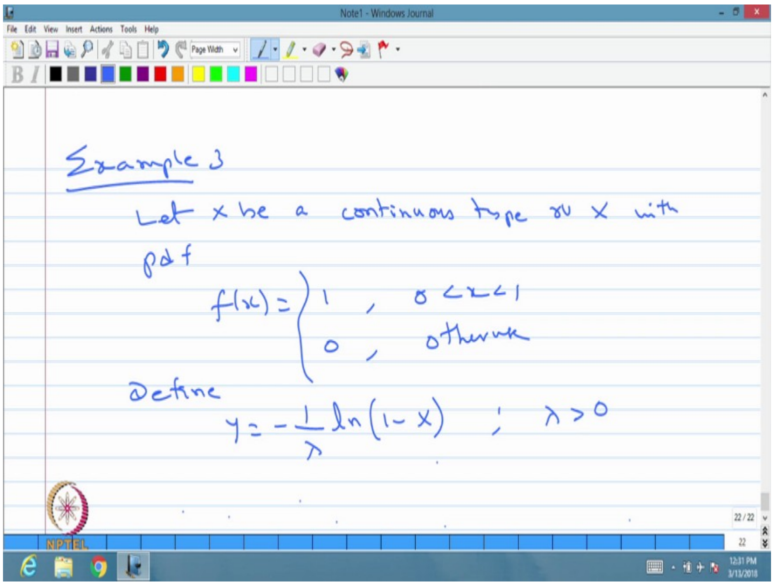
You can prove. I will give the proof for the one part since I am saying strictly monotonic it could be increasing or decreasing. So, we will do the one part, then similarly one can do the other part. So, you can find CDF of the random variable Y, that is nothing but the Y is replaced by g(X) less than or equal to y; whenever I write capital letter; that means, it is a random variable whenever I write the small letter; that means, it is a variable values.

So, I am directly finding the CDF of the random variable Y. I am replacing Y = g(X); that is same as the $P\{X \le g^{-1}(y)\}$. This is valid only g is a strictly monotonic function. In this case it is a strictly increasing function. Once I know the CDF, I can find the probability density function by differentiating both side with respect to y. By differentiating CDF with respect to y, I can get the probability density function of Y; that is f_Y. That is nothing but once you do the differentiation in right hand side with respect to y, you can use the chain rule.

Therefore, it is the probability density function evaluated at the point $g^{-1}(y)$; then differentiate the $g^{-1}(y)$ with respect to y. So, this is valid when g is strictly increasing function. Suppose, g is decreasing function, then also you can do the similar calculation. Then you will get the values with the negative sign, then you can take the negative-negative positive.

So, you can go for absolute of this in general whether it is a strictly increasing or strictly decreasing, you can take absolute of this derivative terms and substitution the probability density function of X with the g inverse of y will give the probability density function of Y whenever Y = g(X), otherwise it is 0. You can go for example for how to apply this theorem.

(Refer Slide Time: 25:51)



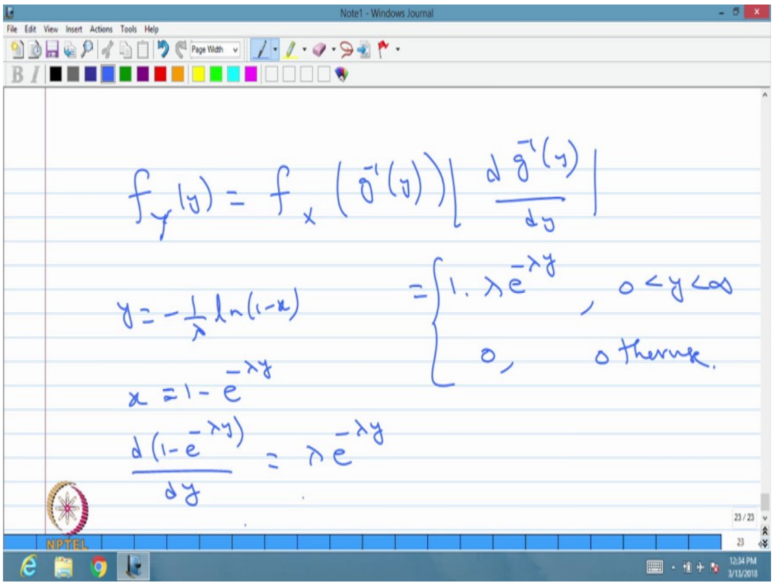
So, easiest example, example number three because we have already discussed two examples.

The third example, let X be a continuous type random variable with the probability density function f(x) takes a value 1 between 0 to 1 otherwise it is 0; that means, it is a constant probability density between the interval 0 to 1, otherwise its 0. If you integrate you will get the value 1, then greater or equal to 0; therefore, this a probability density function.

Now, we are defining a new random variable Y $= -\frac{1}{\lambda}\ln(1-X)$. I am defining new random variable Y ¿$-\frac{1}{\lambda}\ln\left(1-X\right)$. You can prove it is a continuous function; therefore, it is a Borel-measurable function; therefore, it is a random variable.

Here, the lambda has to be strictly greater than 0. So, X is a continuous type random variable with the probability density function 1 between the interval 0 to 1; 0 otherwise and Y you define it as$\frac{-1}{\lambda}\ln\left(1-X\right)$. Now you can verify whether this theorem can be applied. X is a continuous type random variable, Y is this function; it is a strictly increasing function therefore, and differentiable also. So, therefore, we can apply the theorem and you can directly get the probability density function.

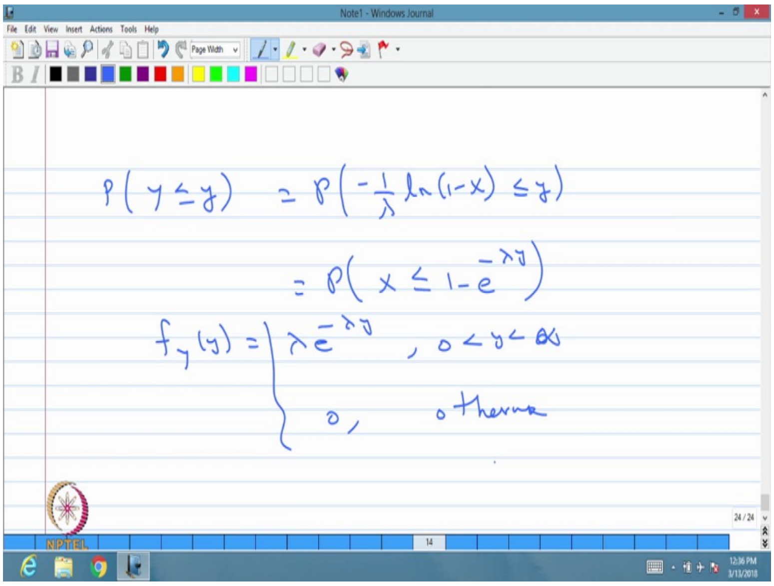
(Refer Slide Time: 27:57)



So, the probability density function of Y that is a probability density function of X by replacing x by $g^{-1}(y)$. Then $\frac{d\,g^{-1}\left(y\right)}{dy}$ absolute. So, we have $y=\frac{-1}{\lambda}\ln\left(1-x\right)$. So, you can find x, you can find x. So, the x is going to be $1-e^{-\lambda y}$; that is a $g^{-1}(y)$. So, you can find the derivative of $1-e^{-\lambda y}$ with respect to y, you will get that is $\lambda e^{-\lambda y}$.So, you need g inverse of y as well as you need derivative. Therefore, this is going to be the probability density function substituted x by $g^{-1}(y)$, but since the probability density function is a constant between the interval 0 to 1. Therefore, it is going to be again 1 and the derivative with absolute that is $\lambda e^{-\lambda y}$ and if you do the little home work when x lies between 0 to 1, you will get y lies between 0 to ∞. Therefore, the probability density function is between 0 to ∞; one times so one can be avoided. So, 0 otherwise.

So, whenever x takes the value 0 to 1, y takes the value 0 to ∞ in that the probability density function is $\lambda e^{-\lambda y}$, otherwise it is 0. So, since this theorem is satisfied, we are able to get the probability density function directly. You can find the CDF of the random variable Y directly without using the theorem also. By using the theorem, we got the probability density function.

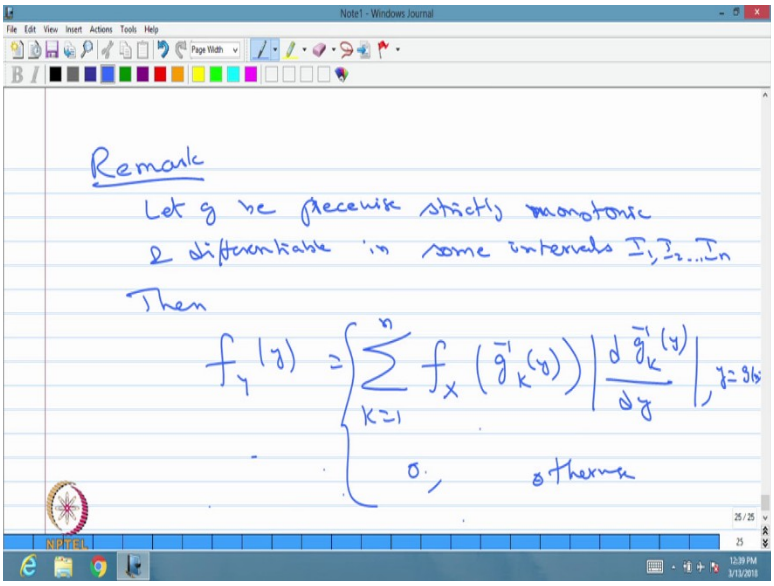
(Refer Slide Time: 30:29)



But, the another method you can directly compute what is the CDF of the random variable Y; that is same as P{ $\frac{-1}{\lambda}\ln(1-X) \le$ y}. Then you can do the little simplification, you will get P{X ≤ 1 - $e^{-\lambda y}$}. If you do the little simplification, you will get a P{X ≤ 1 - $e^{-\lambda y}$}. Therefore, I can go for probability density function by differentiating both side that is going to be $\lambda e^{-\lambda y}$ when Y is lies between 0 to ∞. I am skipping in between some steps that you can work out separately. So, there are two ways; either you can apply the theorem or you can find the CDF first by differentiating you can get the density function.

Sometimes, it may be not strictly increasing or strictly decreasing; in some interval the function may be monotonically increasing, some interval it may be monotonically decreasing, but still you can use the theorem.

(Refer Slide Time: 31:54)



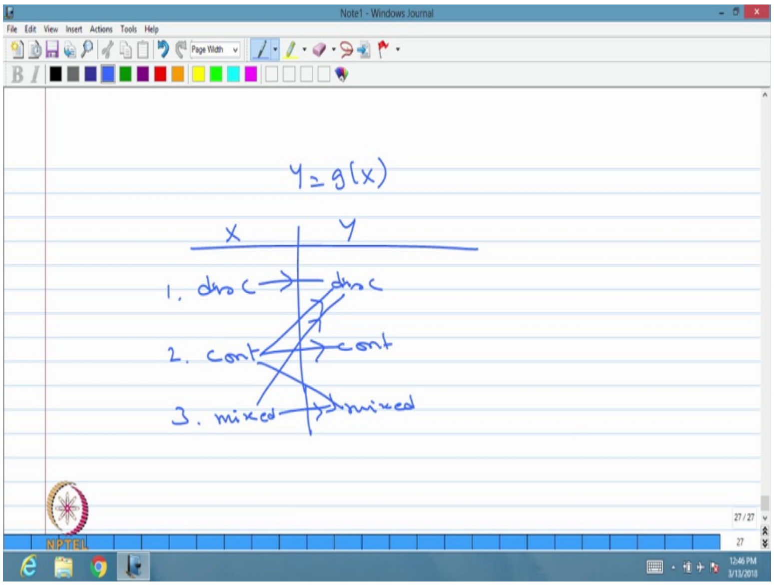
So, I am coming back to the remark of the theorem. Whenever you have an issue in which the function g be a piecewise strictly monotone and differentiable in some intervals. Earlier in the whole interval, it is strictly increasing or strictly decreasing. Now it is piecewise strictly monotonic and differentiable in some intervals. We label these intervals as the I_1, I_2 and so on, suppose I_n intervals.

Then, we can find the probability density function of the random variable Y. Then, we can find the probability density function of Y as sum of n intervals in which you get what is the sum of f_X at different k g^{-1}; then find out the absolute of derivative of $g^{-1}(y)$, the kth one with respect to y. So, this is going to be the probability density function which is greater than 0 whenever you get the points which is y = g(x), otherwise it is 0.

So that means, there is a possibility you may have a function in which it is increasing or decreasing or decreasing or increasing or it may be in many intervals. Then, you can add all the intervals corresponding density function, if you sum it up; then that is going to be the probability density function of Y.

So, we have seen many examples in which discrete to discrete, continuous to continuous or continuous to discrete and so on. So, the way we create the random variable Y accordingly you will end up with the different types of the random variable for Y for a given random variable X.

(Refer Slide Time: 35:05)



So, in conclusion whenever X is given and you are interested to find out the distribution of Y as a function of X in the g, in conclusion we can go for the table form when X is of discrete type. I am just writing disc, discrete type or X is continuous type random variable or X is of the mixed type random variable; the way we defined the Y = g(X), the random variable Y that is also of the form discrete or continuous or mixed. You can think of many functions in your mind, then you can make a line whether you will get a discrete to discrete or discrete to continuous or mixed and so on. So, in 3 to 3, out of 9 which possibility is possible that you can make out.

So, discrete to discrete is possible, I have given the example also. Y ¿ X^2 example where X is a discrete; therefore, Y is also discrete. From discrete you can go for from the discrete type of random variable, you cannot go for the continuous, you cannot go for the mixed type random variable. Whereas, if you have a continuous type random variable by default you can always make a function Y is equal to X or something.

Therefore, from continuous to continuous, it is easy. From continuous to discrete is possible, I have given 1 example, X is a continuous type random variable, Y is equal to 1 and - 1. So, that Y is equal to 1 and - 1 that example is the continuous to discrete type example.

Since, continuous to continuous and continuous to discrete is possible; therefore, continuous to mixed is also possible because the continuous means the unit mass is distributed over the interval, over the real line. So, you can make a many to one function; therefore, some density can be a mass at some points and whereas, the few density you can keep density as it is. Therefore, continuous to mixed is possible.

I will repeat, from continuous type random variable mixed type random variable is possible because the unit mass can be transformed into density in some interval and masses at some points. Therefore, it is a mixed type random variable because continuous to continuous is possible, continuous to discrete possible; therefore, continuous to mixed is also possible

Now, we come to the third type when X is a mixed type random variable mixed means it has the density as well as mass. So, by making a many to one function, all the density you can put it mass at some points. Therefore, mixed to discrete is possible. Very important observation mixed to continuous is impossible, mixed to continuous type random variable is impossible because mixed has mass at some points and density between some interval. So, mixed to continuous is not possible. Whereas, by default by 1 to 1 mapping, you can always have mixed to mixed.

Therefore, whatever you do the different problems in the distribution of function of random variable, at the end you will end up with, these are all the only possible ways you will get the different type of random variable for Y and once you know the distribution of X, you can find the distribution of Y in the form of a CDF or if you know that it is a discrete type random variable in the form of probability mass function, if you know Y is continuous type random variable in the form of probability density function.

With this, we are completing the random variable with the three topics; one is definition and the CDF and the second one is types of random variable and the third topic is distribution of function of random variable.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module – 03
Moments and Inequalities
Lecture – 12

So, in this model, we will discuss the Moments and the Inequalities. So, we planned to give three lectures in this model. So, the first lecture we will discuss the first two moments mean and variance and in the second lecture, we discuss the high order moments and moment inequalities and in the third lecture, we will discuss 3 important generating functions; that is a probability generating function, moment generating function and characteristic function.

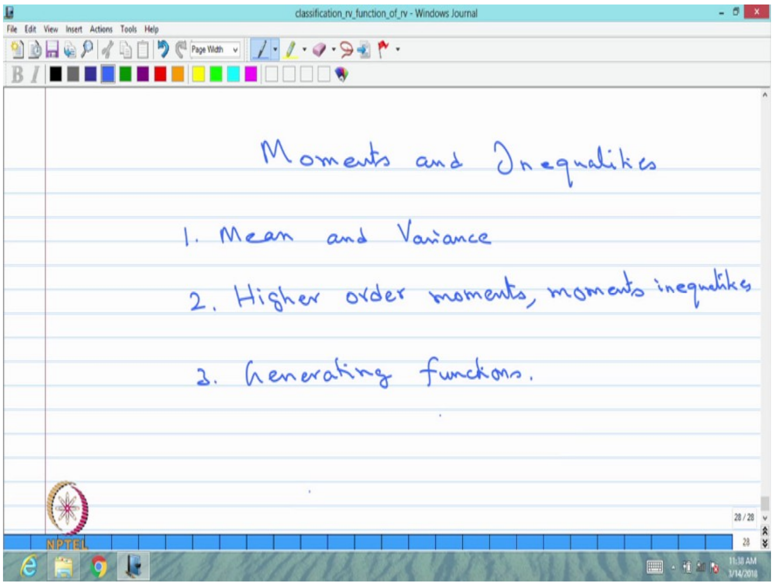
Moments and inequalities; it's a very important topic in the probability or in particular random variable because most of the time finding the distribution of the random variable is a tedious task or it is a very difficult to find out the distribution of the random variable. But it is easy to compute or easy to find the value of the mean or expectation that is basically the first order moment. Similarly finding the variance that is of the second order moment that is also easy comparing to the finding the distribution of the random variable.

Therefore, in the probability theory course we discuss the distribution first, then we discuss the moments; then, later at the end of the course how much the moments are going to play an important role, when we are going to use central limit theorem for approximating some of the random variable into normal distribution with some conditions. So, in that place, the moments are playing an important role and also the moments are the easy measures; it is easy to compute in the real world problems not the distribution.

Once you know the distribution, if the moments exist, we are going to tell the existence of the moments. If the moments exist, we can always find it. If you know the moments of many orders; then you can able to identify what could be the distribution of that random variable. That means, for a random variable if the moment exists, we can able to find.

Through the moments of some unknown random variable; it is possible to fit the corrected distribution for that moments matches.

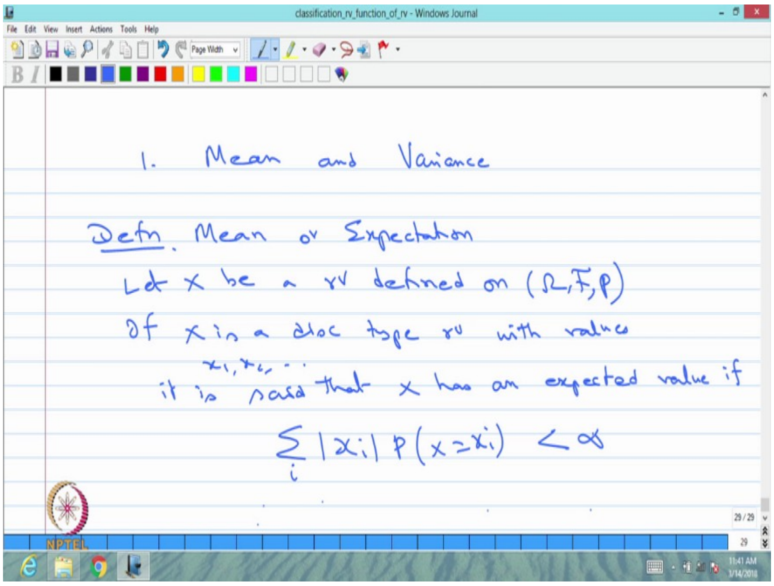
(Refer Slide Time: 03:09)



So, in that way I am going to introduce in this module that is Moments and Inequalities. In this lecture, we start teaching first Mean and Variance. Mean is the first order moment and the Variance is the second order moment. Then, we discuss higher order moments and we also discuss moment inequalities and the third, we discuss generating functions.

So, in this model we are going to have three lectures. First one, mean and variance; second, higher order moments and moment inequalities; third, generating functions. Now, we will move into the first topic that is called Mean and Variance. This is a very important measure. So, let me start with a definition of mean.

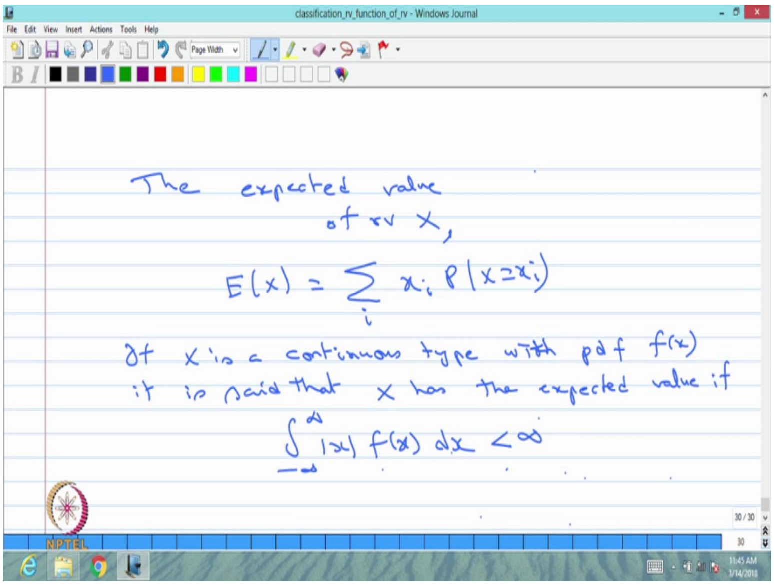
(Refer Slide Time: 04:27)



There is a. another name for mean that is called Expectation. In many times we find the value of average of some n values that is also going to be sort of mean and expectation with some conditions of that random variable. So, that we are going to discuss little later. So, as it is now, I am going to write mean or expectation.

Let, X be a random variable; rv means random variable defined on the probability space (Ω, F, P). If X is a discrete type random variable with values x_1, x_2 and so on. It is said that X has an expected value, if the $\sum_i |x_i| P(X = x_i)$ if the summation is finite quantity.

(Refer Slide Time: 06:52)



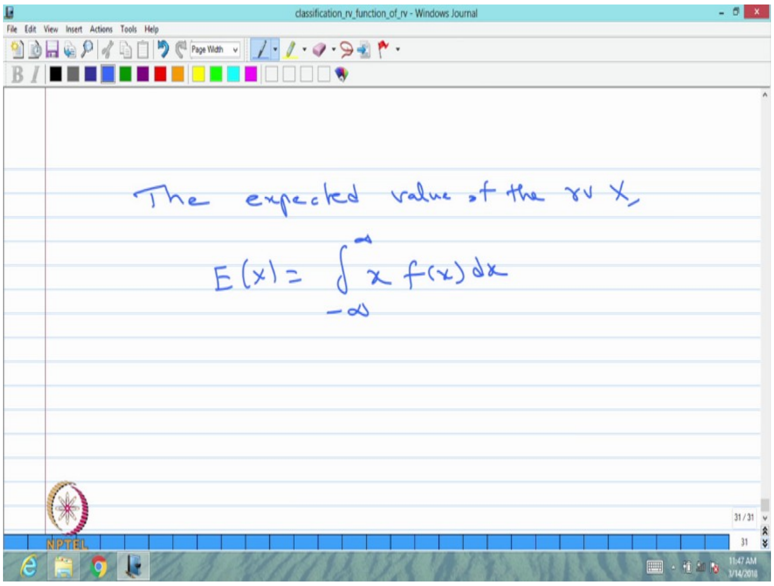
The expected value of the random variable X denoted by E(X), is defined as E(X) that is same as the $\sum_i x_i P(X=x_i)$. What we are saying is whenever you have a random variable defined on the probability space (Ω, F, P), if it is a discrete type random variable whose values are x_i is x_1, x_2 and so on. It is said that X has expected value if the $\sum_i ¿ x_i \vee P(X=x_i) ¿$s a finite value. Why the absolute is there? There is a possibility the x_i's could be positive or negative.

So, in the absolute sense, if it is going to be finite; without the absolute that is going to be the expected value of the random variable provided in absolute sense that summation is a finite quantity. In other words, this is nothing but the series you can think of $\sum_i x_i P(X=x_i)$ as the A_i's, whatever you have studied in the calculus course, this series converges in absolute sense.

Then, without absolute sense the similar series, you made it $\sum_i x_i P(X=x_i)$ that is going to be the expected value. If the series is convergence in absolute sense, then it is going to be converges here. So, we are using the calculus sequence convergence concept to conclude, expected value of a discrete random variable exist.

So, now I am going for if the random variable X is continuous type. If X is a continuous type random variable with the probability density function is F(x), it is said that X has the expected value if the integration $\int_{-\infty}^{\infty} |x| f(x)\, dx$is finite quantity in this case.

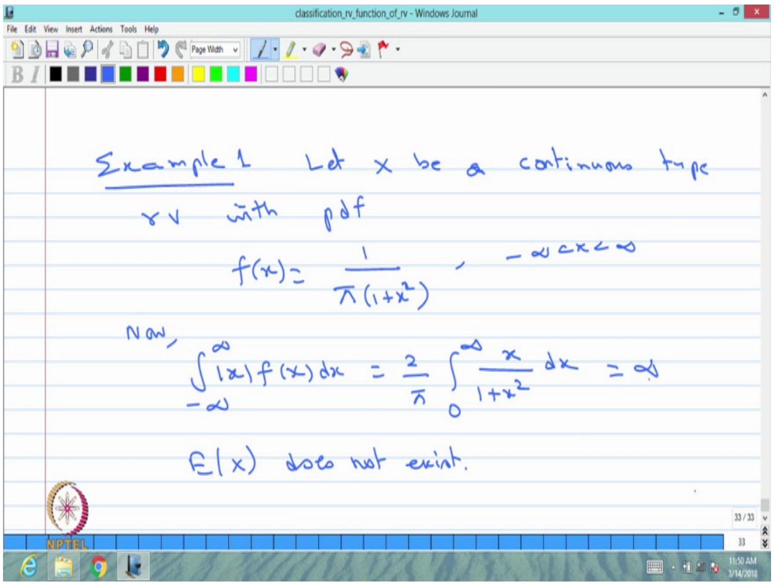
(Refer Slide Time: 10:18)



In this case, the expected value of the random variable X denoted by E(X) = $\int_{-\infty}^{\infty} x f(x)\, dx$; note that $\int_{-\infty}^{\infty} f(x)\, dx =$¿¿ 1. Because f(x) is a probability density function; whereas, here if the $\int_{-\infty}^{\infty} |x| f(x)\, dx$is a finite quantity, then we can say that the expected value of the random variable X exist and that value is same as E(X) is a notation,n that is same as $\int_{-\infty}^{\infty} x f(x)\, dx$.

So, the provided condition is very important, if the provided condition is not satisfied even you get the value of this integration without absolute that is not going to be call it as an expected value. It is very important to conclude in the absolute sense, the summation for a discrete type random variable, integration in absolute sense for the continuous type random variable, if it is finite quantity without absolute sense that is going to be the expected value of the random variable. I am going to give one example for random variable in which the expectation does not exist and one or two examples for expectation exist.

(Refer Slide Time: 12:04)

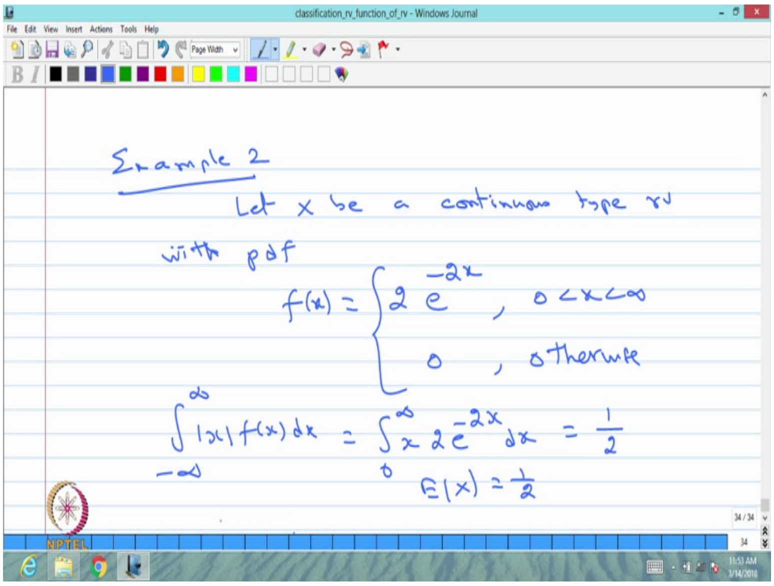


Example 1, let X be a continuous type random variable with probability density function f(x) that takes a value $\frac{1}{\pi(1+x^2)}$, where x lies between - ∞ to + ∞. The probability density function is $\frac{1}{\pi(1+x^2)}$, where x lies between - ∞ to + ∞. You can verify whether this is a correct probability density function by integrating f(x) from - ∞ to + ∞, we will get the value 1. This is greater or equal to 0.

Therefore, it is a probability density function of continuous type random variable. Now you can compute the $\int_{-\infty}^{\infty} |x| f(x)\, dx$. If you compute this quantity that is same as $\frac{2}{\pi}$, it is an even function. So, $\frac{2}{\pi}\int_{0}^{\infty} \frac{x}{(1+x^2)} dx$. If you do the simple calculation, you will come to the conclusion this quantity is going to be ∞. You can make it as a homework computing this integration that is going to be ∞.

So, since this quantity is going to be ∞, you can conclude the expectation of X does not exist for this random variable. Even though, you will get the value from $\int_{-\infty}^{\infty} x f(x) dx$ for this problem, but since in absolute sense this integration value is infinite. Therefore, the mean does not exist.

(Refer Slide Time: 14:38)



I will take a next example, where, expectation exist. Let, X be a continuous type random variable with probability density function that is f(x) = $2e^{-2x}$, when x is lies between 0 to ∞, otherwise it is 0. You can verify this a probability density function or not. If you integrate you will get the value 1 and between the interval 0 to ∞ because from - ∞ to 0, it is 0 and it is a positive function. Therefore, it is going to be a probability density function.

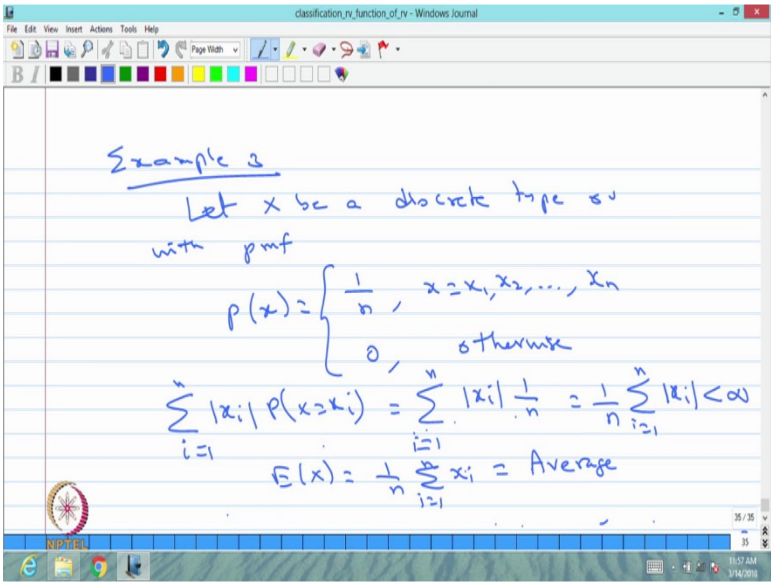
Since, the possible values of x are from 0 to ∞, we do not need to check whether this is going to be a finite quantity in absolute sense you can directly compute; if that is would be a finite quantity, then that is same as the expectation. Let us compute $\int_{-\infty}^{\infty} |x| f(x) dx$that is same as since f(x) is going to be 0 between - ∞ to ∞. Therefore, this is same as $\int_{0}^{\infty} f(x) dx$ is $\int_{0}^{\infty} 2e^{-2x} dx$; that is same as if you do the simple calculation you will come to the answer that is 1/2.

So, since this is going to be a finite quantity. Therefore, we can conclude expectation of X that is going to be 1/2 because the f(x) is going to be greater than 0 between the 0 to ∞. Therefore, in the absolute sense answer is finite that is same as without absolute also.

Therefore, the E(X) = $\int_0^{\infty} x f(x)\, dx$ that is going to be 1/2. So, in this case the expectation exists for this continuous type random variable.

Later, we are going to see this random variable is exponential distributed with the parameter 2; that we are going to discuss later. Therefore, the expectation of exponential distributed random variable is reciprocal of the parameter. So, here the parameter is 2 that is 1/2; it is an expectation for the random variable. We can go for one more example.

(Refer Slide Time: 17:42)



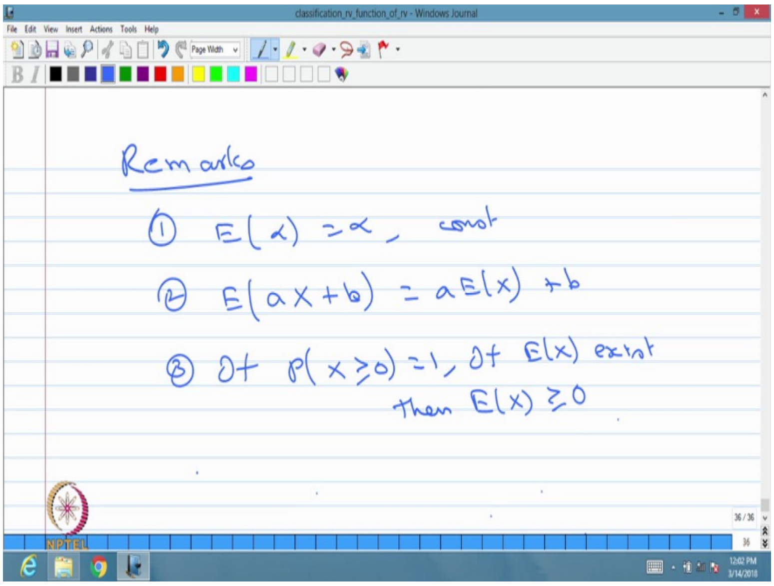
Let, X be a discrete type of random variable with probability mass function is given by p(x) that is 1/n, when x takes a value x_1, x_2 and x_n; that is going to be 0 otherwise. So, this is a probability mass function of a discrete type random variable, with the possible values are x_1, x_2 and so on, x_n and this could be positive values or negative values or it could be 0 also.

Now, we will find out whether expectation exist for this one? So, if you compute $\sum_i ¿ x_i \vee P(X=x_i)$. So, here it is a summation is i is equal to 1 to n. This is same as absolute of x_i's. P{X = x_i} = 1/n; i is equal to 1 to n. This is nothing but $\frac{1}{n}\sum_{i=1}^{n} x_i$. Since, we have taken all the possible values from - ∞ to ∞, $\sum_{i=1}^{n} x_i$ that is also going to be finite quantity.

Therefore, the E(X) = $\frac{1}{n}\sum_{i=1}^{n} x_i$. This is going to be the expected value of the random variable X, when X is a discrete type random variable with the probability mass function 1/n for x_i's are n values. But you know that $\frac{1}{n}\sum_{i=1}^{n} x_i$that is nothing but average. So, whenever the random variable has is equiprobable at finite number of points in that case the mean or expectation that is same as average.

Later, we are going to conclude this random variable is called as discrete type uniform distribution. That means, for a discrete type uniform distribution the mean or expectation that is same as average; that is what many times when you have n values, you make a average out of it. So, the average is same as considering that as a random variable whose probabilities same as 1 divided by the total number of information or total number of values in that case a mean or expectation that is same as the average.

(Refer Slide Time: 21:39)



Now, we will discuss few important results for the mean as remarks. The first remark the expectation of a constant is a constant. If you find the expectation or mean for a constant that is going to be a constant; this result can be proved easily. You can treat a constant as a random variable whose probability is 1 taking only that value. Therefore, expectation of constant is nothing but constant times probability of X takes a value constant that is 1. So, 1 into α; therefore, it is α. It can be proved easily.

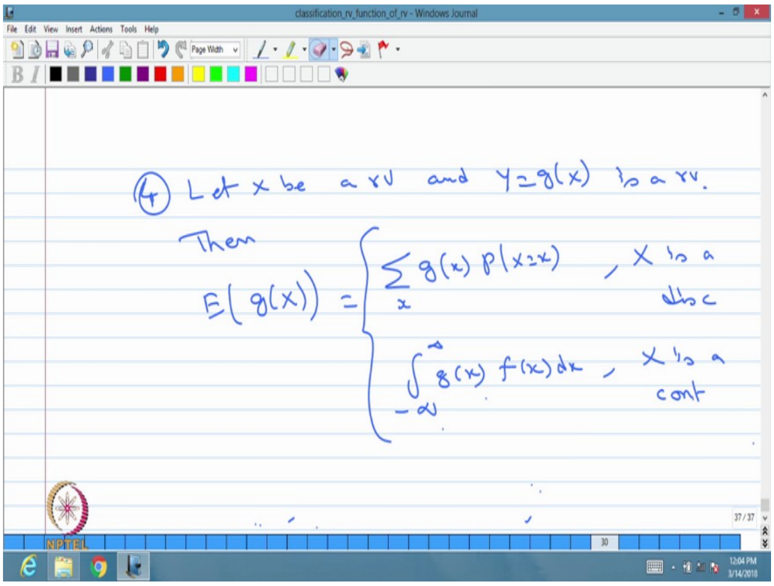
As a second remark, suppose you have some constant times random variable + another constant. Here a and b are constants; in that case just now we made it expectation of a constant is a constant. But here, it is a linear combination and the constant can be taken out with the expectation of X + expectation of a constant that is same as constant.

That means, expectation of constant times random variable + another constant is same as constant can be taken out. Third remark, if $P\{X \geq 0\} = 1$. The unit mass is distributed between the interval 0 to ∞. In that case, if expectation exist when the above two results also sorry above the second result also if the expectation exist then this is too. So, if E(X) exist, then the $E(X) \geq 0$.

This intuitively also you can say. If the $P\{X \geq 0\} = 1$ and if expectation X exist, then the expected expectation or mean value of that random variable is also going to be greater or equal to 0. This also can be proved either from the definition the same way. Since whether you are going to use X is a discrete type or continuous type or mixed type, you can use the similar same definition for the existence in the absolute sense, then you can conclude if it is exist, then that value is going to be greater or equal to 0.

When I give the definition, I have given the definition for the expectation for a discrete type and the continuous type; I have not given the definition for mixed type. But the mixed type is the combination of a discrete and continuous. Therefore, wherever there is a mass you can do the summation; wherever there is a density you can do the integration. Therefore, the expectation formula is a combination of both discrete and continuous; therefore, I have not given the definition.

(Refer Slide Time: 25:17)



The fourth remark. Remark number 4, let X be a random variable and Y = g(X) is a random variable. This we discussed in the last model, X is a random variable; g is a Borel-measurable function in particular area, you can think of continuous or piecewise continuous function.

Therefore, Y is also going to be a random variable. Then, you can find the E(g(X)) = $\sum_x g(x) P(X=x)$.

If the random variable X is a discrete type. If X is a continuous type random variable - $\int_{-\infty}^{\infty} g(x) f(x) dx$ when X is a continuous type random variable. This can be proved provided the expectation exist, provided the expectation exist you can find expectation value of g(x) with a summation if it is a discrete type with the integration if it is a continuous type.

Why this result is very important? Is you are not finding the distribution of g(x), but with the help of the distribution of X you are finding the expectation of g(X), provided it exist. Therefore, this is a very important result in the sense you do not need to find the distribution of g(X) whatever be the distribution of g(X) as long as if it exists.

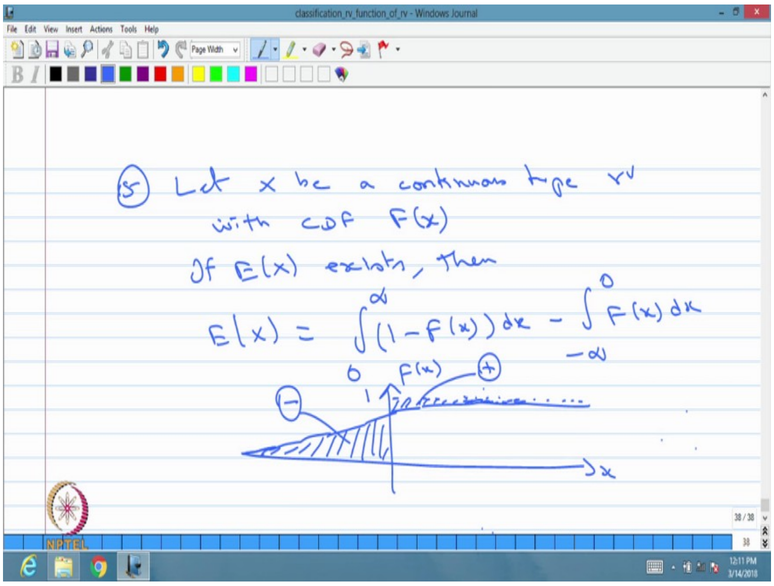
You can find the value of E(g(X)) with the help of the distribution of X means if it is a discrete type random variable, if you know the probability mass function; if it is a continuous type random variable the probability density function of X. You can find the expectation by the summation or the integration without knowing the distribution of g(X).

So, therefore, we will be using this property again and again whenever you want to find out the expectation of a function of random variable; not only this definition similarly later we are going to introduce many random variables or random vector of size n with n random variables. There also you can go for expectation of a functions of random variables in the similar way. Therefore, this is a very important result one can find it.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 13

(Refer Slide Time: 00:02)



The next remark, all these remarks can be proved, but I am just giving as a remark it is for the understanding. The next remark, let X be a continuous type random variable, let X be a continuous type random variable with CDF F(x).

If the E(X) exist that means, the provided condition is satisfied. Then you can always find the value of E(X), that is E(X) = $\int_{0}^{\infty}(1-F(x))dx-\int_{-\infty}^{0}F(x)dx$. This a very important result, whenever you have a continuous type random variable with CDF, F(x) and the expectation exist.

Then you can always find the value with the help of CDF you do not need the probability density function of that continuous type random variable, from the definition of a expectation when X is a continuous type random variable if you know the probability density function if the expectation exist then E(X) = $\int_{-\infty}^{\infty}xf(x)dx$, where that f(x) that is the probability density function. But now what I am saying is you do not need a probability density function with the help of CDF itself you can able to find the expectation provide that it exists.

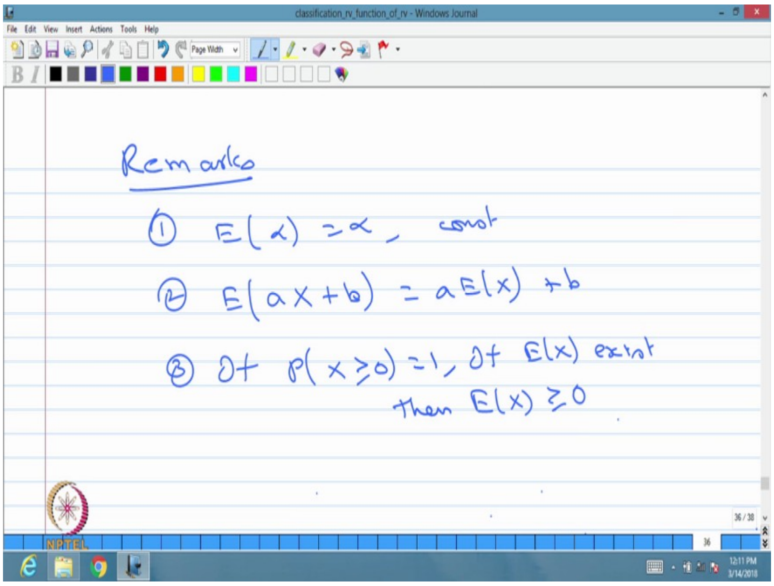
Let me give a one nice pictorial representation of how to compute the expectation for this, when X is a continuous type random variable. Recall, when X is a continuous type ransom variable the CDF is the continuous function in the whole real line from - ∞ to ∞. When X is a continuous type random variable, the CDF of the continuous type random variable is a continuous function in the whole real line - ∞ to + ∞. Therefore, I am just making a one CDF it asymptotically touches one at ∞.

So, this is the CDF of some continuous type random variable. $\int_0^\infty (1-F(x))dx$that is same as an area below 1- F(x) between the interval 0 to ∞. So, you can shade. So, this quantity the shaded quantity is nothing but the $\int_0^\infty (1-F(x))dx$, 1 it is a line F(x), 1 – F(x) 0 to ∞ that is area be between the F(x) with the line 1 that shaded area is a $\int_0^\infty (1-F(x))dx$.

Next $\int_{-\infty}^0 F(x)dx$, that means this part. So, when you whenever the expectation exists for a continuous type random variable this sign is with the plus sign, this sign is with the negative sign. So, that area plus sign area minus that minus sign area that value is going to be the expectation value. Therefore, the expectation can be negative or positive based on the area between - ∞ to 0 that is going to be more than or less than the area between $\int_0^\infty (1-F(x))dx$.

Suppose X takes, X is a non negative random variable that is a P{X ≥ 0} = 1, using the first remark, yeah third remark, sorry.

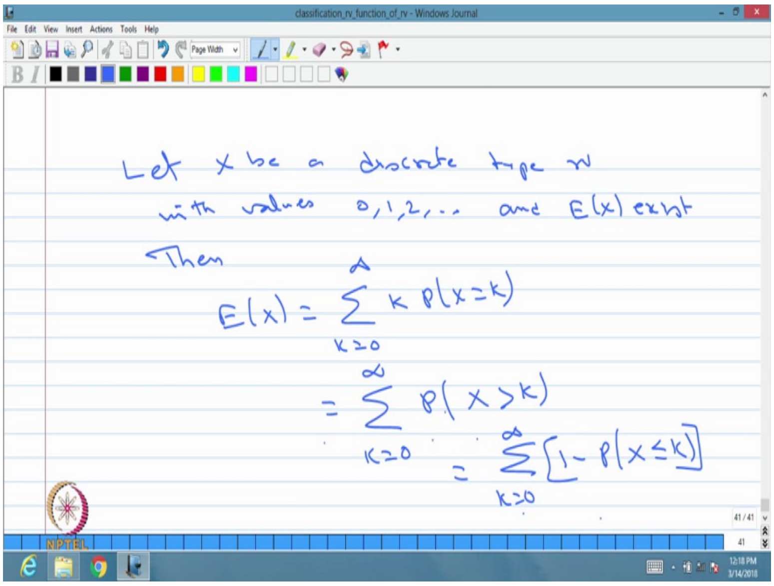
(Refer Slide Time: 05:40)



If P{X ≥ 0} = 1 and if expectation exist then the E(X) is going to be greater than or equal to 0 that you can visualize, $\int_{-\infty}^{0} F(x)\,dx$ that quantity is going to be 0 because P{X ≥ 0} = 1. Therefore, you will get the positive quantity from the first integration and the second integration values is 0. So, this can be visualized.

There is an another remark over this remark, I started with the continuous type random variable you can think of a discrete type random variable also. Suppose it takes only the positive values, then one can make, if the expectation exists, then the expectation is same as in the summation form.

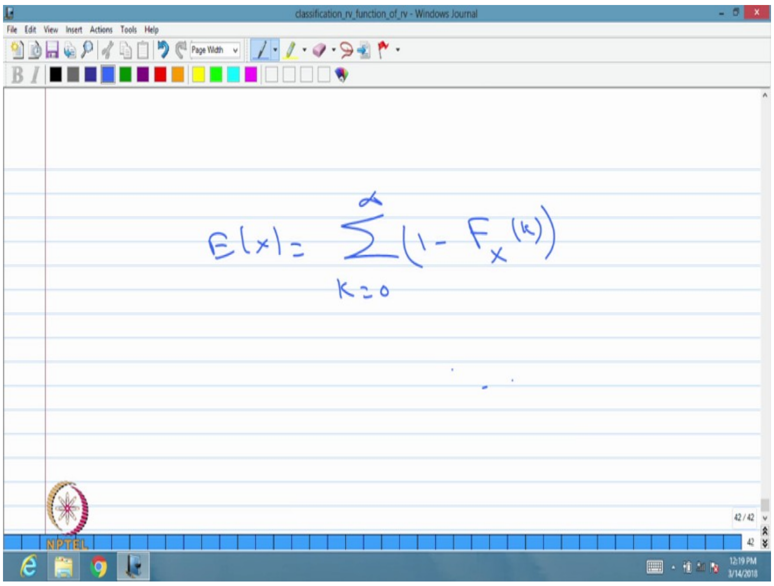
(Refer Slide Time: 06:50)



Let X be a discrete type random variable with values 0, 1, 2 and so on, with taking values as 0, 1, 2 and so on and E(X) exist then we can find the expectation of X. By the definition it is $\sum_{k=0}^{\infty} kP(X=k)$.By using the previous remark, we can conclude that is same as $\sum_{k=0}^{\infty} P(X>k)$.

By using the previous result, v that is same as $\sum_{k=0}^{\infty}(1-P(X \le k))$. I want to use similar logic of 1 – F(x). So, this is $P(X \le k)$is nothing but a F of random variable X at the point k.

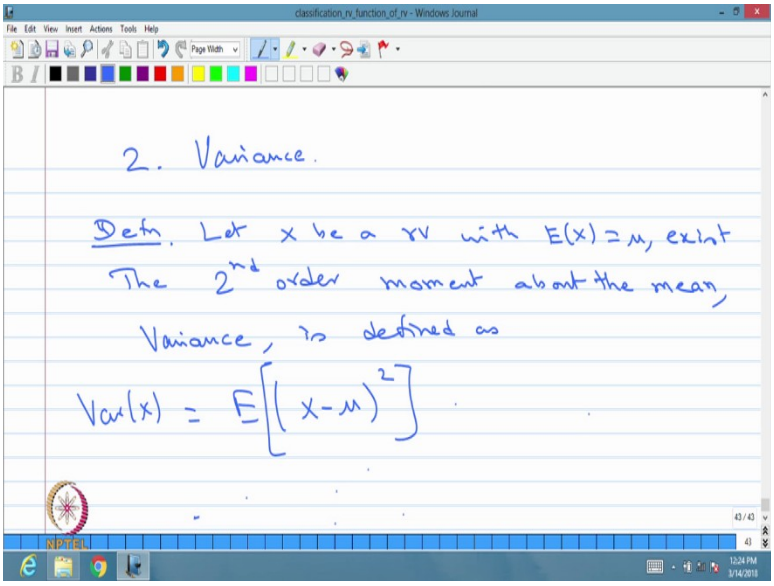
(Refer Slide Time: 08:32)



So, this is same as E(X) = $\sum_{k=0}^{\infty}(1-F_X(k))$.

So, if you see the previous remarks, when X takes the only the non negative values then the second integration vanishes therefore, you will get the first one. So, first I give the remark, so for the continuous type random variable, for a discrete type random variable this is going to be expectation of X provided X takes the value 0, 1, 2 and so on. So, this is also very important remarks for the expectation.

(Refer Slide Time: 09:23)



Now, we will move into the second moment that is called variance. Let me give the definition. Let X be a random variable with expectation of X that you call it as a μ which exist. I am going to give the notation for E(X) = μ which exist.

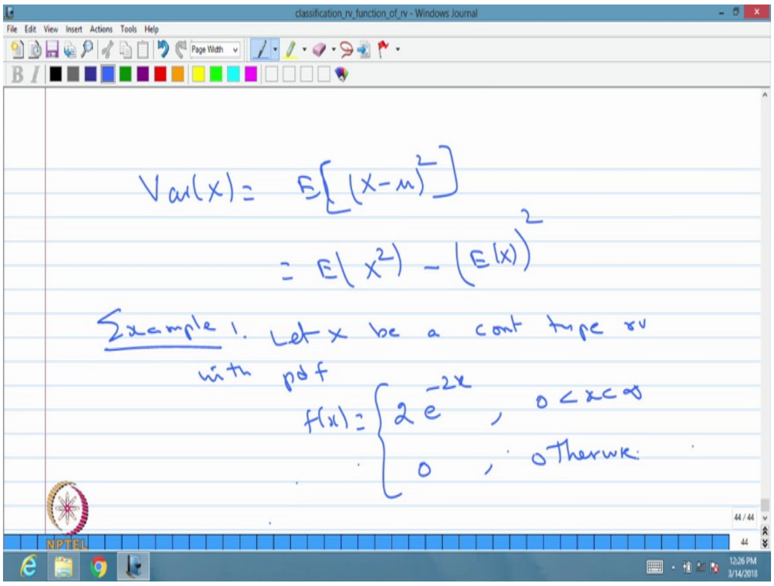
The second order moment about the mean, that is nothing but variance, is defined as $E[(X-\mu)^2]$ that is denoted by variance of X. So, finding the value of $E[(X-\mu)^2]$ that is going to be call it as variance that is called a second order moment about the mean.

Since we are writing the right hand side, $E[(X-\mu)^2]$ that means, provided the right hand side exist; that means, you can treat $(X-\mu)^2$ where μ is expectation of X. So, $(X-\mu)^2$ you can treat it as the function of X the g(X). So, right hand side is nothing, but E[g(X)], where g(X) = $(X-\mu)^2$. You can use the previous result provided the E[g(X)] exist then the expected value that E[g(X)], where g(X) = $(X-\mu)^2$ that is going call it as a variance of X.

So, this second order moment one can define after the existence of the first order moment. If the first order moment does not exist then one cannot define the second order moment about the mean. So, here I have given the second order moment about the mean, later I am going to introduce a second order moment about the origin, that is nothing but the $E(X^2)$ by making μ = 0 or $E(X^2)$ will be the second order moment about the origin. Similarly one can define nth order moment about the origin.

So, now we will discuss the variance first then later we will go for the higher order moments. So, the variance of X that is denoted by var of X that is $E[(X-\mu)^2]$. So, as I said earlier you do not need to find out the distribution of $(X-\mu)^2$ as long as you know the distribution of X and the E(X) exist, and $E[(X-\mu)^2]$ exist then the value is going to be $E[(X-\mu)^2]$ is same as variance.

(Refer Slide Time: 13:13)

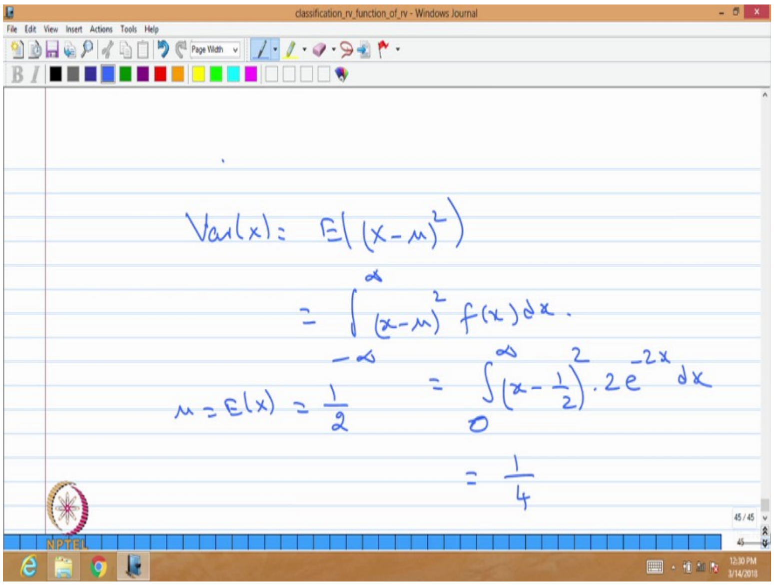


You can rewrite the variance of X as first you have $(X-\mu)^2$, you can expand $(X-\mu)^2$. therefore, after expansion you will get $E(X^2)-E(X)^2$ the E(X) = μ. So, whether you compute $E[(X-\mu)^2]$ or you find out the $E(X)^2$ first that is μ^2, then find out the second order moment about the origin that is $E(X^2)$ then this difference is going to be the variance of X.

Similar to the expectation exist or not same way for some random variable the variance may not exist even though the first order moment exists. If the first order moment does not exist you cannot define the second order moment. So, even the first order moment exists there are some random variable in which the second order moment does not exist. Therefore, the variance does not exist.

So, we can go for simple examples the first example. Let X be a continuous type random variable with the probability density function f(x) the same example $2e^{-2x}$, where x is lies between 0 to ∞, 0 otherwise.

(Refer Slide Time: 15:08)

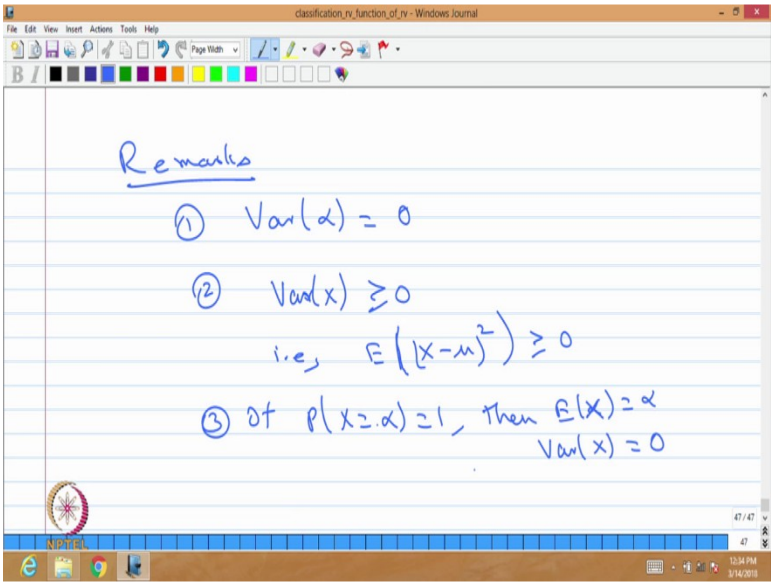


So, how to compute variance of X for this problem is variance of X is same as E [$(X-\mu)$¿¿2]¿, that is same as $\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$. With this problem, the μ is same as expectation of X which we got it already that is 1/2. Therefore, this is going to be - ∞ to 0 the f(x) is 0, therefore, you can directly go for $\int_0^{\infty} \left(x-\frac{1}{2}\right)^2 2e^{-2x} dx$.

We are not finding the distribution of $(x-\mu)^2$, then we are finding the expectation, no we are using an expectation of g(X), which I said it in the remark. So, $\left(X-\frac{1}{2}\right)^2$ probability density function which is greater than 0 between 0 to ∞ therefore, - ∞ to 0 will vanishes. So, this is a must 0 to ∞ of this, if this quantity is going to be find it then the existence also taken care, since X takes a non-negative values you are finding this.

If you do the simplification you can get the value that is 1/4, if you do the simplification of this integration you will get the value 1/4. Therefore, for the random variable which is continuous type whose probability density function is $2e^{-2x}$, when x lies between 0 to ∞ the mean is going to be 1/2 and the variance is 1/4. Later we are going to conclude this is going to be exponential distribution with the parameter 2. So, the mean is 1 divided by the parameter and the variance is a 1 divided by the square of the parameter. So, here the parameter is 2 therefore, it is 1/4.

(Refer Slide Time: 17:42)



Let us go for a few remarks over the variance. The first remark variance of a constant is 0 whereas an expectation of a constant is a constant. Variance means what is a variation about the mean. So, here the mean of a constant is a constant. Therefore, there is no variation about the mean that is α therefore, the variance is 0. So, intuitively you can see the variance of constant is 0.

The second remark, variance of X if it exist then that value is always going to be greater than or equal to 0, if the variance exist for a random variable X then the variance is always greater than or equal to 0. That is $E((X-\mu)^2) \geq 0$.

This you can say by using the remarks on the expectation you go back when the random variable X whose probability is P{X ≥ 0}=1, then the expectation is going to be greater than or equal to 0 with that logic, $E\left(\left(X-\mu\right)^2\right)$ whose probability is always $\mathrm{P}\left(\left(X-\mu\right)^2 \geq 0\right)=1$ because it is a non negative random variable. Therefore, the $\mathrm{E}\left(\left(X-\mu\right)^2\right) \geq 0$.

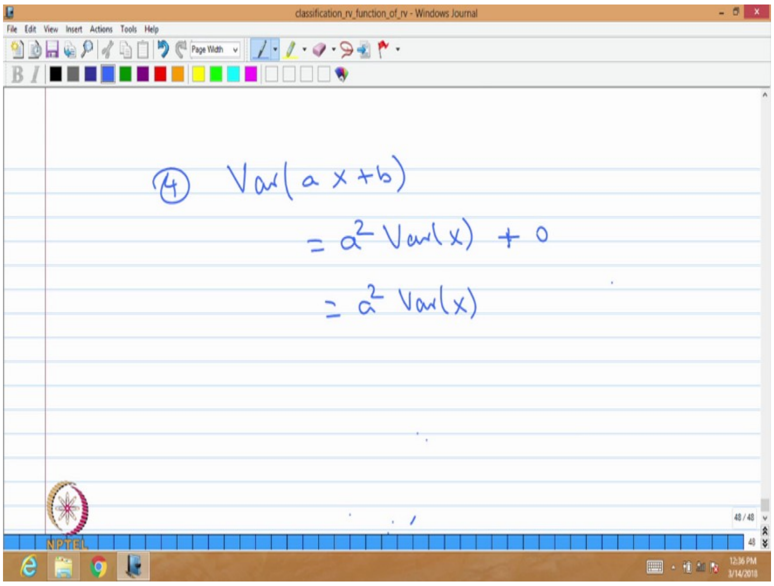
Because, whether it is a discrete random variable or continuous type random variable this is nothing, but if it is a discrete it is a $\sum \left(x_i-\mu\right)^2 P\left(X=x_i\right)$and the probability mass function is greater than or equal to 0, x_i - μ whole square is a positive quantity therefore, if it exist then the summation is going to be positive.

If X is a continuous type random variable then $\left(X-\mu\right)^2$ probability density function integration from - ∞ to ∞ again integrant is greater than or equal to 0 therefore, the if the expectation exist sorry if the variance exist then these quantities also in going to be greater or equal to 0.

You can combine the remark number 1 and 2 in the form of a third remark if P{X = α} = 1 that means, it is a constant. P{X = α} where α is a sum number that is 1. Then the mean is same as mean of the random variable same as the α and the variance of the random variable is going to be 0. That means, for a constant random variable the expectation is same as constant and the variance is going to be 0. And this is if and only if condition if the variance is 0 then you can conclude that is a constant random variable. When I say constant random variable the probability of X takes a value that constant is 1. Therefore, it is called constant random variable. Later I am going to discuss in detail.

So, here the remark is if the variance is 0 you can conclude the probability of X takes that constant is 1, if a constant then its expectation is the same and the variance is 0. Now, I can go to the 4th remark.

(Refer Slide Time: 22:12)



Variance of some constant times X + b if the variance exist, if the variance of X exist that is same as if you do the little calculation you can prove that is same as a square times variance of X plus variance of constant is 0. Therefore, it is going to be a square times variance of X. Whereas, E(a X + b) = a E(X) + b. Here, variance (aX + b) when variance of X exist then it is same as $a^2 Varince(X)$ because variance of constant is 0.

Introduction to Probability Theory and Stochastic Processes

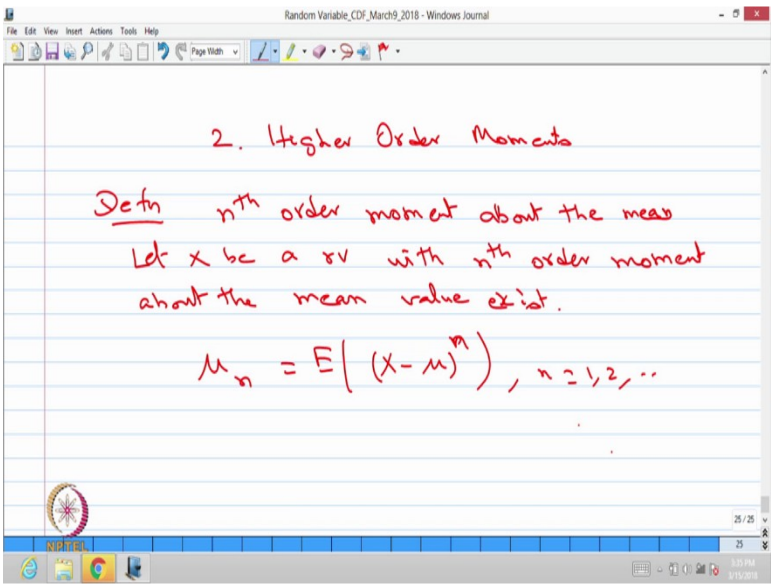
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 14

In this week we started Moments and Inequalities. Already we discussed in the last lecture mean and variance, mean is nothing but the first order moment and variance is nothing but the second order moment of the random variable. In the last class, we have discussed the first order moment and the second order moment with the examples. In this lecture we are going to discuss higher order moments.

Since, we have already discussed first and second order moment now we are going to discuss any nth order moment for the random variable if it exist followed by we are going to discuss the moment inequalities. So, let me start with definition of higher order moments.

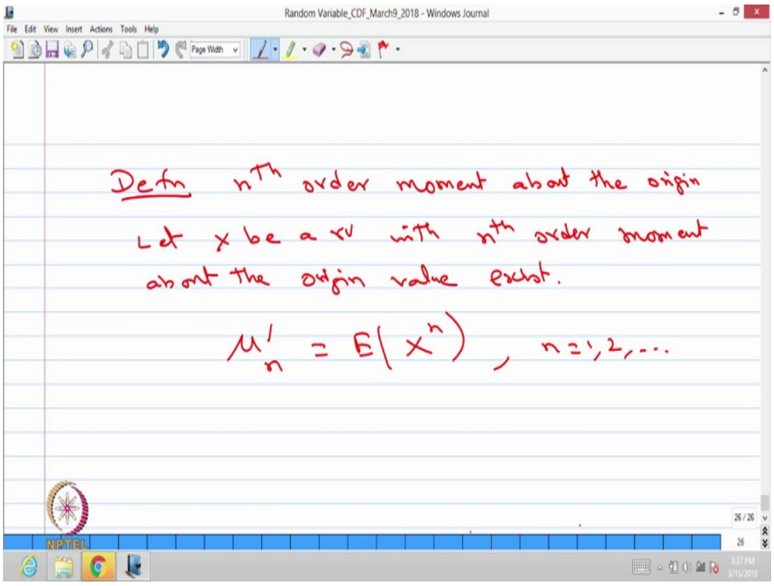
(Refer Slide Time: 00:58)



Higher order moments, the definition, that is a nth order moment about the mean. Let X be a random variable with nth order moment about the mean value exist. Then one can define with the notation μ_n that is nothing but expectation of X minus, the expectation of X is denoted by μ that is mean, $E\left[(X-\mu)^n\right]$that is going to be the nth order moment about the mean. Whenever it exist, it can be denoted by μ_n, whenever it exist, that is the right hand side expectation exist then you can denote by the letter μ_n that is $E[(X-\mu)^n]$, where n can takes the value it could be 1, 2 and so on.

Obviously, if you take the value n is equal to 1 that is nothing, but the μ_1 is E(X - μ) that is same as E(X) - μ that is going to be 0, when n is equal to 2, then it is nothing but the variance of the random variable X. So, provided the right hand side expectation exist, then one can define the nth order moment about the mean with the notation μ_n.

(Refer Slide Time: 04:08)

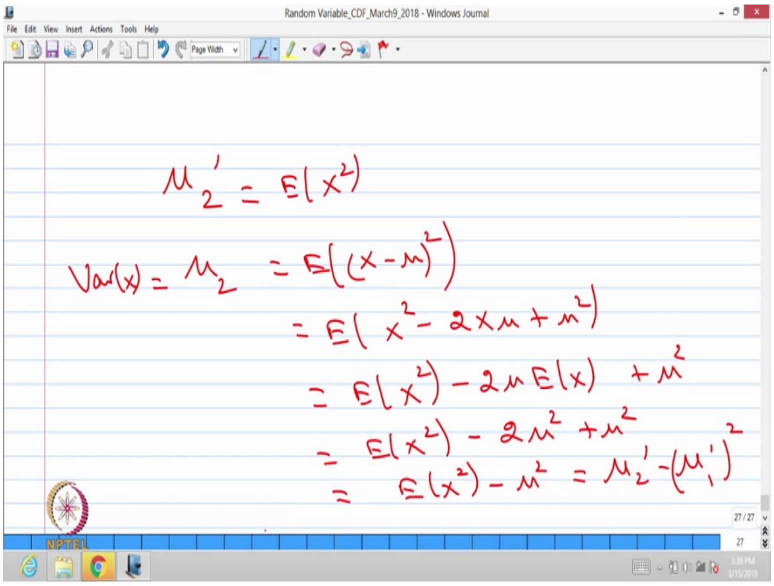


The same way I can define the nth order moment about the origin nth order moment about the origin or some books they use a word 0, both are one and the same.

Let X be a random variable with nth order moment about the origin value exist. Then one can define with the notation μ_n' that is nothing, but $E(X^n)$. Here again n can take the value 1, 2 and so on. So, when n is equal to 1 this is nothing but the mean or expectation of the random variable and two onwards it is going to be called as a nth order moment about the origin, provided the expectation exist, that is very important.

One can relate the second order moment about the origin with the second order moment about the mean.

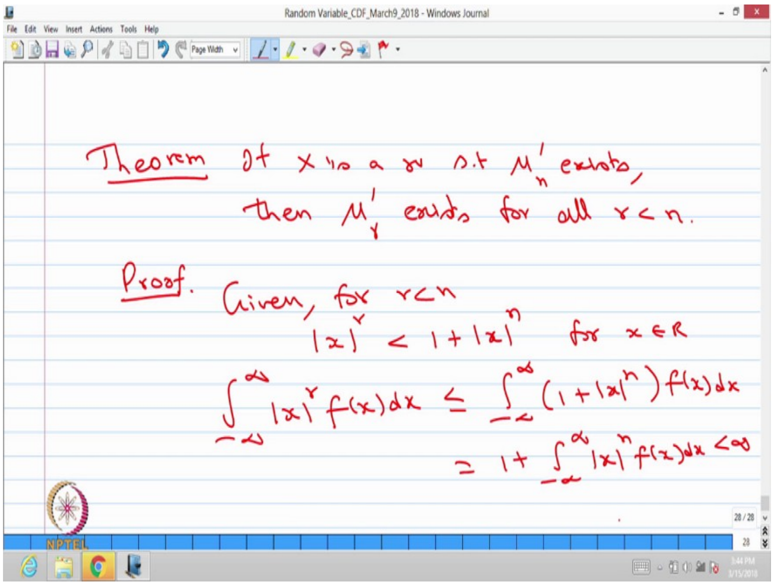
(Refer Slide Time: 05:55)



For example, μ_2' that is $E(X^2)$ and μ_2 that is nothing but the $E[(X-\mu)^2]$. This is same as variance of X. So, the $E[(X-\mu)^2]$ if you expand that is $E(X^2-2X\mu+\mu^2)$, expectation is a linear operator so it is E(X^2) - 2; and μ is constant so it is E(X). And μ is a constant, so $\mu^2 is$ constant $E(\mu^2)$ that is μ^2.

So, when you simplify you will get $E(X^2)$ minus this is 2 μ into μ therefore, 2 $\mu^2+\mu^2$. So, that is same as $E(X^2)-\mu^2$ that is same as $\mu_2'-(\mu_1')^2$. That means, $\mu_2=\mu_2'-(\mu_1')^2$. So, that means, one can write central moment about the mean in terms of central moment about the origin.

(Refer Slide Time: 08:01)



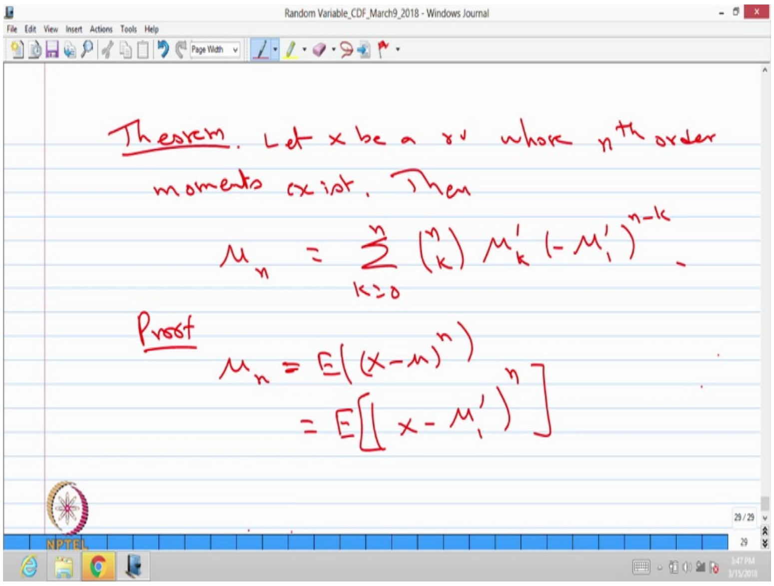
Next, I am going to give the one important result as a theorem. What the theorem says, if X is a random variable such that μ_n' exist, then μ_r' exists for all r < n that is a theorem.

Whenever for a random variable if the nth order moment about origin exist, then all the rth order moment about the origin exist for all r < n. You can give the proof of this theorem, given $|x|^r < 1+|x|^n$, this is for all x belonging to real.

We can conclude, suppose you consider X as a continuous type random variable, $\int_{-\infty}^{\infty} |x|^r f(x)dx$, where f(x) is the probability density function of a continuous type random variable that is $\leq \int_{-\infty}^{\infty} (1+|x|^n) f(x)dx$that is same as 1 + $\int_{-\infty}^{\infty} |x|^n f(x)dx$. And since nth order moment about the origin exist. Therefore, this quantity is going to be finite, the integration quantity therefore, this whole quantity is going to be finite. This implies $\int_{-\infty}^{\infty} |x|^r f(x)dx$ is a finite that is for all r which is less than n.

So, this is given, you can include one more statement, for r < n, given for r < n, $|x|^r < 1+|x|^n$ this is true therefore, both side you can do the integration by multiplying f(x) and given that its nth order moment about the origin exist, therefore, for all r < n the moment of r exist also.

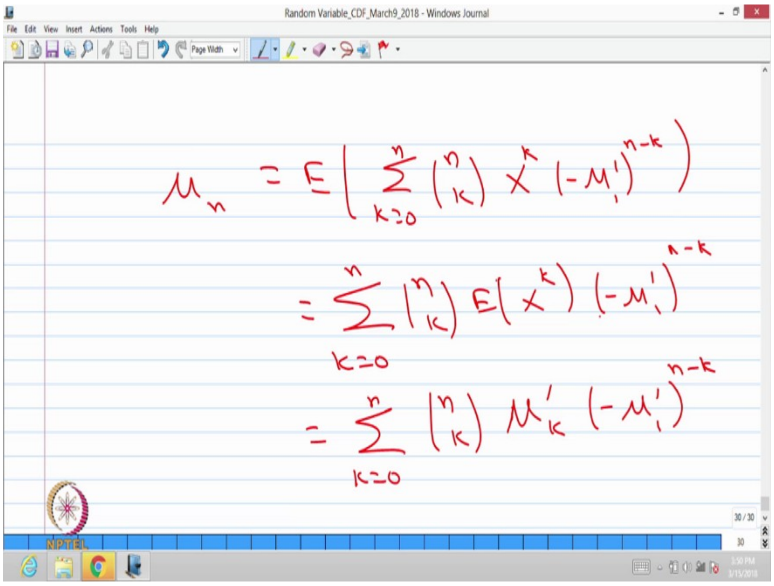
(Refer Slide Time: 11:47)



The next result as a theorem. Let X be a random variable whose nth order moment exists. Then one can write $\mu_n = \sum_{k=0}^{n} {}^{n}C_k \mu_k' (-\mu_1')^{n-k}$. This can be proved whenever nth order moment exist, then one can write the nth order moment about the mean is same as a function of all the previous order moments about the origin.

The proof is as follows you start with μ_n that is nothing but nth order moment about the mean, that is $(X-\mu)^n$, that is same as the E($(X-\mu_1')^n$) ,one can write μ as a μ_1'. Now, you can go for the binomial expansion of $(X-\mu_1')^n$, that is same as E($\sum_{k=0}^{n} {}^{n}C_k X^k (-\mu_1')^{n-k}$) that is same as the n C k is a constant that is not a random $(-\mu_1')$ that is also not random. Therefore, expectation can be taken inside that is $\sum_{k=0}^{n} {}^{n}C_k E(X ¿¿ k)(-\mu_1')^{n-k} ¿$.

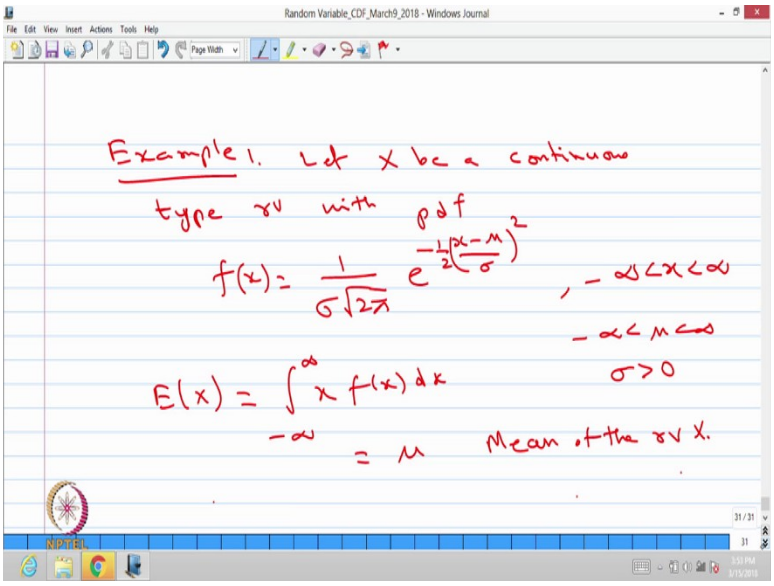
(Refer Slide Time: 14:15)



I can rewrite $E(X^k)$ as the kth order moment about origin. Therefore, this is going to be $\sum_{k=0}^{n} {}^nC_k \mu_k' (-\mu_1')^{n-k}$. Because of the previous theorem when the nth order moment about mean exist that means, all the previous order also exists. Therefore, this is a valid statement. With the help of previous moments about the mean you can always find the moment of nth order about the origin.

In conclusion with the previous starting from first to nth order moment about the origin one can get nth order moment about the mean, one can go for one easy example of how to find the nth order moment for some random variable which is of the continuous type.

(Refer Slide Time: 16:31)

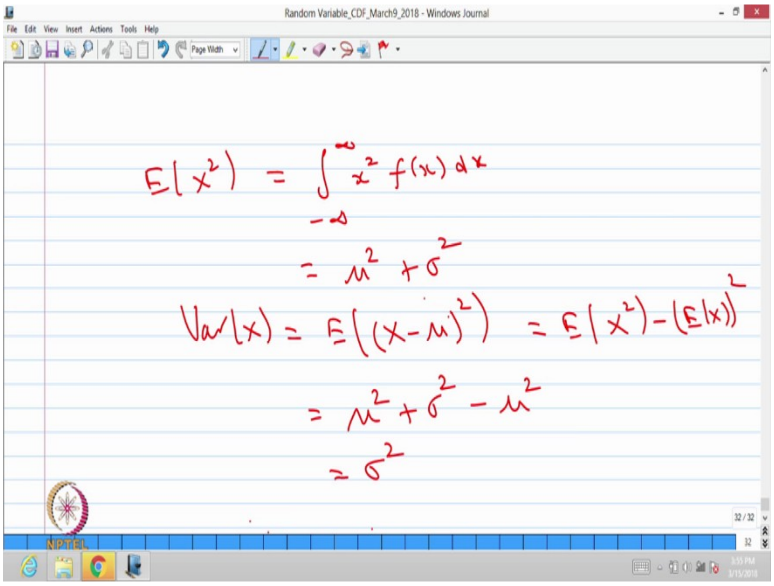


Let X be a continuous type random variable with probability density function f(x) = $\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, where x lies between -∞ to ∞. So, this is a probability density function of a continuous type. Later we are going to call it as a normal distribution when we are discussing standard distributions.

So, now, we will keep it as a continuous type random variable with the probability density function, $\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Always the σ and μ values are given. One can say the μ value can lies between -∞ to ∞. Whereas, the σ quantity is always greater than 0. What is a meaning of μ and σ? That also can be discussed.

In this example, if you find out E(X) that is minus infinity to infinity x times probability density function with the assumption that the expectation exist we will try to find the value $\int_{-\infty}^{\infty} x f(x) dx$. This is going to be after simplification you can get the answer that is μ. I am not going for the simplification of this integration as it is. If you substitute the f(x), $\int_{-\infty}^{\infty} x f(x) dx$ you can get the value μ, this μ is going to be call it as a mean. That is called the mean of the random variable X here.

(Refer Slide Time: 19:32)

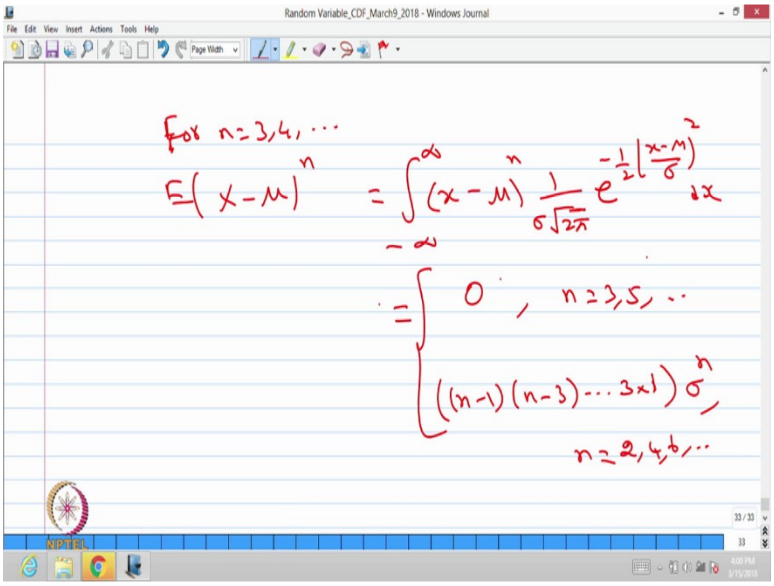


Similarly, if you compute $E(X^2)$ nothing but $\int_{-\infty}^{\infty} x^2 f(x)dx$. One can able to get by, after some simplification you can get $\mu^2+\sigma^2$ substituting f(x) is $\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. Therefore, the variance of the random variable X that is $E\left[(X-\mu)^2\right]$that is same as $E(X^2)-E(X)^2$.

Just now we got it $E(X^2)$ is $\mu^2+\sigma^2$and E(X) that is mean that we got it as a μ that is $-\mu^2$ therefore, you get variance is σ^2. That means, for this continuous type random variable the mean is going to be μ and the variance is going to be σ^2.

We have another measure that is a positive square root of variance that is called as standard deviation. So, here the sigma is the standard deviation because sigma square is a variance and the positive square root of variance that is called standard deviation. For this continuous type random variable, the sigma is the standard deviation and sigma square is the variance. So, we got first moment that is the mean, variance we got it a sigma square, now we can go for higher order moments.

(Refer Slide Time: 21:56)



That is E$\left[(X-\mu)^n\right]$ for n is equal to 3 onwards, because for n is equal to 2 that is famous variance, we got it already. So, we are computing the nth order moment about the mean from 3 onwards, that is same as $\int_{-\infty}^{\infty} (x-\mu)^n \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$.

If you see the integration very carefully when n is a odd positive integer then the integration value is going to be 0 because $e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $is\, an$ even function when n is odd positive integer the whole integration values is going to be 0. Therefore, you can immediately conclude, this is going to be 0 for n is equal to 3, 5 and so on. Now the question is what is a value when n is going to be the even positive integer. By even positive integer one can simplify this integration and you can get the answer that is ((n-1)(n-3)…3.1)σ^n when n is going to be 2, 4, 6 and so on.

So, for this continuous type random variable which is nothing but the normal distribution with the mean μ and the variance σ^2, we are finding the nth order moment about the mean for all for all the odd powers it is going to be 0 for the even you get the expression is ((n-1)(n-3)…3.1)σ^n, when n takes even positive integers.

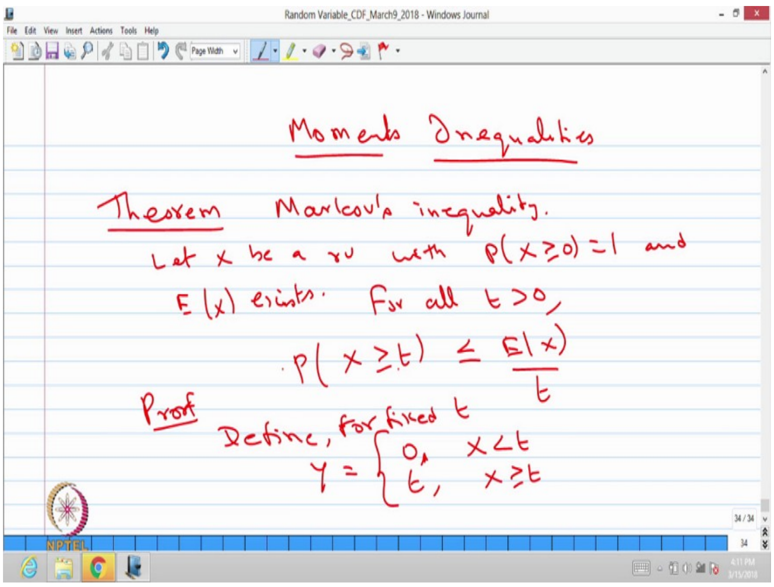
Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 15

So, even though we have done only one example for finding the nth order moment, there are many more problems we may need to find out the nth order moment when we started discussing some standard distribution. So, for time being we will stop it with only one example of finding nth order moment, later we will do the similar problems after we discuss the standard distributions.

(Refer Slide Time: 00:40)



So, now we are moving into the next topic that is called moments inequalities. It is a very important topic in the sense sometimes it is very difficult to find the distribution, but you may have the moments with you. That means, whenever the distribution of the random variable is not known to you, at the same time we know the fewer moments for example, first order moment or second order moment or the nth order moment.

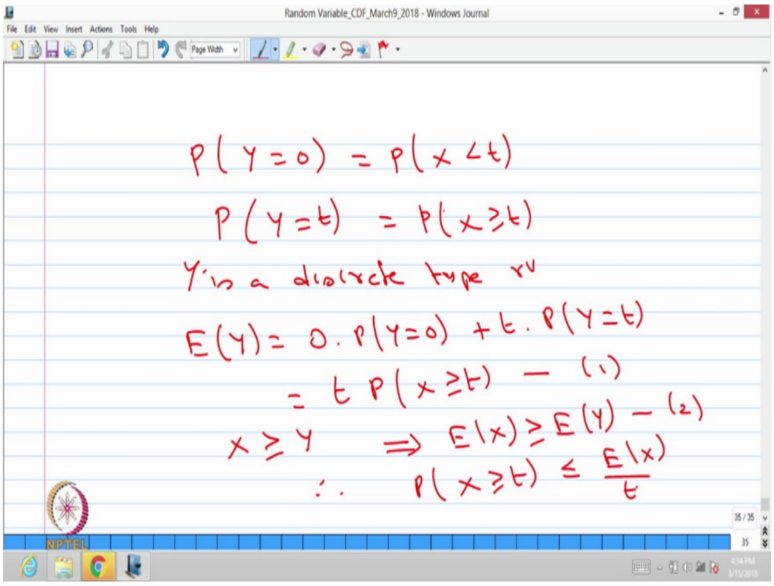
Then it is possible to find out the probability of some events not exactly in the form of lower bound or upper bound. I am repeating the statement, when we know the distribution of the random variable. That means if it is a continuous type you know the probability density function, if it is a discrete type you know the probability mass function of the random variable you can always check whether the moments exist first. After checking the moments exist then you can find the moments either it is first order moment or second order moment or any nth order moment, that is if you know the distribution of the random variable. But sometimes you may not know the distribution of the random variable you may know the moments as much as possible. That means, till the nth order that n could be 1 or 2 or whatever be the number positive integer. In that case you can find the lower bound or upper bound for the probability of some events that is what we are going to do.

So, let me start with first result as a theorem that is called Markov's inequality. What this Markov inequality says? Let X be a random variable, let X be a random variable with P{X ≥ l0} = 1. I am saying that is a non-negative random variable and the expectation of X exist. What the theorem says for all t > 0, the P{X ≥ t} that always has the upper bound that is E[X]/t. The theorem says you do not know the distribution of the random variable which is a non negative random variable. That means the P{X ≥ 0} = 1. And you know that the expectation exists and you know the expectation value also that is E[X], the value of mean or the expectation is known to you.

Then for all t greater than 0 the P{X ≥ t} ≤ $\frac{E[X]}{t}$, that is nothing but the probability of right tail, the right tail probability has the upper bound which is a mean divided by small t for every t, t > 0, this can be proved easily.

The proof is as follows I am going to define another random variable that is Y which takes a value 0 for X less than t and this value is going to be t when X is greater than or equal to t and this definition is for fixed t. I am defining the random variable Y for fixed t, either it takes a value 0 whenever X is less than t or it takes a value t when X is greater than or equal to t. I have not said X is a discrete type random variable or continuous type random variable or mixed type random variable, I said only it is a non negative random variable. That means, probability of X is greater than or equal to 0 is 1. The way I define Y as a function of X it takes a value 0 or t for fixed t therefore, you can say the P{Y = 0} = P{X < t}.

(Refer Slide Time: 06:29)

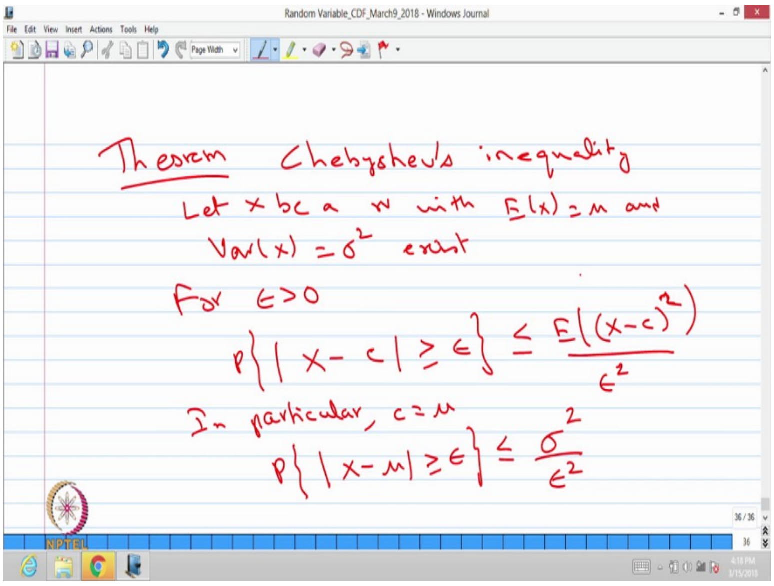


Similarly, you can say the $P\{Y = t\} = P\{X \geq t\}$. That means, the random variable Y has a mass at 0 and small t this is for fixed t. So, in material of the random variable X is a discrete or continuous or mixed, the Y is going to be a discrete type random variable. Y is a discrete type random variable with the values taken 0 and t whose probability mass function is same as $P\{X < t\}$ and $P\{X \geq t\}$.

Now, one can find the E[Y], that is nothing but since it is a discrete type random variable, $0P\{Y = 0\} + tP\{Y = t\}$. This is same as t $P\{Y = t\}$ that is same as t $P\{X \geq t\}$.

We know that $X \geq Y$, for fixed t, $X \geq Y$ this implies the $E[X] \geq E[Y]$. So, I am going to use the result 1 and the result 2, $E[Y] = t\,P\{X \geq t\}$ and $E[X] \geq E[Y]$. Therefore, by using the result equation 1 and 2, by using equation 1 and the inequality 2 we get $P\{X \geq t\} \leq E[X]/t$. That means, the right tail probability for some fixed t which has the upper bound mean divided by the t, where t is greater than 0. So, this inequality is very important in the sense without knowing the distribution as long as you know the mean or expectation, one can find the upper bound for the right tail probability.

(Refer Slide Time: 09:41)

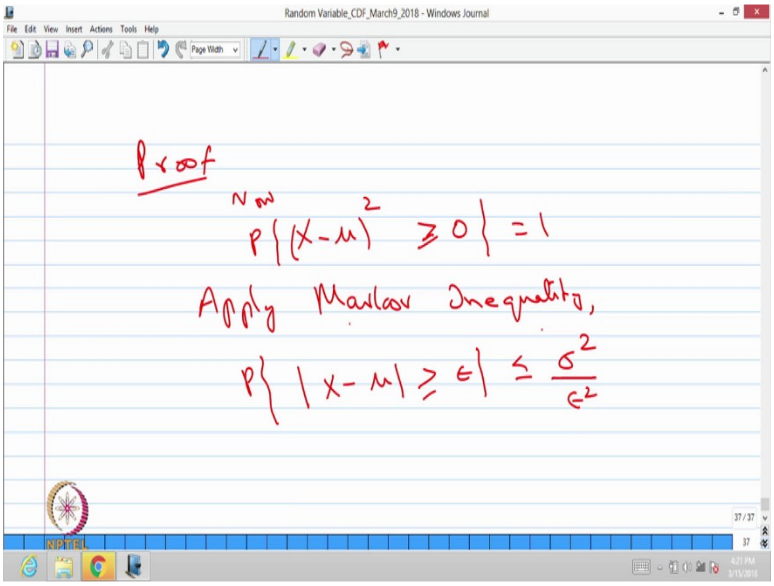


The second inequality, I am going to give it as a theorem that is called Chebyshev's inequality. What this theorem says let X be a random variable with both mean and variance exists in the earlier inequality only mean exist. We are not making the restriction of second order moment exist, now we are making a variance also exist, but here we are not making a condition it is a non-negative random variable. These are all the two changes with the Markov inequality and the Chebyshev's inequality.

What this theorem says for epsilon greater than 0 the $P\{|X-c|\geq\epsilon\}$, this probability has the upper bound that is $\frac{E\left[(X-c)^2\right]}{\epsilon^2}$, where c is a constant. In particular, the c is same as mean of the random variable μ, then the above result $P\{|X-\mu|\geq\epsilon\}$ less than or equal to, when c becomes a μ that is a mean or expectation of the random variable X, the $E[(X-\mu)^2]$ that is nothing but the variance of the random variable X, that is sigma that is same as $\frac{\sigma^2}{\epsilon^2}$.

The previous result is a right tail probability whereas, this one is the tail probabilities $P\{|X-\mu|\geq\epsilon\}$ that means, it is tail probabilities has the upper bound $\frac{\sigma^2}{\epsilon^2}$ for $\epsilon\geq 0$. This also can be proved easily.

(Refer Slide Time: 12:43)

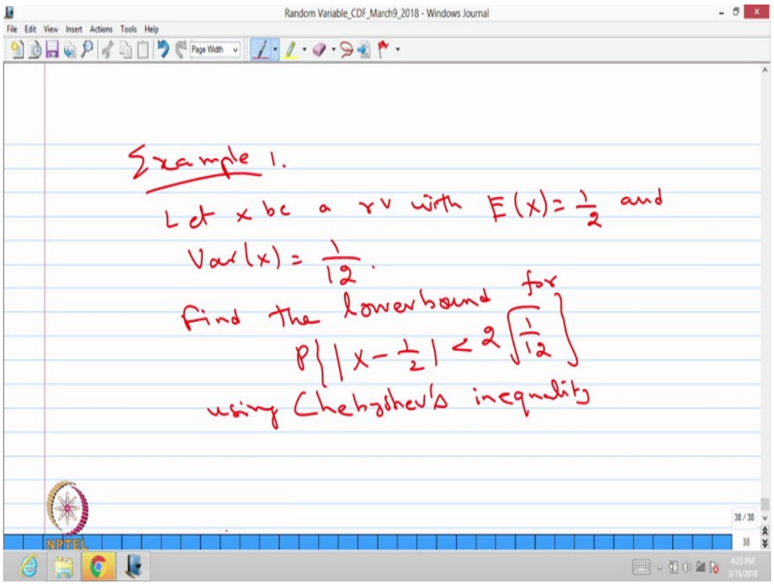


Proof: I can use the result of Markov inequality to prove the Chebyshev's inequality. That is a X is a random variable whose mean and the variance exist. Now I can use $(X-\mu)^2$ is a random variable not only it is a random variable $(X-\mu)^2$is a non-negative random variable, immaterial of the random variable X. So, $(X-\mu)^2$which is going to be greater than or equal to 0 whose probability is going to be 1. The P{$(X-\mu)^2 \geq 0$} =1 that is a non-negative random variable. Therefore, I can apply the Markov inequality because I treat $(X-\mu)^2$as a random variable, whose mean exists and since the variance of X exist that is same as mean of $(X-\mu)^2$.

Therefore, now I can apply Markov inequality because $(X-\mu)^2\geq$ 0 whose probability is 1 and E[$(X-\mu)^2$] that is same as variance of the X that is also exist. Therefore, this is going to be P{|X - μ| $\geq \epsilon$ }$\leq \frac{\sigma^2}{\epsilon^2}$. I am skipping in between two steps you apply first $(X-\mu)^2$and take the absolute; therefore, you will get for $\epsilon >$¿0 P{|X - μ| $\geq \epsilon$ }which has the upper bound $\frac{\sigma^2}{\epsilon^2}$ by using the Markov inequality.

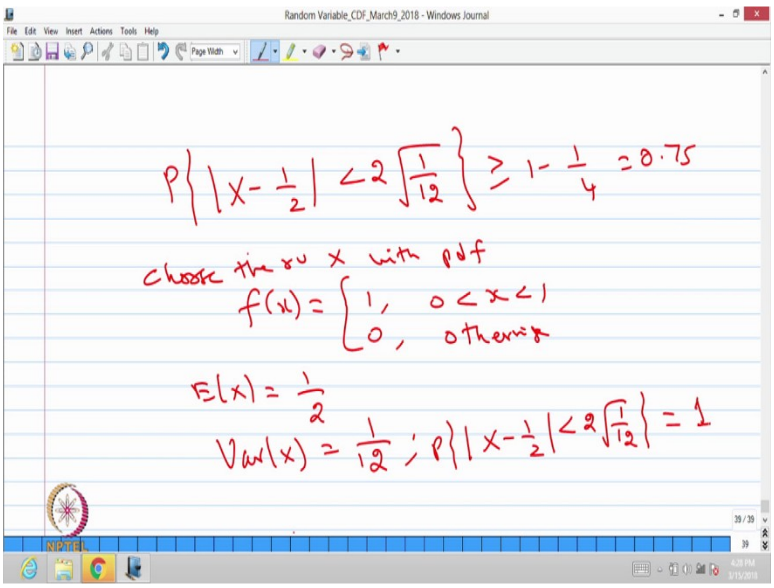
Now, we will go for one simple example how one can apply the Chebyshev's inequality.

(Refer Slide Time: 15:38)



That is example 1, let X be a random variable with mean of the random variable is 1/2 and variance is 1/12. Let us find the lower bound for the event probability of the event that is $P\{|X - 1/2| < 2\sqrt{(\frac{1}{12})}\}$. Find the lower bound for probability of this using Chebyshev's inequality. Using Chebyshev's inequality find the lower bound for the $P\{|X - 1/2| < 2\sqrt{(\frac{1}{12})}\}$. Because we have separate random variable mean and variance you do not know the distribution of this random variable.

(Refer Slide Time: 17:34)



One can find the lower bound by using the Chebyshev's inequality. That is P{|X – 1/2| < 2 $\sqrt{(\frac{1}{12})}$} ≥, 1/2 is a mean. I have given the Chebyshev's inequality for the tail probability and this is the probability of negation of tail probability therefore, the inequality changes and the value become 1 minus of variance divided by ϵ^2here the $\epsilon =¿2\sqrt{(\frac{1}{12})}$. So, if you do the simplification you will get 1 – 1/4 that is nothing but 0.75.

So, what this Chebyshev's inequality say is the negation of tail probability has a lower limit 0.75. And you know that the probability lies between 0 to 1. The lower bound 0.75 means if you choose the distribution or if you know the distribution of the random variable X, the exact probability will lies between 0.75 to 1, whereas without knowing the distribution of the random variable X, by knowing the mean is 1/2 and the variance is 1/12, one can conclude the lower bound for this probability is 0.75.

Let us fit some distribution for this random variable X whose mean is going to be 1/2 and the variance is going to be 1/12. Then we will find out what is the exact probability of this event. Let me repeat by using a Chebyshev's inequality without knowing the distribution we got the lower bound for this probability of the event that is 0.75. Suppose we choose the random variable X with some distribution in which the mean is going to be 1/2 and the variance is going to be 1/12, one can find the exact probability of this event.

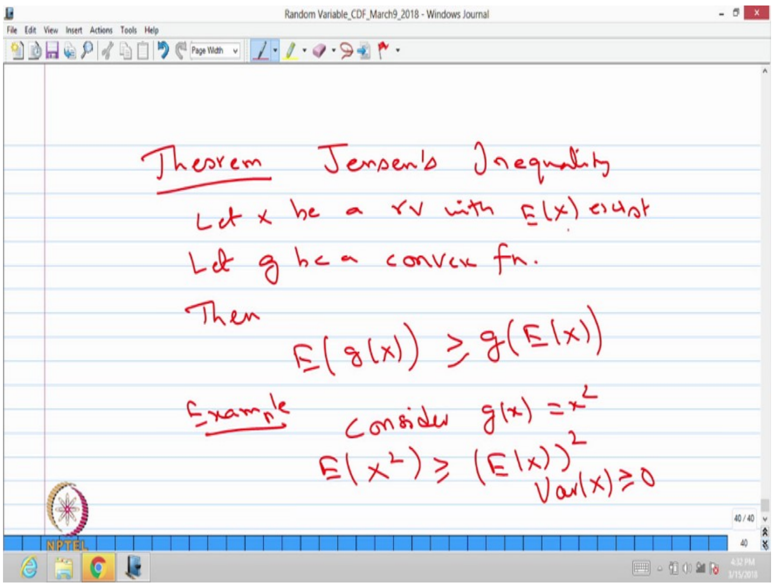
So, let us choose the random variable, the distribution of the random variable X such a way that the mean is going to be 1/2 and the variance is going to be 1/12. You can make out the probability density function of this random variable is 1, between 0 to 1 and 0 otherwise. Choose the random variable X with probability density function. That means, I am choosing a continuous type random variable with the probability density function that is f(x) is a 1 between the interval 0 to 1, otherwise 0.

You can verify if we find the mean for this random variable that is from $\int_{-\infty}^{\infty} xf(x)\,dx$, then substitute f(x), then integrate between 0 to 1 x, therefore, you will get the answer 1/2. Similarly, if you find out the variance of X for that either you can do $E[\left(X-\frac{1}{2}\right)^2]$ or you find out $E[X^2]$, you know already E[X] then variance of X is $E¿$]-$E[X]^2$. Either way you do you can get the answer variance of X is going to be 1/12. Therefore, it matches with the given problem the random variable X if you choose the probability density function these, you will get the mean is 1/2 and the variance is 1/12.

Now, you can find the same event $P\{|X - 1/2| < 2\sqrt{(\frac{1}{12})}\}$, if you find out the probability of this event which is going to be we can do the simplification and so on you can get the answer 1. That means, exact probability is going to be 1, by choosing a random variable whose probability density function is this whereas, without knowing the probability density function oh sorry without knowing the distribution of the random variable you are getting the lower bound for the probability that is 0.75. Instead of this distribution if you would have chosen some other distribution you may get some other value which also lies between 0.75 to 1, both are inclusive. So, the Markov inequality and Chebyshev's inequality are very useful to get the lower bound or upper bound for the probability of some events whenever the distribution of the random variable is unknown. If you know the distribution you can get the exact probability.

Now, we will move into the next inequality that is called Jensen's inequality as a theorem.

(Refer Slide Time: 23:21)

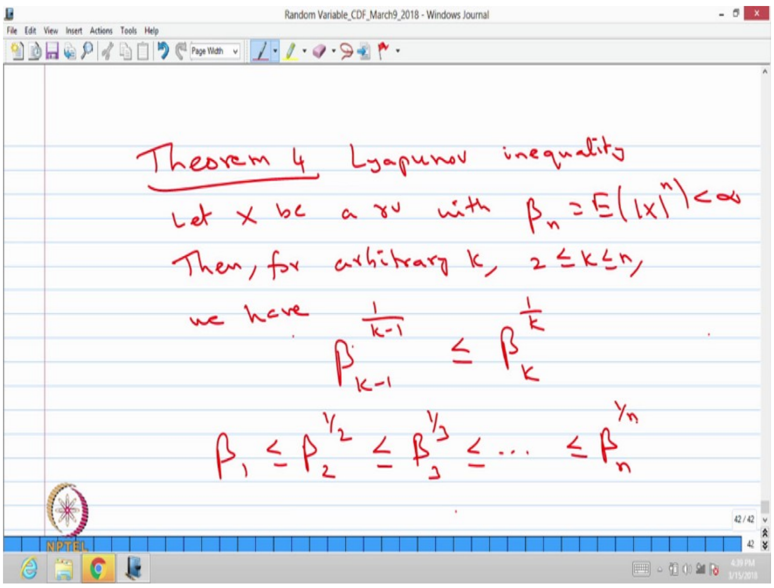


Jensen's inequality. What this inequality says let X be a random variable with the E[X] exist, let g be a convex function then what the theorem says, the $E[g(X)] \geq g(E[X])$. The Jensen's inequality says whenever the random variable whose expectation exist for a convex function g, E[g(X)] ≥ g(E[X]). There are many real world problems one can use this inequality to get the bound for expectation, whenever you interchange the expectation and the g interchanged.

Let us give a simple example for this. I am not going for the proof of this theorem, but you can go for the example. Consider $g(x)=x^2$, consider $g(x)=x^2$, you apply the theorem here that means, $E ¿] \geq E[X]^2$. This you know the result because this is nothing but variance of X that is always greater than or equal to 0. This implies this greater or equal to 0 that is same as variance of X is always greater than or equal to 0. So, like that you can think of any convex function to apply the Jensen's inequality.

Now, I am going for one more inequality of the nth order. The Markov inequality involves the first order, Chebyshev's inequality involves the second order, and the Jensen's inequality involves any convex function with the first order. Now, we are going for nth order moments in the form of inequality.

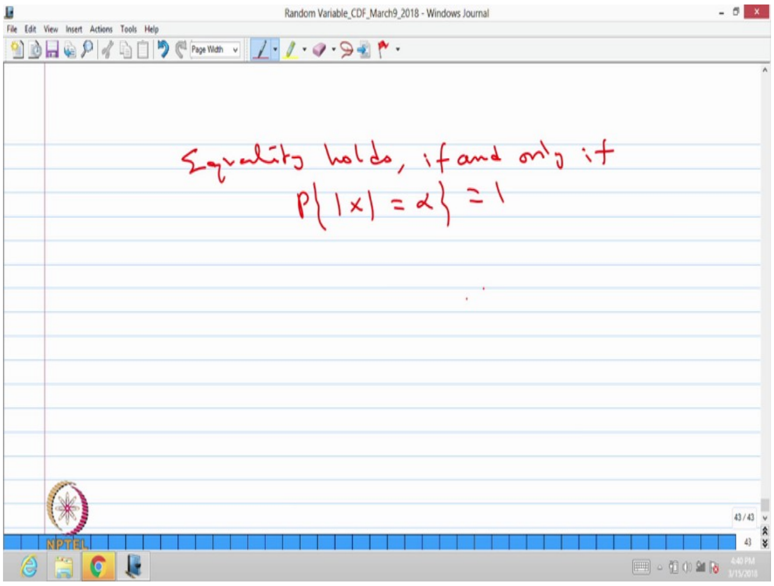
(Refer Slide Time: 26:54)



We will move into the 4th inequality as a theorem 4 it is called Lyapunov inequality. What this inequality says, let X be a random variable with $\beta_n$that is nothing but $E[|X|^n]$ which is a finite quantity that I denoted as a β_n which exist, then for arbitrary k, where k lies between 2 to n. This theorem says the $\beta_{k-1}^{\frac{1}{k-1}} \leq \beta_k^{\frac{1}{k}}$.

So, this is valid whenever the β_n exist, that is $E\left[|X|^n\right]$ exist for any n, n can be 1, 2, 3 and so on. So, if any β_n exists, then this inequality holds for k lies between 2 to n. That means, I am not going for the proof of this theorem. That means, the $\beta_1 \leq \beta_2^{\frac{1}{2}}$, that is always less than or equal to $\beta_3^{\frac{1}{3}}$ and so on till less than or equal to $\beta_n^{\frac{1}{n}}$.

(Refer Slide Time: 29:14)



So, the equality holds, equality holds if and only if the probability of |X| take some constant value that is going to be 1. Equality holds if and only if. So, I am not going to give the proof of this theorem whereas, we have given the proof of Markov inequality and Chebyshev's inequality.

In the later section we are going to include, we are going to solve some problems related to the Chebyshev's inequality and Markov inequality.

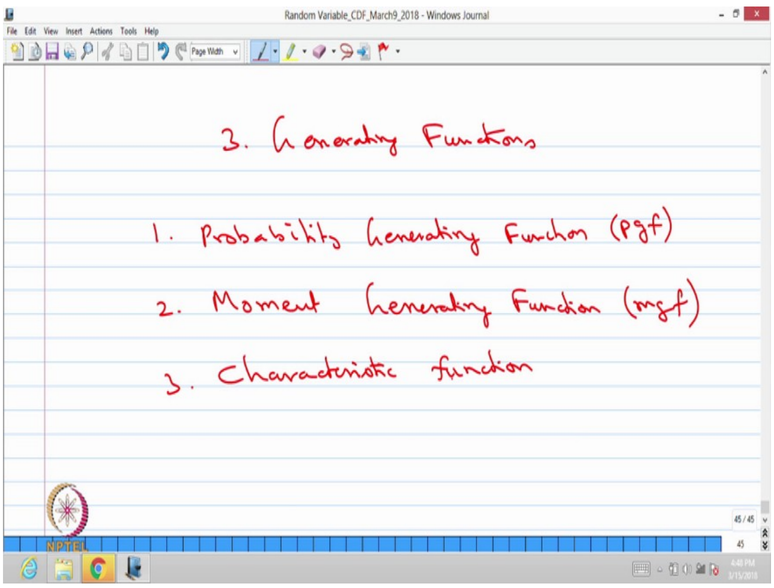
Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 16

So, we have already discussed Mean and Variance in the first lecture, Higher Order Moments and Moments Inequalities in the second lecture, now we are moving into the third lecture on Generating Functions.

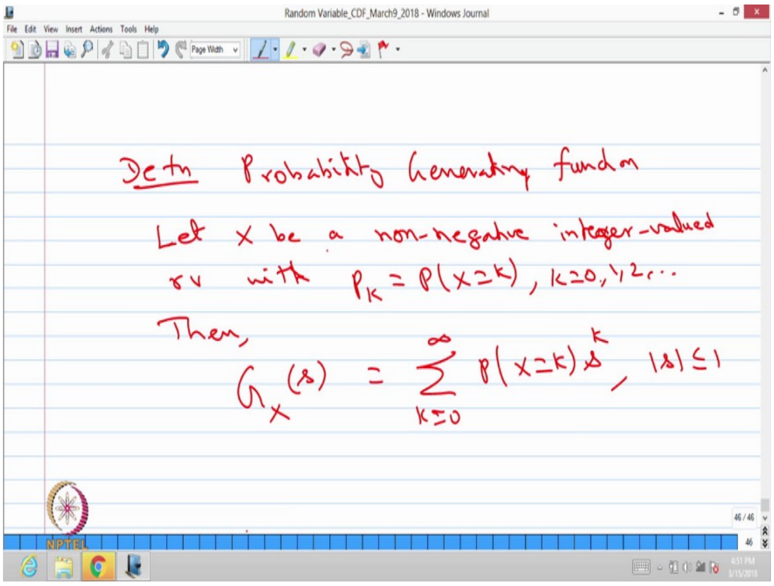
In this lecture we are going to discuss three important generating functions namely probability generating function, moment generating function and characteristic function. We will give the definition, some properties of these, and then one or two examples. And later we will find out the generating functions for some standard distributions in detail. So, as such now we will give the definition and the properties and one or two easy examples. So, lets start with third lecture that is generating functions.

(Refer Slide Time: 01:00)



In this we are going to discuss probability generating function in short it is pgf. The second we are going to discuss moment generating function that is mgf. The third we are going to discuss characteristic function, this is also generating function.

(Refer Slide Time: 02:16)

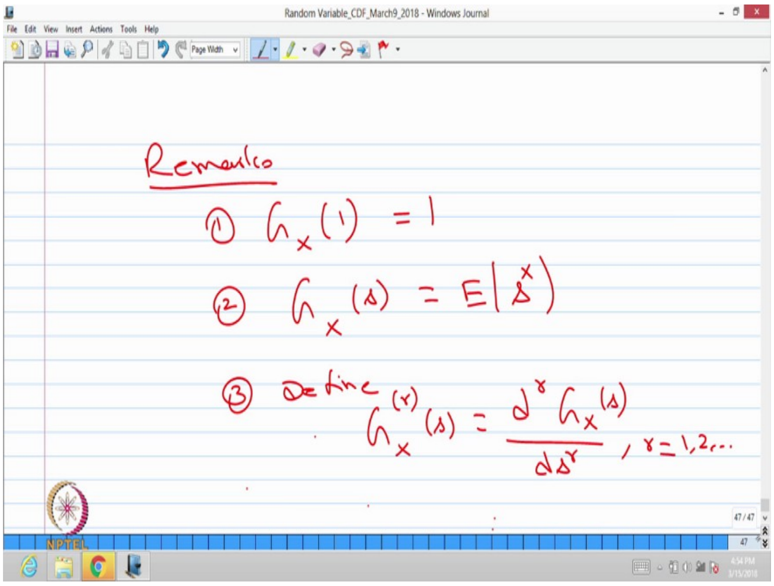


So, let us start with a first one probability generating function, the definition. Probability generating function: Let X be a non-negative integer valued random variable that is basically a discrete type random variable. In particular it is a non negative integer valued random with the probability of X takes the value k we denote it as a p_k, where k takes a value 0, 1, 2 and so. There is a possibility it may take a finite number of values also. Then one can define the probability generating function as $G_X(s)$ that is nothing but $\sum_{k=0}^{\infty} P(X=k) s^k$, where s is the real variable which lies between -1 to 1.

So, this is a real valued function, it is a function of s in the form of series with the probability mass at the point k multiplied by s^k. So, the right hand side is a series, the left hand side we are denoted by the $G_X(s)$. Since, the right hand side is a series, this series converges within the interval $|s| \leq 1$. We are not bothering about whether this series converges or not outside this interval, what we are saying is within the interval s lies between -1 to 1, the right hand side series converges, gives the function of s that is called the probability generating function of the random variable X, where X is a non negative integer valued random variable. That means, whenever X is a non negative integer valued random variable one can define the probability generating function as a function of s.

We will go for few remarks then we will give some examples.

(Refer Slide Time: 05:37)

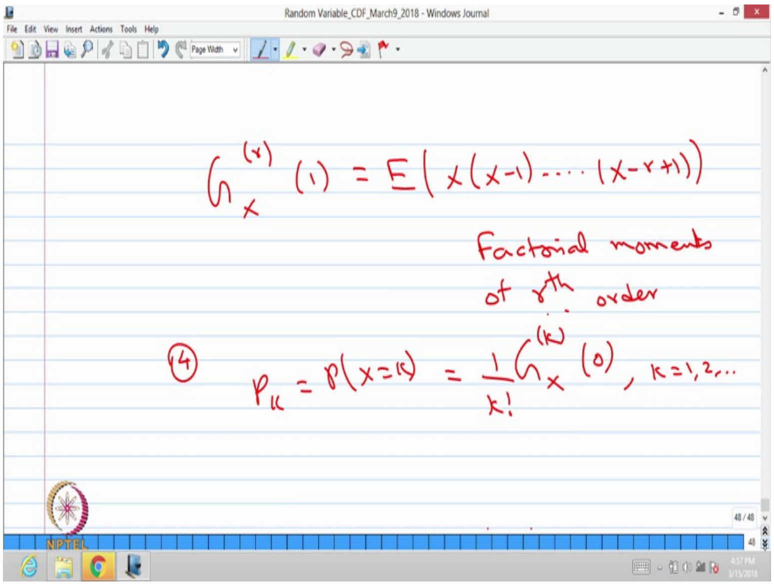


As a remarks, the first remark, since I said that the right hand series is converges between -1 to 1, if you substitute the value s =1, you will get the $\sum_{k=0}^{\infty} P(X=k)$ that is nothing but 1 because that is a probability mass function. So, this value is going to be 1.

The second remark you can relate probability generating function with expectation of function of a random variable that is the probability generating function of X is same as the way it is summation of probabilities multiplied by s^k, one can write it is nothing but $E[s^X]$.

Third remark one can relate the probabilities with the derivative of generating function. Suppose I define; suppose I define the $G_X^{(r)}(s)$ is nothing but $\frac{d^r G_X(s)}{ds^r}$.Suppose I define $G_X^{(r)}(s)$ is the rth successive derivative of probability generating function with respect to s where r can take the value 1, 2 and so on.

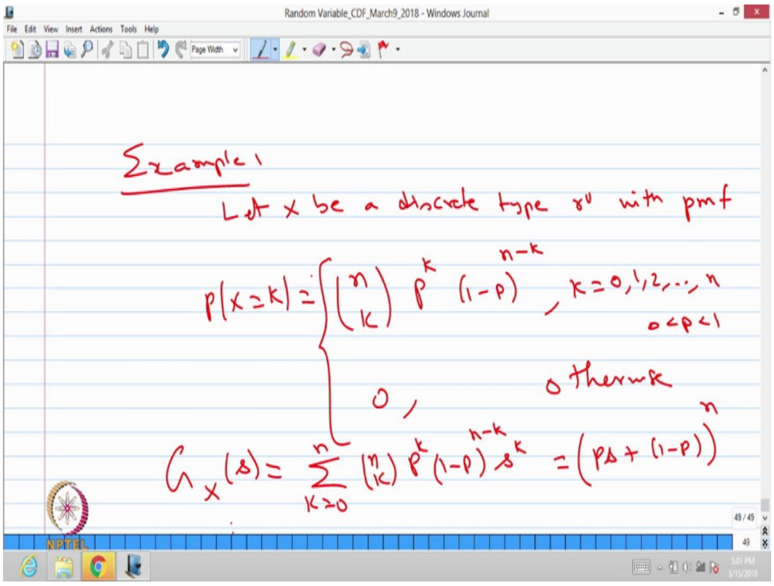
(Refer Slide Time: 07:58)



I can write down rth derivative evaluated at the point 1 that is nothing but E[X(X-1)…(X-r+1)]. The way I defined $G_X^{(r)}(1)$that is same as E[X(X-1)…(X-r+1)]. The right hand side is called factorial moments of rth order. Sometimes, if you want to find out the variance you can find the factorial moment of a second order through that you can find the second order moment or variance.

Similarly, I can find from the probability generating function by taking the derivative I can get the probability mass at the point k, that is p_k that is nothing but the P(X = k) that is same as the $\frac{1}{k!}G_X^{(k)}(0)$,that is going to be the P(X = k), here k can take the value 1, 2 and so. After you get the kth successive derivative of probability generating function substituting value s = 0 multiplying by $\frac{1}{k!}$ will give P(X = k). So, this result is by knowing the probability generating function you can get the probabilities. Whereas, the definition says if you know the probabilities you can get the probability generating function for a non negative integer valued random variable.

(Refer Slide Time: 10:55)

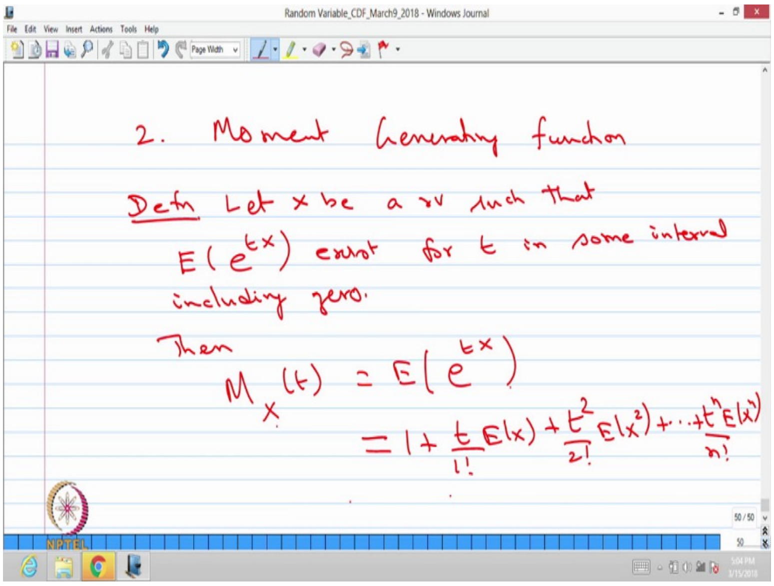


Let us go for finding the probability generating function for the random variable which is of a non negative integer valued. Example 1, let X be a discrete type random variable with probability mass function P(X = k) = ${}^nC_k\, p^k(1-p)^{n-k}$, where k takes a value 0, 1, 2 and so on till n, where n is a positive integer, otherwise the probability mass function is going to be 0. So, this satisfies the properties to define the probability generating function. So, one can go for finding the probability generating function for this random variable.

So, $G_X(s)$ that is nothing but $\sum_{k=0}^{\infty} P(X=k)s^k$, $\sum_{k=0}^{n} {}^nC_k\, p^k(1-p)^{n-k} s^k$. So, here the p value lies between 0 to 1 that I did not specified earlier. And if you see the summation this is always converges $\sum_{k=0}^{n} {}^nC_k\, p^k(1-p)^{n-k} s^k$. So, you do not need to expand and then do the simplification. $p^k s^k$ you can keep it together, then it becomes $(p\,s)^k(1-p)^{n-k}$ and this is nothing but the binomial summation. Therefore, one can easily write $(ps+1-p)^n$. By combining $p^k s^k$ you can easily get the result $(ps+1-p)^n$ that is going to be the probability generating function for this random variable.

Later we are going to introduce this random variable as a binomial distribution with the parameters n and p or with the parameters p and n. p is lies between 0 to 1 open interval, n is a positive integer then the probability generating function is $(ps+1-p)^n$.

(Refer Slide Time: 14:27)

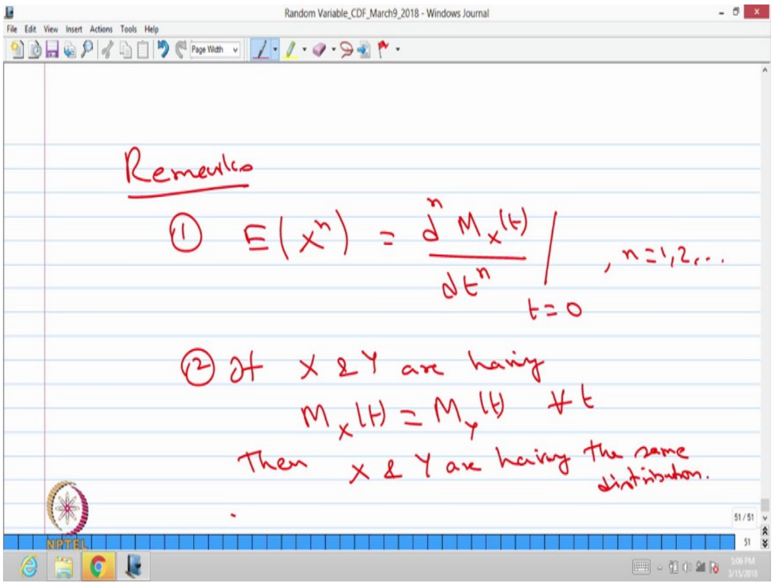


Now, we will move into the second generating function that is moment generating function. In short it is mgf. The definition is as follows. Let X be a random variable such that the $E[e^{tX}]$ exist for t in some interval including 0.

So, as long as $E[e^{tX}]$ exist for t in some interval which include 0. In that case one can define the mgf or moment generating function of the random variable X. As, is a notation $M_X(t)$that is nothing, but $E[e^{tX}]$. Since I have written $E[e^{tX}]$, e^{tX} can be expanded therefore, you will get that is same as, that is same as $1+\frac{tE[X]}{1!}+\frac{t^2E[X^2]}{2!}+\ldots+\frac{t^nE[X^n]}{n!}$. Since we said $E[e^{tX}]$exist.

Therefore, all the moment of order n exist about the origin and this series also converges, then only one can find the mgf of the random variable X. It is a function of a t. If fewer moments exist and other moments does not exist one cannot define the mgf of the random variable.

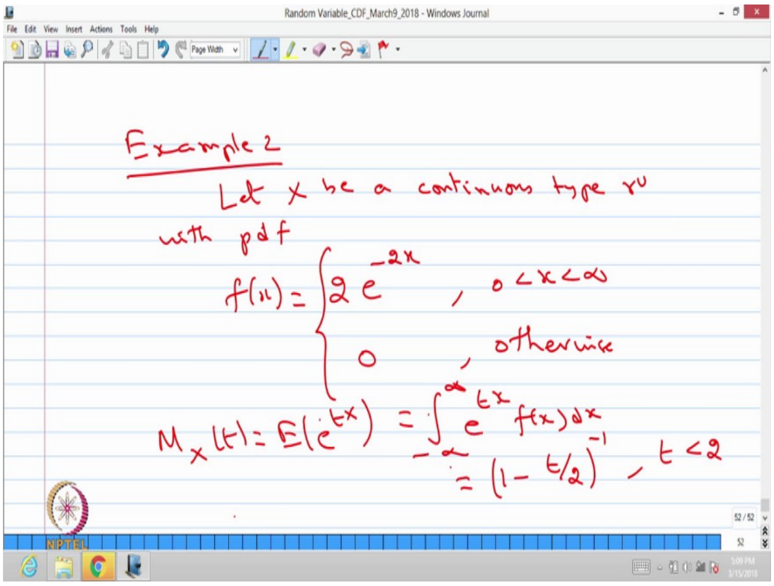
(Refer Slide Time: 17:31)



As a remark, the first remark, one can get the nth order moment from the mgf by successive derivative of moment generating function n times with respect to t then substituting t equal to 0. One can get the nth order moment about the origin from the moment generating function by $\frac{d^n M_X(t)}{dt^n}\Big|_{t=0}$ that is same as $E[X^n]$.

The second if two random variables or their mgf's are same for all t. If two different random variables whose mgf's are same for all t, then one can conclude both the random variables are having the same distribution. This result is valid for the probability generating function also, I have not mentioned. If two random variables of probability generating functions are same for all s then both the random variables are having the same distribution.

(Refer Slide Time: 19:47)



Let us go for the example finding the mgf. Let X be a continuous type random variable with probability density function f(x) = $2e^{-2x}$ when x is lies between 0 to infinity, 0 otherwise. The same example which we have taken earlier for finding the mean and variance. Now, we are finding the mgf of same random variable that is nothing but mgf as a function of t that is $E[e^{tX}]$, that is same as $\int_{-\infty}^{\infty} e^{tx} f(x)\,dx$. Here we are doing with the assumption that the mgf $E[e^{tX}]$exists. Then we are finding the mgf.

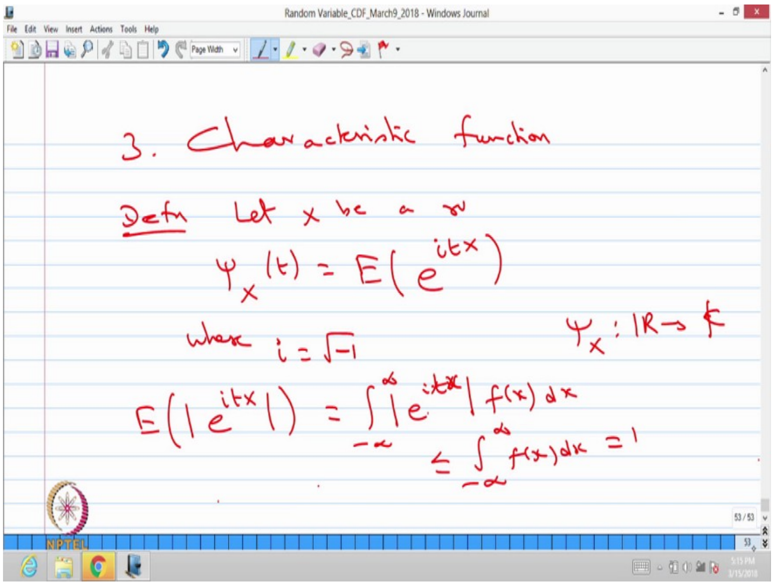
Substitute f(x) = $2e^{-2x}$between the interval 0 to infinity n. After simplification you can get the answer that is $\left(1-\frac{t}{2}\right)^{-1}$, whenever this result is valid when t < 2; that means, the e^{tX} is a finite quantity for t < 2 then only we can define the mgf and that mgf quantity is going to be $\frac{1}{1-\frac{t}{2}}$.

So, this is a very important, not for all t between minus infinity to infinity the mgf exist. For this random variable the mgf exist between minus infinity to 2 and the mgf value is $\frac{1}{1-\frac{t}{2}}$.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 17

(Refer Slide Time: 00:02)



We will go for third generating function that is Characteristic Function. This is also generating function. The definition for characteristic function is as follows. Let X be a random variable. The characteristics function of the random variable X as a function of t that is $\psi_X(t)$ is a notation is defined as $E[e^{itX}]$, where i is $\sqrt{-1}$is a complex one.

You compare the definition of probability generating function, moment generating function and characteristic function. Probability generating function can be defined only for the random variable which is non negative integer valued random variable. The moment generating function is valid for the random variable in which$E[e^{tX}]$ is a finite for all t in some interval including 0. Whereas, the characteristic function there is no restriction for the random variable. That means, for any random variable or for all random variable one can define the characteristic function, that is $E[e^{itX}]$.

Even though we use the expectation of function of random variable, we are not making the provided condition expectation exist. That means, the right hand side quantity always exist.

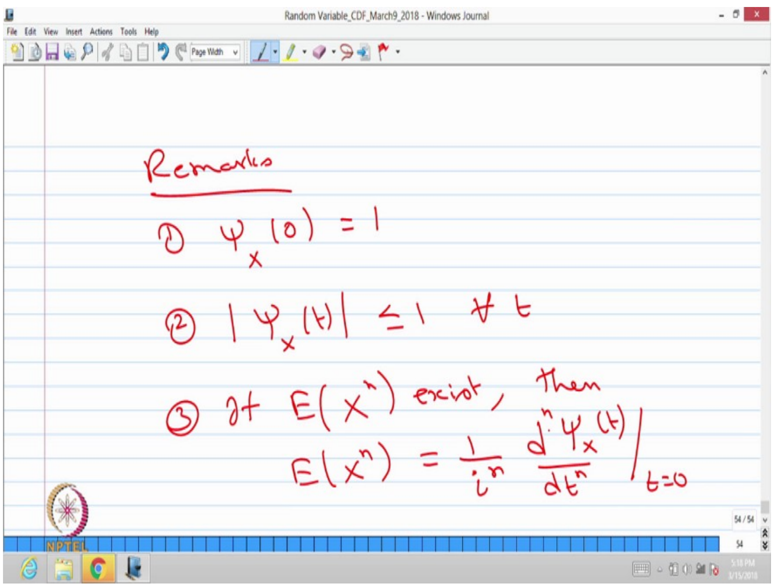
That can be proved, because once you say the expectation of a function of random variable exist provided in absolute sense which should be finite, we can verify $E¿$.

So, here the provided condition is not necessary because the $E¿$. This quantity is same as $\int_{-\infty}^{\infty} ¿e^{itX} \vee f(x)dx$. And we know that i is $\sqrt{-1}$ and $¿e^{itx} \vee \leq 1$. Therefore, this is less than or equal to $\int_{-\infty}^{\infty} f(x)dx$. Here, when I go for integration from minus infinity to infinity, I assume that it is a continuous type random variable. Therefore, I am going for integration with the probability density function as a multiplication. Suppose if it is a discrete type random variable then it is a summation absolute probability mass function.

So, here I have considered X is a continuous type random variable. And this quantity is going to be 1. So, always $E[¿e^{itX}|] \leq 1$, that is a finite quantity. Therefore, there is no need of provided condition for characteristic function.

One more observation the earlier two functions are the real valued function whereas, this is a complex valued function. That is a ψ characteristic function it is a complex valued function.

(Refer Slide Time: 04:33)

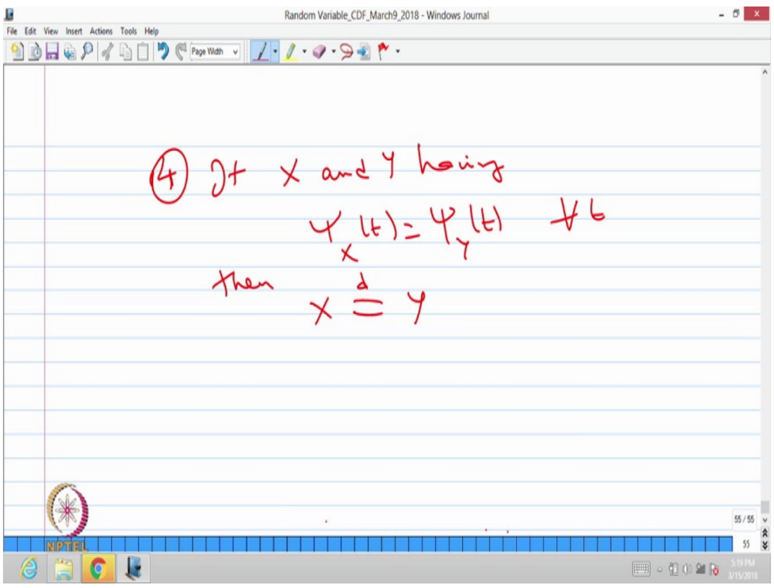


Now, we will go for few remarks over characteristic function. The first remark, if you substitute t = 0 in the characteristic function, that is same as E[1] = 1.

The second remark, if you find out, since it is a complex valued function you can go for absolute of characteristic function, that is for all t if you take an absolute of this that is same as absolute of $E[e^{itX}] \leq E¿$. Just now we got that result is 1. Therefore, that is going to be 1, this is for all t. The absolute of characteristic function is always less than or equal to 1 for all t.

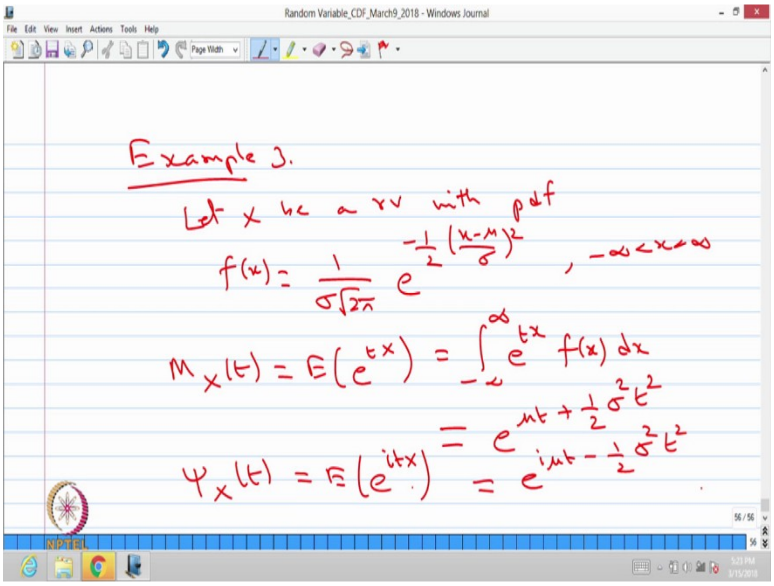
The third remark, if $E\left[X^n\right]$ exists, because the characteristic function that does not say the moment generating function exist or not. If the nth order moment exist then one can find the nth order moment about the origin by $\frac{d^n \psi_X(t)}{dt^n}$. Then substitute t equal to 0 then multiplied by i^n. If the moment of nth order about the origin exist from the characteristic function by $\frac{1}{i^n} \frac{d^n \psi_X(t)}{dt^n} \Big|_{t=0}$, one can get the nth order moment about the mean.

(Refer Slide Time: 07:17)



The next remark, if two random variables, if two random variables having characteristic functions same for all t then you can conclude both are identically distributed, is a notation, both are random variables X is equal to Y having the same distribution. Therefore, I write a d above the equal symbol. If two random variables having the same characteristic function for all t then we can conclude both are having the same distribution, so all these three generating functions this result valid, if two random variables having the same thing then both are going to have the same distribution.

(Refer Slide Time: 08:33)



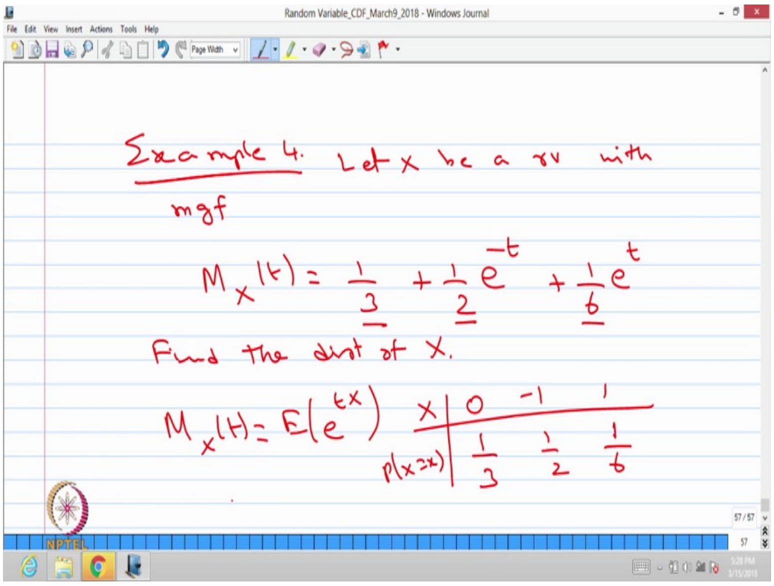
As an example, already we discussed two examples, this is third example. Let X be a random variable with probability density function f(x): $\frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2}$, where x lies between -∞ to ∞; μ lies between -∞ to ∞ that is a parameter and the σ is greater than 0. So, this is a probability density function of continuous type random variable, later we are going to say this is a normal distribution with the mean μ and the variance σ^2. You can find the mgf of the random variable. You cannot find the probability generating function because it is a continuous type random variable. So, this is nothing but the $E[e^{tX}]$ that is same as $\int_{-\infty}^{\infty} e^{tx} f(x)\, dx$ you substitute the above probability density function

So, in this example we are finding both moment generating function as well as characteristic function. If you do the simplification it is going to be $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. I am skipping all the calculation, substitute the f(x) then do the integration after simplification you can get the answer it is $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. This is a moment generating function of the random variable X whose probability density function is given. This, you can find the characteristic function that is

$E[e^{itX}]$. You see that the difference between moment generating function and the characteristic function is replacing t by it, where i is $\sqrt{-1}$. Therefore, the characteristic function for the normal distribution is $e^{i\mu t-\frac{1}{2}\sigma^2 t^2}$ by replacing t by it.

So, this result of characteristic function of a normal distribution is going to be used later. Therefore, I am introducing finding the mgf as well as characteristic function for a normal distribution with the parameters μ and σ^2, where μ is the mean and σ^2 is the variance.

(Refer Slide Time: 12:05)



Let us go for one more example. Let X be a random variable with mgf is given by $\frac{1}{3}+\frac{1}{2}e^{-t}+\frac{1}{6}e^{t}$. Find the distribution of X this a reverse problem or inverse problem by giving the mgf, we have to find the distribution. Till now we know the distribution we can find the mgf provided it exist. So, here the mgf exist and the mgf is given by $\frac{1}{3}+\frac{1}{2}e^{-t}+\frac{1}{6}e^{t}$. Find the distribution of X, it could be discrete type, it could be continuous type or it could be mixed type.

There is a one important result if the mgf exist it is unique, and the characteristic function always exists and it is unique if the probability generating function exist then it is unique for the random variable. With that concept if the mgf exist it is a unique and it is going to give a unique distribution of the random variable X we know that the mgf is nothing but for any random variable that is $E[e^{tX}]$ if it is a discrete type then it is $\sum e^{tx} P(X=x) \vee \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$.

By seeing the definition and by looking at the mgf you can conclude 1/3, 1/2, 1/6 if you add that is going to be 1 all are with the positive symbol and multiplied by 1 term nothing and another term e^{-t} and the other term is e^{t}. And you see that it is an $E[e^{tX}]$, from that we can conclude the random variable X takes a value 0, -1 and 1 with the probability of X takes a value that is going to be 0 is 1/3; -1 is 1/2 and 1 is 1/6.

You can verify if this is a probability distribution then the mgf is going to be $\frac{1}{3}e^{0t}+\frac{1}{2}e^{-1*t}+\frac{1}{6}e^{t*1}$ that is same as a $\frac{1}{3}+\frac{1}{2}e^{-t}+\frac{1}{6}e^{t}$. Therefore, the distribution of X is this table with the probability mass at the point is 0 is 1/3; -1 is 1/2 and 1 is 1/6. So, this is a discrete type random variable that gives the mgf $\frac{1}{3}+\frac{1}{2}e^{-t}+\frac{1}{6}e^{t}$.

With this, 4 examples we have completed generating functions namely probability generating function, moment generating function and characteristic function. So, with this we are completing the module on moments and inequality starting with mean and variance then higher order moments and moment inequalities, and finally generating functions.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 04
Standard Distributions
Lecture - 18

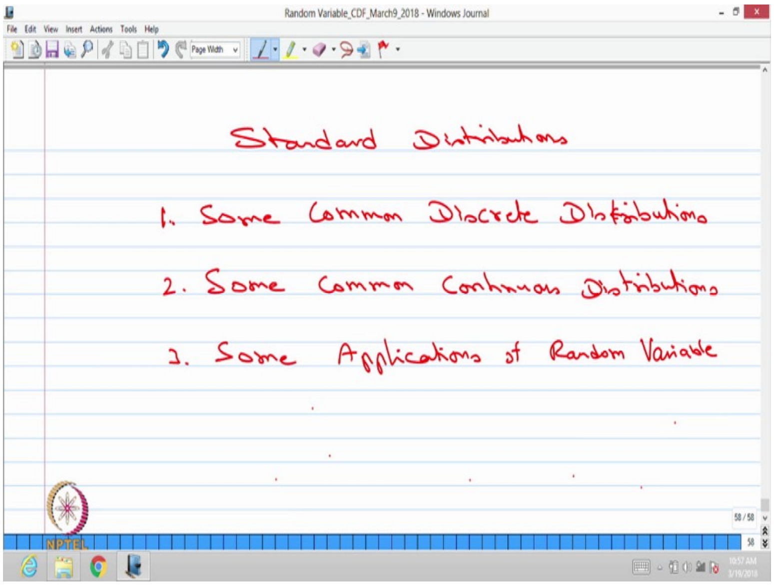
We have already finished 3 models; starting with the basics of probability, then the second model is random variable, and a third model is a moments and inequalities. Now, we are moving to the 4th model that is called as standard distributions.

In this model we are going to discuss various standard distributions which includes standard discrete random variable distributions, and distributions of standard continuous type random variables. Standard means whenever we solve the real-world problems, some distributions comes very often. So, those distribution we call it is a common distribution or standard distributions or frequently we come across the same distributions again and again so those distribution has some name therefore, we call those distribution as a standard distribution or some common distributions.

So, first let me discuss few standard common discrete type random variables whose distributions and also the moments, in particular mean and variance. And similarly, we will discuss later some common continuous type random variables and their distributions, also the mean and variance for those distributions. Then we will discuss some of the problems which has the underlined distributions.

So, in this model we are going to make 3 lectures, that is the title of the model is standard distributions.

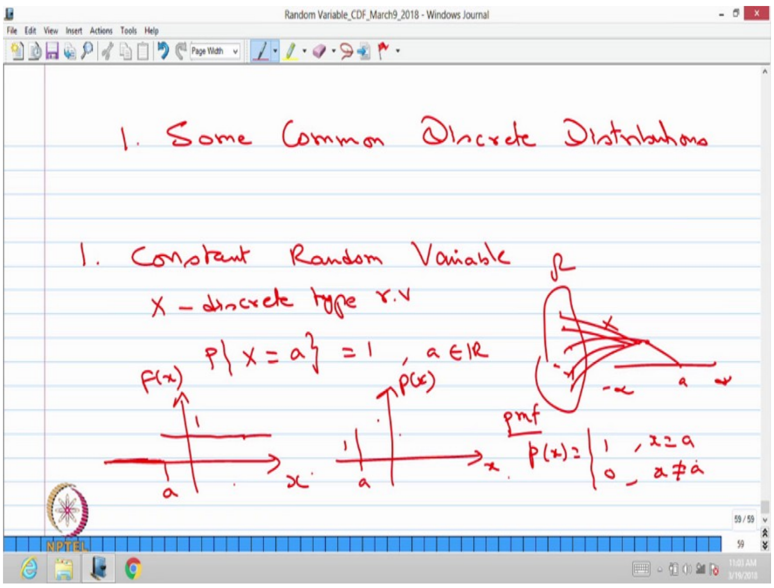
(Refer Slide Time: 02:22)



In this, first we are going to discuss some common discrete distributions, second, we are going to discuss some common continuous distributions, third we are going to discuss some applications of random variable. That means we are going to discuss some problems which are related to these common discrete and continuous distributions.

Let us start with some common discrete distributions.

(Refer Slide Time: 03:40)



The first one, which we are going to discuss that is called constant random variable. A random variable X is said to be a discrete type random variable, X is a discrete type random variable. It is said to be a constant random variable whenever the probability mass function of that random variable X takes one single value a with probability 1; a can be any in the real, that means, the whole unit mass is accumulated at the point X = a.

That means, in a random experiment you may have a sample space with the finite elements or countably infinite or uncountably many elements, the way the real valued function is mapped from Ω to R, it is a many to one function. Therefore, the P{X = a} is nothing but the collection of all possible outcome which gives the value a, i.e., this is Ω it has so many elements.

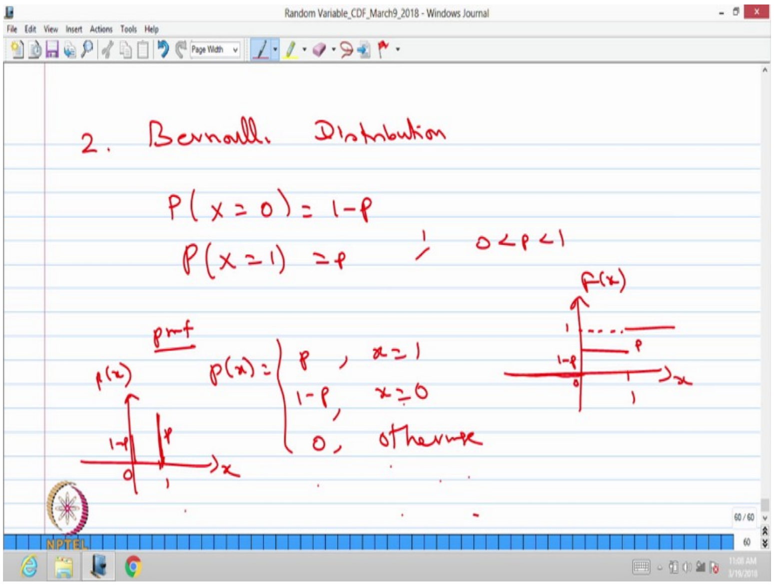
So, the mapping from Ω to R at the point a for all the points for all possible outcomes the mapping is a, where a is belonging to real line; that means, the P{X = a} is nothing but it is a probability of P(Ω). From the Kolmogorov axiomatic definition, we know that the P(Ω) = 1 therefore, it is going to be P{X = a} = 1. That means, all other points if you go for the inverse image whose mass is going to be 0.

You can draw the CDF of this random variable. Suppose a takes some negative value then the CDF is going to be 0 till a at the point it becomes 1, and it will be 1 till ∞ that means, the CDF has only one jump and the jump value is 1. It satisfies all the properties of the CDF, therefore you can conclude this is a CDF of the random variable and the if you draw the probability mass function for this discrete type random variable at X = a it has the value 1 and all other point the mass is 0.

That means in the whole unit mass is accumulated at only one point. Therefore, it is called a constant random variable. That means, any constant can be represented as a random variable with the probability 1 at that point. It is a very important result the CDF has only one jump and the probability mass function is at only one point with the value 1 otherwise it is 0. Therefore, the probability p(x) can be written in an easy way it is equal to 1 when x = a, it is equal to 0 when x ≠a.

So, this is a probability mass function and this is a CDF and this diagram gives the probability mass function in a graphical form.

(Refer Slide Time: 08:18)



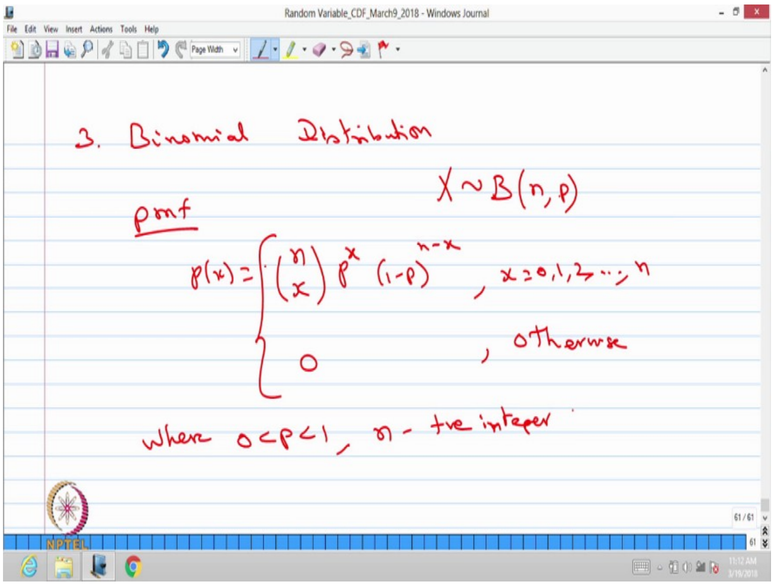
Now, we will go to the second one that is called Bernoulli distribution. A discrete type random variable is said to be Bernoulli distributed random variable when the P{X = 0} = 1 - p and the P{X = 1} = p, where p lies between 0 to 1 otherwise it is 0. That means, the probability mass function p(x), it takes a value p(x) is p when x takes a value 1 and 1 - p when x takes a value 0; otherwise it is 0.

You can draw the CDF this discrete type random variable. x takes a value 0, so till 0, it is 0 at 0 there is a jump of a height 1 - p. At the point 1 it has another jump that jump value is p. Therefore, it touches 1. That means, for this discrete type random variable the CDF has two jumps, one jump is at 0 with the jump value 1 - p and the next jump at the point 1 with the jump value p. Therefore, this is a discrete type random variable and you can draw the probability mass function also the similar way.

So, at x = 0 the height is 1 - p; at x = 1 it has a jump p. It depends on p is going to be less than 1/2 or greater than 1/2 accordingly you will have a 1 - p with the shorter height and the p is taller comparing to the mass at the point is 0. Whenever, the random experiment has two possibilities we call it as a Bernoulli trial. Suppose you treat one possibility as a success and other possibility as a failure then we usually denote the success probability with the probability p and the failure probability with 1 - p then that type of trial is called a Bernoulli trials. A discrete type random variable with the probability mass function of this form we call it as that random variable is Bernoulli distributed. Always the p has to be open interval 0 to 1 if p = 0 or 1, then it becomes a constant.

Therefore, it has to be always the open interval 0 to 1, which gives a Bernoulli distributed random variable X.

(Refer Slide Time: 11:56)



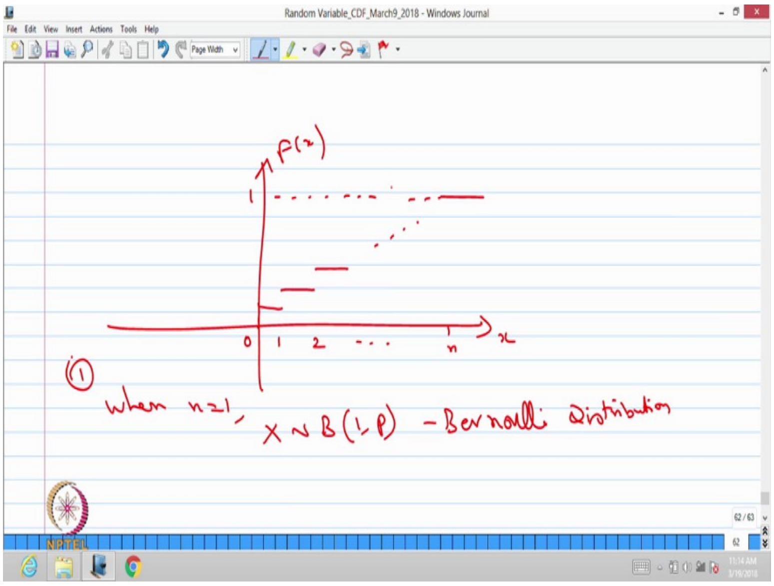
Now, we will go to the third one that is binomial distribution. When we say the given random variable is a binomial distributed, this also discrete type whenever a discrete type random variable whose probability mass function is of the form p(x) ${}^nC_x\, p^x(1-p)^{n-x}$, when x takes a value 0, 1, 2 and so on till n; otherwise it is 0. Then we call or we say the random variable which is a discrete type random variable is binomial distributed random variable.

Since this is a probability mass function, this is always greater or equal to 0 and if you make a summation over x from 0 to n that is going to be 1. And here the p value always lies between 0 to 1 and n is the positive integer, the probability mass function of the form ${}^nC_x\, p^x(1-p)^{n-x}$.

Once you know the value of p and n you know the probability distribution of this random variable. Therefore, we call n and p are the parameters of the binomial distribution. So, we usually write the notation X follows this tilda means of follows capital B; that means, for the binomial distribution and here the parameters are n and p therefore, n comma p. So, whenever we write $X \quad B(n,p)$. That means, the random variable X follows binomial distribution with the parameters n and p. If you supply the value of n which is a positive integer and small p which lies between open interval 0 to 1, then you are known with the probability distribution of this random variable.

The probability mass function has this form. So, you can make out the CDF of binomial distribution has a n +1 jump points and the corresponding jump values are ${}^{n}C_{x}\, p^{x}(1-p)^{n-x}$. So, this is a discrete type random variable. So, you can see the CDF, at x = 0 is a first jump, at x = 1 it has the second jump and x = 2 it has another jump and so on, at x = n it has the last jump and which gives the value 1.

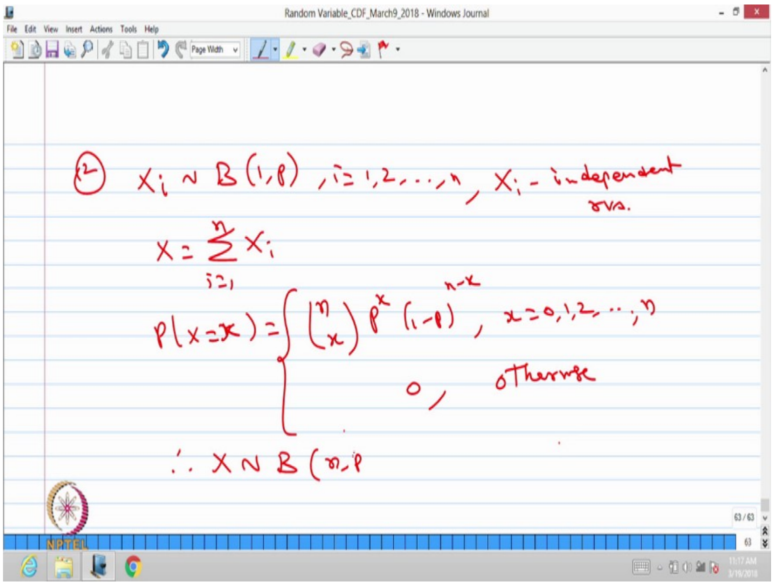
(Refer Slide Time: 14:48)



So, this CDF is a right continuous function with the n + 1 jump points and the jump values are the ${}^{n}C_{x}\, p^{x}(1-p)^{n-x}$. As a special case when n = 1 then the same random variable is going to be say it as a binomial distribution with the parameter 1, p that is nothing but Bernoulli distribution. When n is equal to 1 you will land up a binomial distribution. So, the probability mass function is going to be at x = 0 it is 1 - p; at x =1 the probability mass is a p. Therefore, it is a Bernoulli distributed random variable.

One can create a binomial distribution with the help of Bernoulli distribution, i.e., is whenever you have a random variable, second remark, we can treat this as the first remark.

(Refer Slide Time: 16:43)



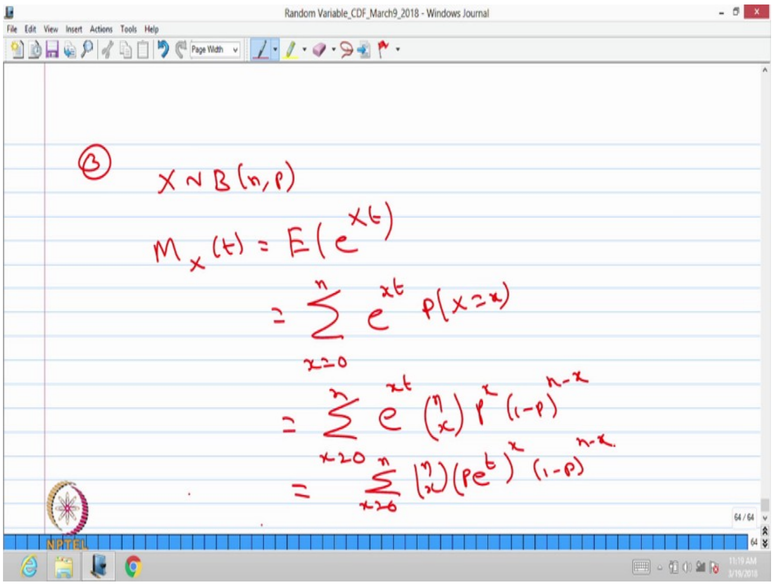
The second remark whenever you have a Bernoulli distributed random variable with the parameter p for n such random variables. If you create a one random variable as a $\sum_{i=1}^{n} X_i$. That means, each random variable is Bernoulli. That means, either it takes a value 0 or 1 and if you create a random variable of a sum of n such Bernoulli distributed random variable. And you can make the one assumption all are independent random variables.

In detailed what is the meaning of those n random variables are independent and what are all the properties going to be satisfied that we will be discuss later, but now you can keep the assumption X_i's are independent random variable; that means, it is a they are mutually independent random variable. Then one can conclude the X = $\sum_{i=1}^{n} X_i$, so now, the possible values of x is going to be 0 to n, because each Xi takes a value 0 or 1 therefore, sum of n such Bernoulli distributed random variable the possible values are going to be x. And you can get the probability mass function it is X takes a value x that is going to be ${}^{n}C_x\, p^x (1-p)^{n-x}$ when x takes a value 0 or 1 or 2 so on till n.

That means, suppose you say that X takes a value x; that means, out of n such Bernoulli trials you are getting x with a probability p and the remaining n - x you got the failure that is $(1-p)^{n-x}$ with the possibilities of ${}^{n}C_x$ ways. Therefore, P{X = x}= ${}^{n}C_x\, p^x (1-p)^{n-x}$. So, this is going to be a probability mass function of binomial distribution. Therefore, one can conclude if you have a n independent Bernoulli distributed random variables with the same parameter p for the Bernoulli distribution then the sum of n independent Bernoulli distributed random variable becomes binomial distribution with a parameters n and p.

So, you can say X follows a binomial distribution with the parameters n, p. So, this is a way one can create the binomial distribution with the help of a Bernoulli distribution. As a third remark we can go for finding what is the mgf of a binomial distribution.

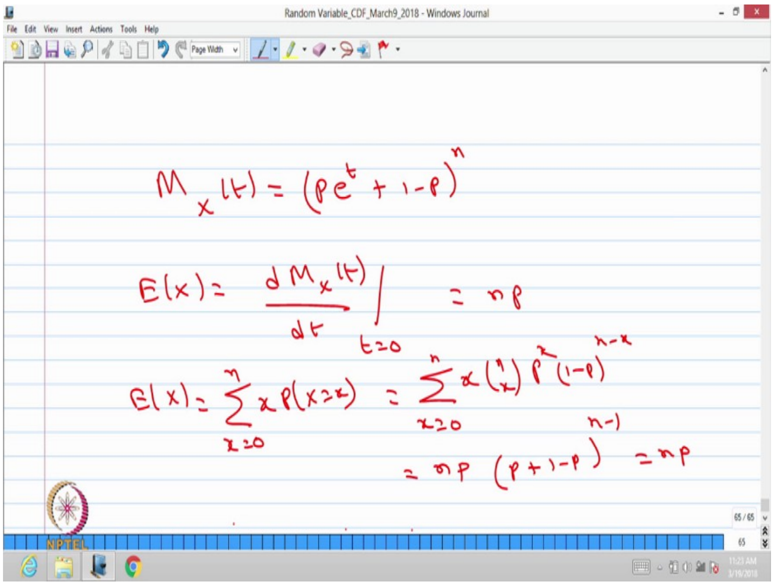
(Refer Slide Time: 19:54)



Suppose X follows a binomial distribution the parameter n and p, we can find mgf of binomial distribution that is nothing but $E\left[e^{tX}\right]$ that is same as since it is a discrete type random variable. So, it is going to be $\sum_{x=0}^{n} e^{xt} P\{X=x\}$. So, you substitute all the values that is $\sum_{x=0}^{n} e^{xt} P\{X=x\}$ and their P{X = x} is a ${}^{n}C_x p^x (1-p)^{n-x}$.

Now, also you do not need to expand and simplify and so on you can keep p^x and e^{xt} together. Therefore, this is nothing but summation x is equal to $\sum_{x=0}^{n} {}^{n}C_x (pe^t)^x (1-p)^{n-x}$.

(Refer Slide Time: 21:23)



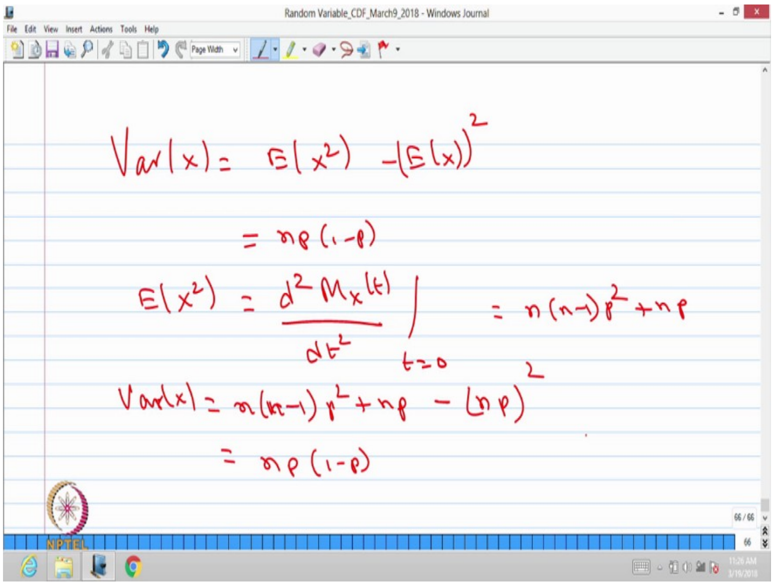
Therefore, the mgf of a binomial distribution is $(pe^t+1-p)^n$ using a binomial summation you can get a mgf of binomial distribution is $(pe^t+1-p)^n$.

What is the use of mgf? Once you know the mgf you can get all the movements by successive derivative. If you want to find out only mean you can use the mean definition and you can get the mean. Suppose you need many moments then better you can find the mgf first then successive derivative and I have already explained from the moment generating function how to get the moments of order n. So, we can use that and find out the first order moment, second order moment, any nth order moment.

So, in particular if you want to find out the mean of this random variable; if you want to find out the mean of this random variable, that is nothing but if you use the derivative one derivative of mgf with respect to t then substitute t equal to 0 will give the mean of the binomial distributed random variable. So, by doing the derivative and so on you can get the value np. This is a one way of finding mean of the binomial distribution or you can find out from the scratch that is a $\sum_{x=0}^{n} x P(X=x)$. So, that is same as a $\sum_{x=0}^{n} x\, {}^{n}C_x\, p^x (1-p)^{n-x}$.

You can cancel x with ${}^{n}C_x$ one term therefore, you will get factorial $\frac{n!}{(n-x)!(x-1)!}$. So, now, you can keep one n and one p outside. Therefore, the simplified quantity becomes ${}^{n-1}C_{x-1}\, p^{x-1}(1-p)^{n-x}$.So, that is nothing but, that is nothing but $(p+1-p)^{n-1}$and you know the that value is 1. Therefore, this is going to be np. So, there are two ways we can find out the expectation either from the scratch by the definition method or by the mgf method.

(Refer Slide Time: 24:32)

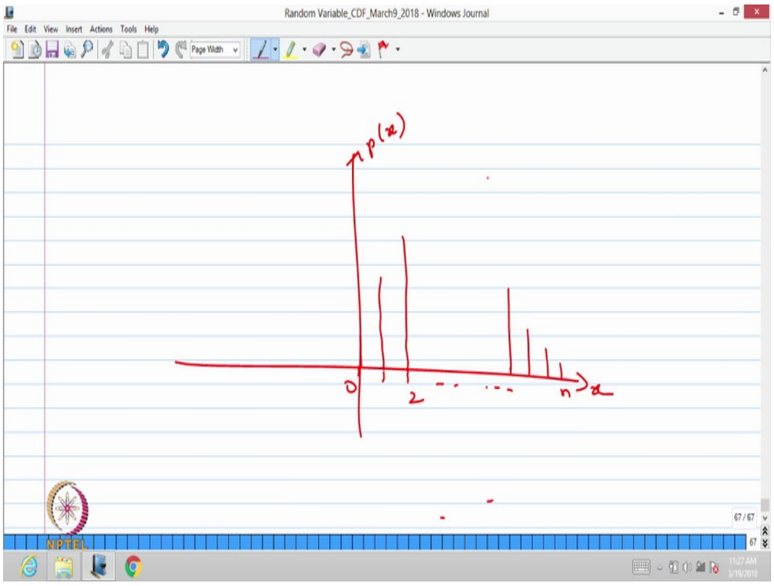


The same way one can compute the variance. So, the variance of X that is going to be either by using the definition, for that you have to find out the $E[X^2]$ and already you know the value of E[X] you can substitute and get the value or you can do the second derivative of the mgf then through that you can get the variance. So, if you do the little simplification you can get the variance of X is going to be np(1-p).

For that you can do the other method that is you can find $E[X^2]$ that is nothing but second derivative of mgf of X with respect to t twice, then substitute a t equal to 0 that some simplification will give the value that is $n(n-1)p^2+np$.

So, once you know the $E[X^2]$ and the substitute in the variance formula that is is $n(n-1)p^2+np-(np)^2$. So, you simplify that you will get the value that is np(1-p). So, this is a way one can get the variance of X for the binomial distributed random variable.

(Refer Slide Time: 26:32)



There is another observation if you draw the probability mass function of a binomial distribution. For x is equal to 0 you will have some height, for x is equal to 1 you will have another height, for x is equal to 2 you will have a more height and so on then it will be keep decreasing.

So, this is based on the value of p you will have increasing for the different values of x, then decrease. For p is equal to less than 1/2 you will have a more increase in the first 1/2 then it goes decrease. For p is greater than 1/2 you will have less heights in the beginning and more heights in the second half. If p is equal to 1/2 then you will have a very symmetric of increasing and decreasing over 0 to n, and for n is equal to even you will have a two heights for n is equal to odd you will have only 1 height. That is the way the values are going to be keep increasing then it will decrease.

I have just to drawn one diagram for probability mass function for some n and some p. The importance of the CDF graph is from the finite number of data and if you draw the cumulative distribution graph, if that graph is same as the CDF of binomial distribution then one can conclude the data follows a binomial distribution. Not only from the DF one can do the observation from the probability mass function also. That means, if you draw the histogram of the values of the data over 0 to n, and the probability the histogram look like the probability mass function at some point if it is almost same then you can conclude the data also follows binomial distribution. Therefore, one should always know the CDF and the probability mass function for a discrete type random variable.

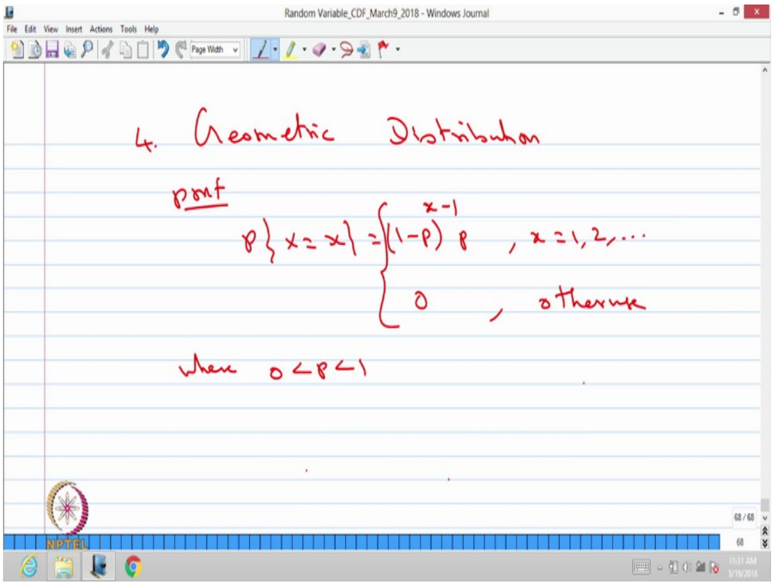
Similarly, for a continuous type random variable one should know the CDF as well as the probability density function of a continuous type random variable. So, both the things are very useful when you have a data in the first then you can identify what could be the distribution.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 19

So, till now we have discussed a constant random variable and Bernoulli random variable, binomial random variable or binomial distribution.

(Refer Slide Time: 00:15)



Now, we are moving into fourth one, that is geometric distribution. This is also common discrete type distribution, whenever a random variable which is discrete type random variable whose probability mass function is of the form, P{X = x} = $(1-p)^{x-1}p$, where x takes the value 1, 2 and so on, otherwise, it is 0.

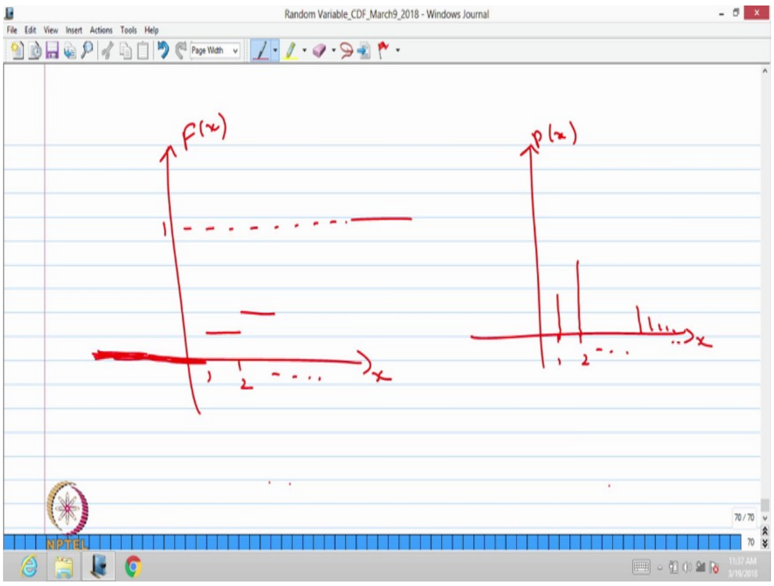
Then we call this random variable X is geometric distributed random variable here also, the p lies between 0 to 1. There is the connection between Bernoulli distribution with the geometric distribution, that connection is whenever you have a Bernoulli trials, the occurrence of the first trial in which you get the success that follows geometric distribution.

Suppose you have a random experiment with infinitely many Bernoulli trials in it, and each Bernoulli trial has a random variable, which is a Bernoulli distributed random variable with the probability of success p. This X is nothing but the trial in which you are getting the first success, that probability is you are not getting the success x - 1 times and the xth trial you are getting the first success. Therefore, it is $(1-p)^{x-1}p$ all are consecutive, and all the Bernoulli trials are independent.

So, whenever you have n independent or sequence of independent Bernoulli trials, the first success in the nth trial that becomes the geometric distribution. So, the difference between Bernoulli, binomial and geometric, the Bernoulli distribution has only 2 jumps the CDF has only 2 jumps, and the binomial distribution has only n + 1 jumps, the geometric distribution has a countably infinite jumps.

So, this is also discrete type random variable. So, let us discuss the CDF and the probability mass function of this random variable geometrically distributed.

(Refer Slide Time: 03:01)



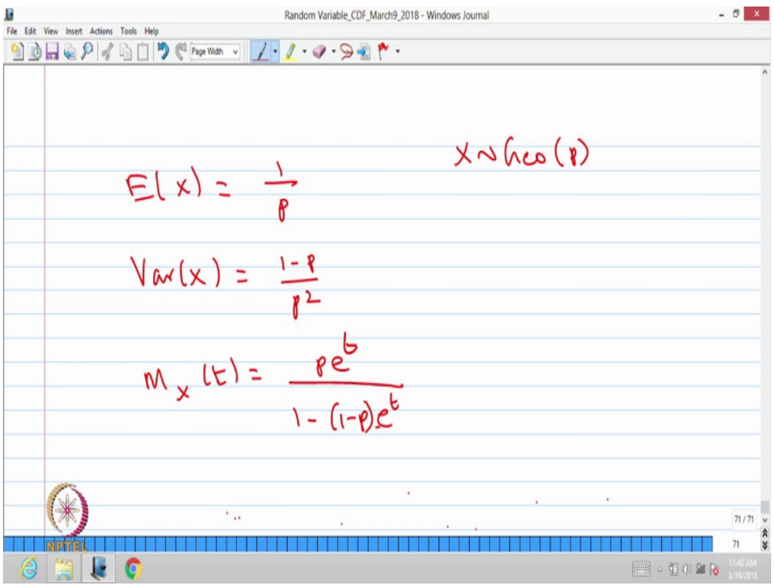
So, it has the jump points 1 2 and so on, therefore, at till x is equal to 1, it has the value 0 at x is equal to 1, it has a first jump at x is equal to 2, it has a second jump and so on. At infinity it touches 1, at infinity it touches 1.

So that means, of the CDF has a countably infinite jumps with the jump points 1, 2 and so on, and jump values are $(1-p)^{x-1}p$. And if you draw the probability mass function, at x is equal to 1, it has some height, at x is equal to 2, it has some other height and so on. Then it will be keep decreasing, then it land up at countably infinite points if you had all the heights that is going to be 1.

So, this is a probability mass function, and this is the CDF of the geometric distribution. The way I have explained through the data, suppose you have a data with the cumulative distribution, it keeps increasing at countably infinite number of points, land up to be some finite value. Or the probability or the histogram of the data that has some heights keep increasing and going down, and it has a countably infinite points in which is it has these values, then you can conclude the data could follows a geometric distribution.

So, in the statistics, we get this type of graphs first from the data, in the probability theory course, we started with the probability mass function then the CDF and so on, in a theoretical way we study whereas in the statistics we start from the data, then we conclude what could be the distribution of those data. So, one can discuss the mean variance and the MGF for this geometric distribution also.

(Refer Slide Time: 05:33)

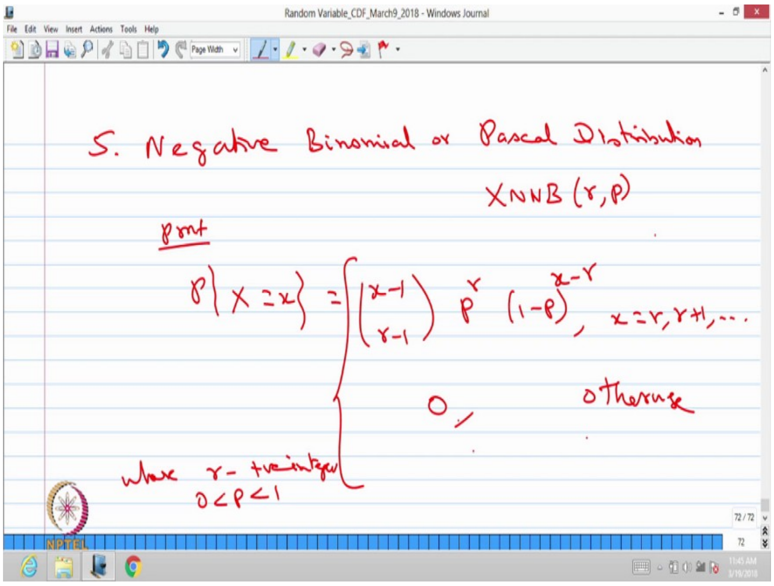


So, the mean for geometric distribution is going to be $\frac{1}{p}$, and the variance of geometric distribution is going to be $\frac{1-p}{p^2}$, and one can get MGF of geometric distribution; that is $\frac{pe^t}{1-(1-p)e^t}$.

So, in notation we use X follows geometric with the parameter p, when we say X $Geo(p)$. That means, this is a geometric distribution with the parameter p, whose probability mass function is $(1-p)^{x-1}p$ where x takes a value one and so on. You can always create another random variable in which the probability mass function starts from 0 onwards instead of one onwards, then that random variable is call it as a modified geometric distribution. In the real-world problem sometimes you come across the possible values are 1, 2 and so on, or sometimes at the values start from 0, 1, 2 and so on.

So, you can use the correct probability mass function so that summation is one. I am not going for the derivation the same derivation what we have done it for the binomial you can use the same thing.

(Refer Slide Time: 07:20)

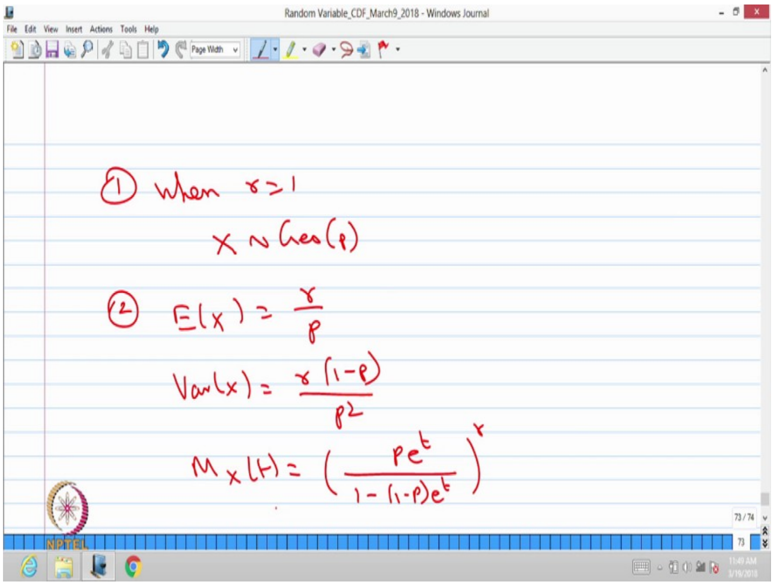


Now, we will move into the fifth one, that is negative binomial. There is another name for this distribution, that is called Pascal distribution. A discrete type random variable is said to be a negative binomial or Pascal distribution, whenever the probability mass function is of the form $P(X=x)={}^{x-1}C_{r-1}p^r(1-p)^{x-r}$; where x takes the value r, r + 1 and so on, otherwise it is 0.

Here r is positive integer, and p lies between 0 to 1. That means, whenever you supply the value of r and p, you know the distribution of this random variable. We use a notation X follows a negative binomial NB, with the parameters r, p; this is also related to the Bernoulli distributed random variable in the form of capital X denotes in the x trail, we are getting first time rth success follows negative binomial, whenever each trails are Bernoulli and they are independent.

So, whenever you have independent Bernoulli trials, obtaining first time rth success that follows a negative binomial distribution with the probability of success is p, and the probability of failure is 1 - p.

(Refer Slide Time: 09:49)



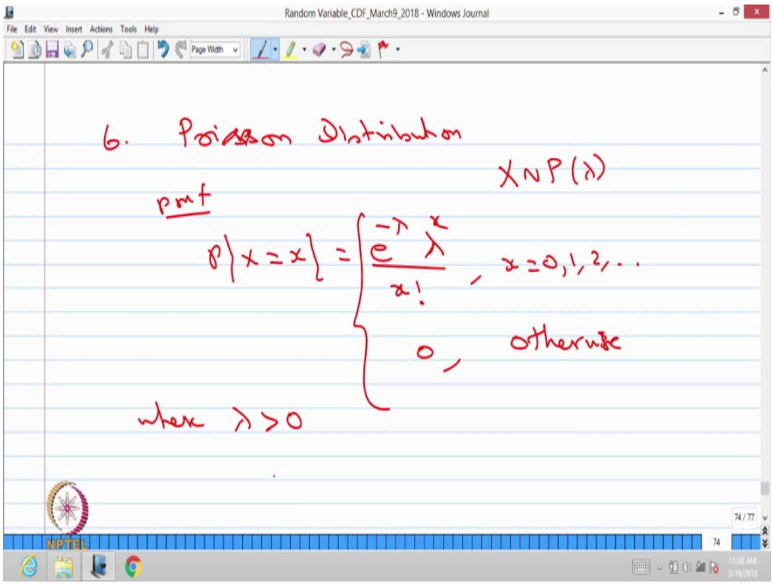
When r = 1 the same random variable X follows a geometric distribution with a parameter p. When the rth success that is when it is a first success in the xth trail, then that follows a geometric distribution. Therefore, geometric distribution is a special case of negative binomial or Pascal distribution with the parameter r = 1.

The probability mass function can be visualized once you are getting r - 1 success out of x -1 bernoulli trials, that follows a binomial distribution followed by the rth success; that means, ${}^{x-1}C_{r-1}\, p^{r-1}(1-p)^{x-1-r+1}$ that can be treated as r - 1 success getting out of x - 1 trials, which follows a binomial distribution multiplied by the rth success getting in the xth trail. Therefore, x can be r. That means, you may get the rth success in the xth trail itself, or you may get rth success in (r +1)th trail and so on.

So, that is a interpretation of a the probability mass function P{X = x} where x takes the value r, r + 1 and so on. So, this is also discrete type random variable, and CDF has countably infinite jumps. So, I am not going to draw the CDF of a negative binomial, but one can visualize the CDF has a countably infinite jumps of this discrete type random variable.

For this random variable also, one can find the mean variance and so on, the mean of this random variable is nothing but $\frac{r}{p}$, and variance of this negative binomial that is $\frac{r(1-p)}{p^2}$. You can verify when you put r = 1 it has to be a same as of geometric distribution. And the MGF of negative binomial or Pascal distribution is $\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$, when r = 1 that is same as the MGF of geometric distribution. I am not going for the derivation, but one can drive and you can get the same answer.

(Refer Slide Time: 13:07)

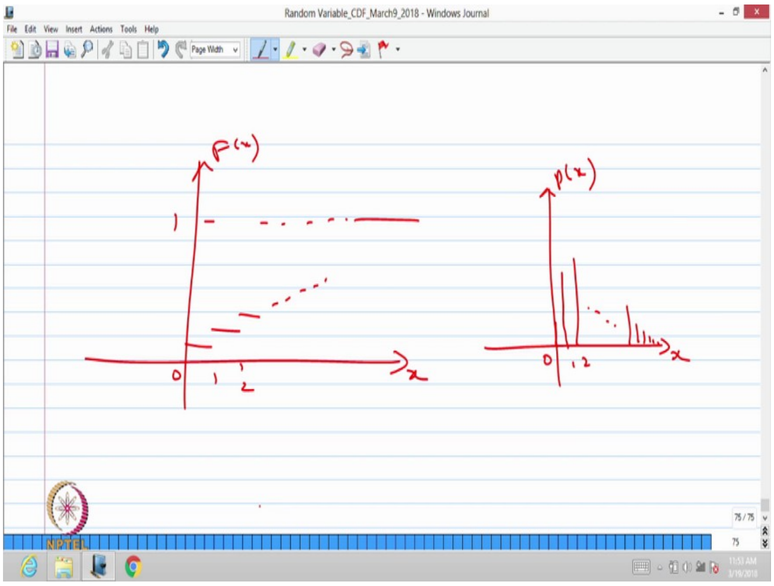


So, the next distribution is Poisson distribution. It is a very important distribution, because this connects probability with a stochastic process in the different level. A discrete type random variable is said to be a Poisson distributed random variable, if the probability mass function of this random variable is going to be of the form $\frac{e^{-\lambda}\lambda^x}{x!}$ where x takes value 0, 1, 2 and so on, otherwise 0. Here the λ has to be strictly greater than 0, it is a constant.

So, whenever any discrete type random variable whose probability mass function of this form, $\frac{e^{-\lambda}\lambda^x}{x!}$; where x takes a value 0, 1, 2 and so on; otherwise the probability mass function is must be 0, then that random variable is call it as a Poisson distributed random variable.

You can verify in this probability mass function, this is always greater or equal to 0, and if you make a summation over x starting from 0 to infinity, $e^{-\lambda}$ is out and $e^{-\lambda}$ is outside then the summation, and that summation quantity becomes e^{λ}. And since the $\lambda > 0$. Therefore, this quantity is going to be 1. Therefore, this is a probability mass function.

(Refer Slide Time: 14:57)



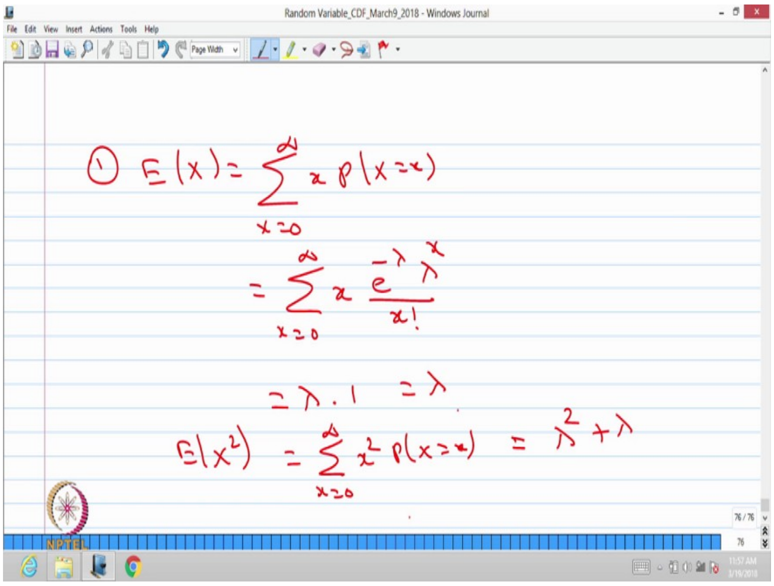
One can draw the CDF for this Poisson distributed, x takes value 0 it as a jump at x equal to 1, it has another jump x equal 2, it has another jump and so on countably infinity jumps. It touches one at infinity. Similarly, if you draw the probability mass function of Poisson distribution 0, it has some height, and one it has another height, 2 it may have another height and so on, and it will be keep decreasing at countably infinite number of points.

So, this is a one diagram in which the λ value is; so, that it is keep increasing decreasing, or there is a possibility it may have at x = 0, it to have a tallest then it may keep going down. And whatever the possibility the summation of probability mass at the countably infinite number of points it is going to be 1. So, the same conclusion, if the data has cumulative distribution graph, or the histogram look like this CDF form or probability mass function form, then one can concluded that data follows a Poisson distribution.

And there is another relation with Poisson distribution with binomial and Bernoulli. If you have a n independent Bernoulli distributed random variable, that summation becomes a binomial, when the n becomes very large, and the p probability of success is very small one can prove the limiting case of a n tends to infinity, and p is very small, then the binomial distribution will tends to Poisson distribution. For binomial distribution the n is always finite quantity, and the p is probability of success in any one Bernoulli trial, and all such n Bernoulli trails are with the probability of success p same, as and all are independent. Therefore, you are getting the binomial distribution.

But for larger n also when p is very small, then the limiting case of the binomial distribution goes to Poisson distribution. Therefore, you have a countably infinite jumps in the CDF, one can visualize the limiting case of a binomial distribution is Poisson distribution. So, that is the connection between Bernoulli distribution, binomial distribution and Poisson distribution.

(Refer Slide Time: 17:45)

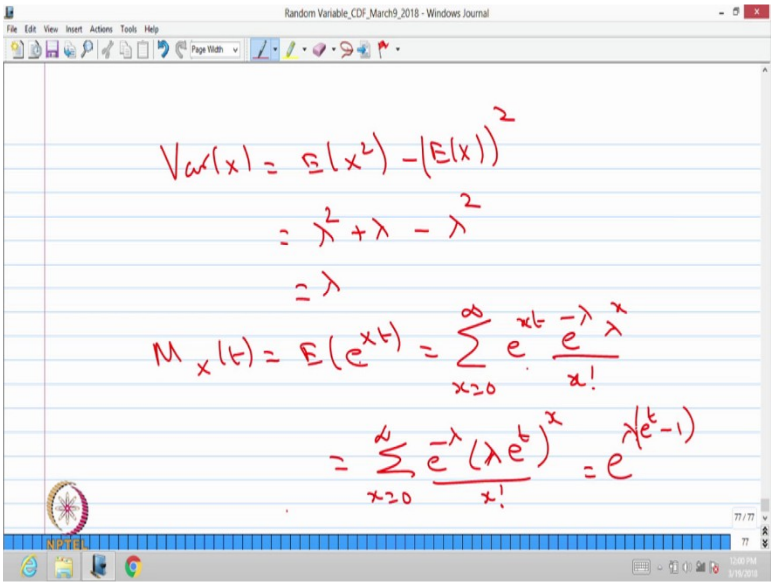


One can get the mean for Poisson distribution, that is nothing but $\sum_{x=0}^{\infty} xP\{X=x\}$. This is nothing but x times P{X = x} = $\frac{e^{-\lambda}\lambda^x}{x!}$ when x takes a value 0 to infinity. x! and x cancel; so, you will get (x - 1)!, you can take one λ outside, the remaining quantity becomes one.

Therefore, the mean is going to be λ. Similarly one can find $E[X^2]$; $E[X^2]$ that is nothing but $\sum_{x=0}^{\infty} x^2 P\{X=x\}$, the similar way one can compute. So, you can get the answer that is a $\lambda^2+\lambda$.

If you do the little simplification by substituting P{X = x} = $\sum_{x=0}^{\infty} xP\{X=x\}$ to the simplification, you will get $\lambda^2+\lambda$.

(Refer Slide Time: 19:06)



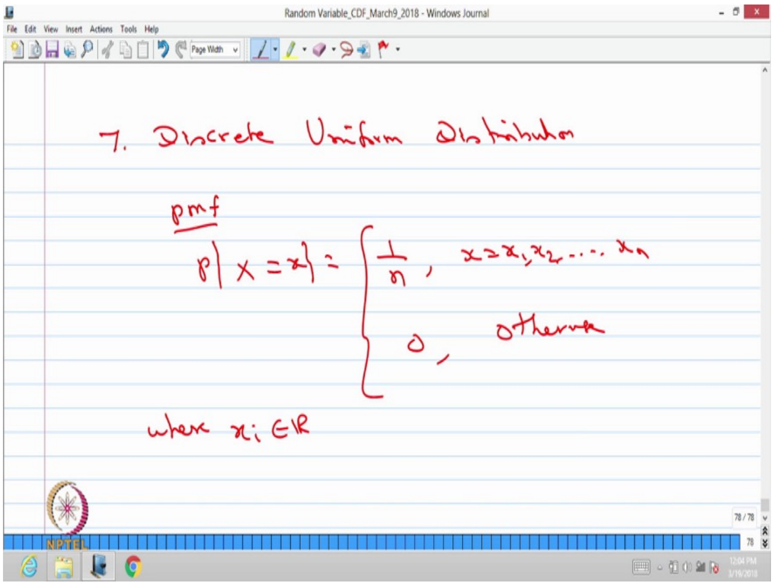
Therefore, you can get the variance of X; that is E$[X^2]-E[X]^2$. And E$[X^2]=$¿ $\lambda^2+\lambda$ and E[X] = λ. Therefore, that is λ^2. So, simplify you will get λ. So, these are very important result the mean and variance of the random variable is same which is λ. So, here λ is a parameter, because once you know the value of λ you are known with the distribution. Therefore, we use a notation X follows the capital P with the parameter λ; that means, this is a Poisson distributed random variable with a parameter λ.

So, once you specify the value of λ, you are known with the distribution of this random variable. So, in Poisson distribution the important result is a mean and variance are same which is same as the parameter. Similarly, one can compute the moment generating function, because through this you can get all the moments of order n. So, if you do the MGF calculation, it is $E[e^{tX}]$ then that is same as a $\sum_{x=0}^{\infty} e^{xt} P\{X=x\}$, that is $\frac{e^{-\lambda}\lambda^x}{x!}$, where x takes a value from 0 to infinity. You can keep λ and e power, sorry, you can keep $\lambda^x \wedge e^{xt}$ together. So, therefore, this is nothing but $\sum_{x=0}^{\infty} \frac{e^{-\lambda}(\lambda e^t)^x}{x!}$.

So, if you do the little simplification, you can get the answer that is same as $e^{\lambda e^{t}-\lambda}$ so, that I can write it as minus 1. So, it is basically $e^{-\lambda}$ is outside. So, this summation is nothing but $e^{\lambda(e^{t}-1)}$. Therefore, it is an exponential of $\lambda(e^{t}-1)$ that is a MGF. So, from the MGF you can always get the; by derivative you can get the E[X], $E[X^2]$, then through that you can get the variance also.

So, since it is a discrete type random variable, you can go for probability generating function; so, even though I have not explained how to find out the probability generating function for all the distribution. So, starting from the Bernoulli binomial geometric Poisson and negative binomial, all this distribution because it is a discrete type and it takes a positive integer values. Therefore, one can go for finding the probability generating function of this stand common standard discrete type distributions.

(Refer Slide Time: 22:39)



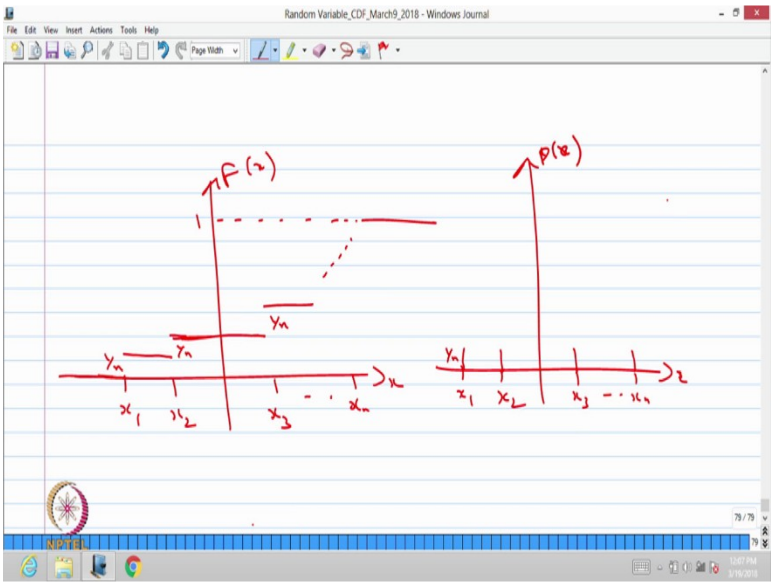
The next one that is number 7, that is a discrete uniform distribution, a discrete type random variable is said to be a discrete uniform distribution, whenever the probability mass function is of the form P{X = x}, that takes a value 1/n, when x takes a value x_1, x_2 and so on x_n, otherwise 0. Here the x_i's are the real numbers.

So, it can be any n points, the probability mass function are those n points is same, which is same as 1/n, and all other points the probability mass function is 0. Such a discrete type random variable is called a discrete uniform distribution, why the word uniform because the probability mass function is same for all such n points.

So, all such n point has to be distinct all should be different distinct n real values in which the probability mass function is a same. And since it is a probability mass function the summation has to be 1. Therefore, the for a uniform distribution the probability mass function is 1/n, then only the summation is going to be one and all are going to greater than equal to 0, one we at those n points.

So, such as discrete type random variable is called a discrete uniform distribution.

(Refer Slide Time: 24:32)



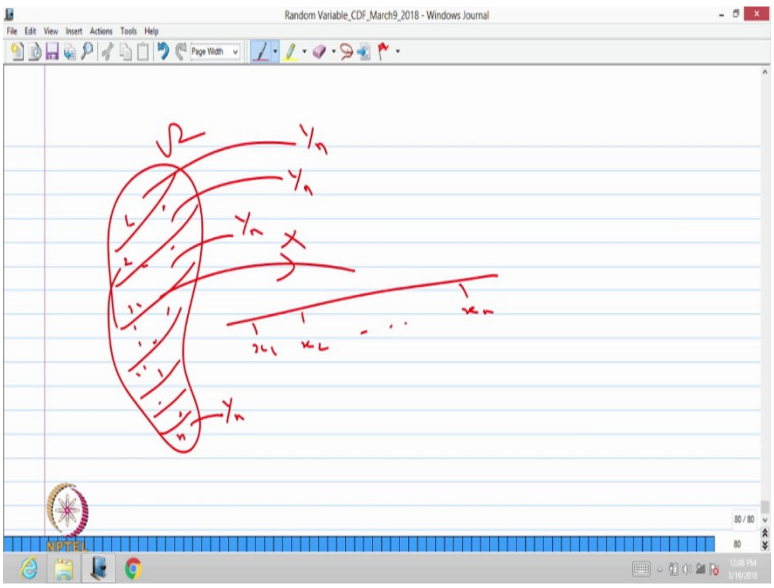
Let me draw the sample CDF, suppose x_1 is here, x_2 is here, x_3 is here, x_n is somewhere here. It need not be equi distance, it can be any n district points. And the CDF I have just list out x_1 is the first value, and x_2 is greater than value x_1, x_3 is a greater than x_2 and so on. So, the CDF is 0 till x_1, at x_1 it has a jump and jump values 1/n.

So, this height is 1/n till x_2 the value is 1/n at $x_2$it as a next jump. Till x_3 it is going to be the same value and this jump is 1/n. And x_3 it has the next jump and this jump is 1/n. All the jump heights are same, at the point x_n it has a last jump, and it becomes one.

If you see the CDF for any discrete type random variable, which has only n jumps, and all the jump values are same, then that random variable is a discrete uniform distributed random variable. You can relate this CDF with the earlier random variable CDF. It may have a one jump or n+1 jumps or countably infinite jumps, but the jump values are different at different point. Whereas, here it is fixed always n jump points and always the n jump values are same which value is 1/n that value is 1/n, then that CDF is corresponding to the CDF of a discrete uniform distribution.

So, if you draw the probability mass function at those n points, the heights are going to be 1/n same heights. That means, if you have a data in which if you draw the histogram. And all the histogram heights are same, with the n number of points or with the way you made groups and so on, you can think of that could comes from the discrete uniform distribution. Or the data if you draw the CDF cumulative distribution and it has same jump heights, and only finite number of jumps then it is a discrete uniform distribution.

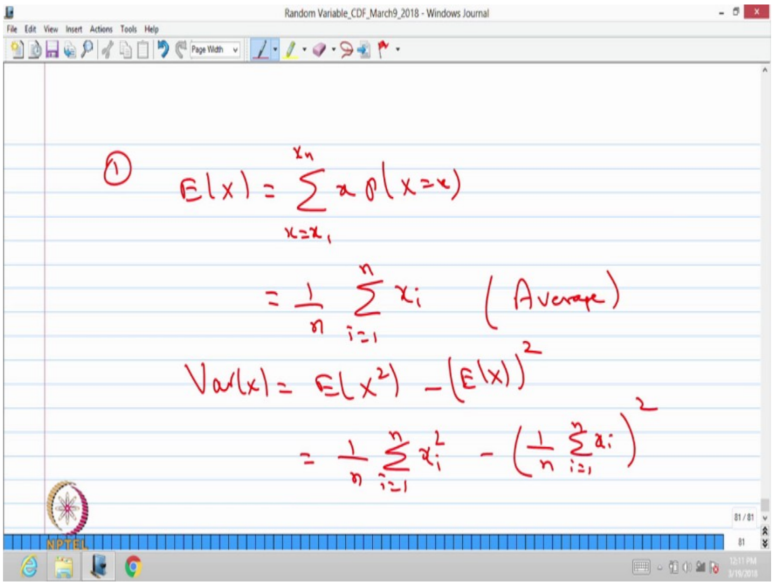
(Refer Slide Time: 27:39)



That means, you can interpret in other way also, you have Ω, Ω consist of finite or countably infinite, or uncountably many samples in it the way the mapping goes maps into $x_1, x_2, \ldots, x_n$, such a way, partitioning Ω into n pieces and each one has a mass 1/n. Each one is attached with one point whose probability mass function is 1/n.

So, you partition is the first partition second partition so on. This is a nth partition whose mass is 1/n. That means, that random variable is discrete uniform distribution.

(Refer Slide Time: 28:38)

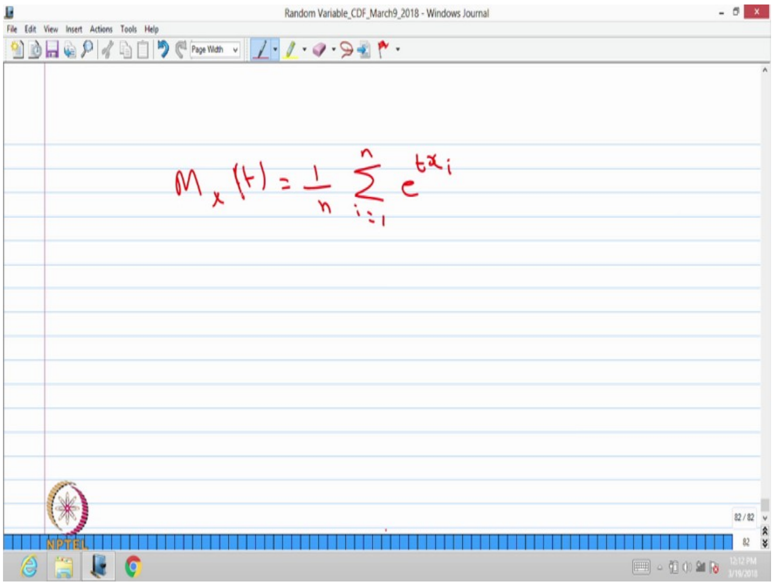


Since it is a discrete uniform distribution, you can find the mean is nothing but $\sum_{x=x_1}^{x_n} xP\{X=x\}$. And the P{X = x} = 1/n. Therefore, the 1/n can be taken out side, and you add all the values i is equal to 1 to n.

So, the mean is nothing but; mean or expectation is nothing but some of those values multiplied by 1/n. That is nothing but it is average. So, whenever the random variable is of the discrete uniform, the mean or expectation which is same as the average, we can go for finding the variance of X that is $E[X^2]-E[X]^2$. So, first you compute the E$[X^2]$, then you substitute in this formula then you can get the variance.

So, since the probability mass function at those points is 1/n. Therefore, this is going to be $\frac{1}{n}\sum_{i=1}^{n} x_i^2$ - $\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$,if you do the simplification you can get it.

(Refer Slide Time: 30:27)



And the similar way you can get the MGF also, the MGF is going to be MGF of X, that is same as $\frac{1}{n}\sum_{i=1}^{n} e^{tx_i}$, because of the P{X = x} = 1/n, that can be taken out.

So, the s$\frac{1}{n}\sum_{i=1}^{n} e^{tx_i}$ may give the moment generating function for the discrete uniform distributed random variable. With this we are completing some common discrete distributions, starting from a constant, Bernoulli, binomial, geometric, negative binomial, Poisson and discrete uniform distributions.

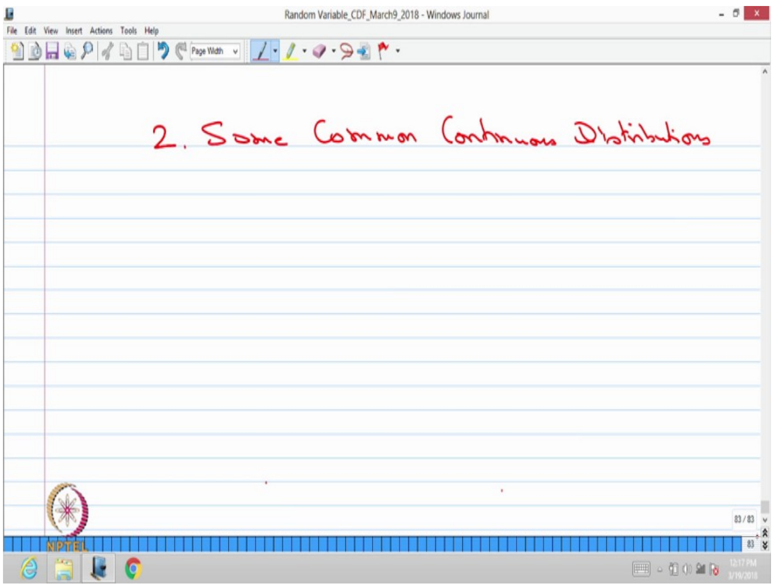
So, there are 7 distributions we have discussed they are all called common discrete type random variable; whose probability mass function and CDF, mean, variance, MGFs are discussed.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 20

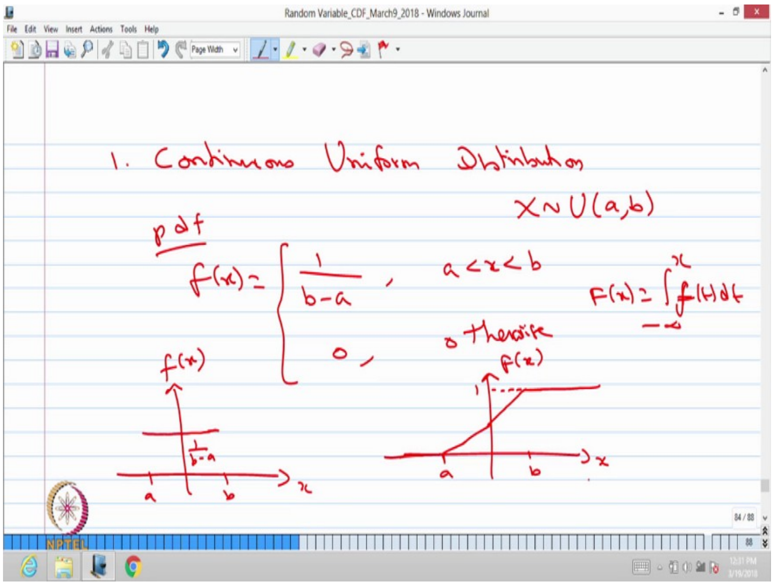
So, in the last class we have discussed some common discrete distributions in Standard Distributions module. Now, we will move into some common continuous distributions. That means some continuous type random variable occur more frequently whenever we come across different problems in the probability. Therefore, we introduce a word called common continuous distributions. That means, I am going to discuss a few or some important continuous type random variable whose the probability mass function, then what is the CDF of those continuous type random variable, then what is the mean, variance, mgf it if it exist, then characteristic function and so on.

(Refer Slide Time: 01:00)



So, the title of these a lecture is, some common continuous distributions. In these we are going to discuss a few continuous type random variables which occur very frequently in the different problems of probability.

(Refer Slide Time: 01:23)



The first one that is continuous uniform distribution. A random variable X which is of a continuous type whose a probability density function is of the form f(x) = $\frac{1}{b-a}$, where x lies between a to b; otherwise it is 0 where a, b is in the real line. So, whenever any continuous type random variable whose probability density function is of the form f(x) = $\frac{1}{b-a}$. That's a probability density function which is greater than 0, otherwise it is 0. Then we call it is a continuous uniform distribution. You can draw the probability density function to visualize.

Suppose, a is a negative and b is a positive and the probability density function is between the interval a to b is $\frac{1}{b-a}$. So, you compute $\frac{1}{b-a}$. So, you draw, so this height you think of $\frac{1}{b-a}$ otherwise; that means, from - ∞ to a the probability density function value 0. Similarly, from b to ∞ the probability density value is 0, only between the interval a to b the values $\frac{1}{b-a}$.

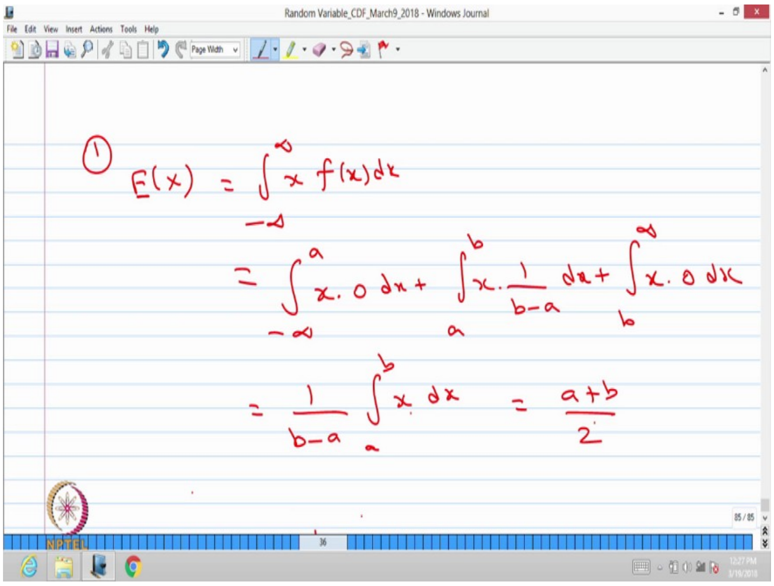
Why it is called a uniform, because the probability density is a uniform it is a constant it is not a function of x between the interval a to b since the density is a constant between the interval, and since it is the probability density function the integration has to be one therefore, that values $\frac{1}{b-a}$. Therefore, this continuous type random variable is called a uniform distributed random variable. Earlier we have discussed discrete a uniform distribution; that means, that is at the probability mass function is uniform are same in all n distinct points. Here it is a continuous uniform distribution; that means, the density function between the interval is a constant which is same as one divided by length of the interval. Therefore, it is called a continuous uniform distribution.

If you draw the CDF of this, the CDF of continuous uniform distribution, till a the value is going to be 0 at a till b you find out F(x) = $\int_{-\infty}^{x} f(t)dt$ as probability density function if you substitute and find out the integration from - ∞ to a the probability density function is 0. Therefore, the value 0. Whereas, from a to b it is you have $\int_{a}^{x} \frac{1}{b-a} dt$. So, if you integrate you will get the slanting line till b, at the point b it becomes 1.

So, this type of random variable is call it as a continuous uniform distribution, the way I have explain to through the data of the CDF and the probability mass function of a discrete type random variable, the same way by seeing the CDF and the probability density function one can conclude if the data has a CDF is 0 till some point after that it is a slanting line. And at some point, it becomes some constant value then it remains constant you can normalize it make it is a similar to the CDF of a continuous type then one can conclude this data follows a continuous uniform distribution.

Similarly, if you draw the histogram and the histogram is between some interval it is a constant and all other value it is 0, then you can visualize that data follows a continuous type uniform distribution. So, here a and b are constant, lies between - ∞ to ∞ and the probability density function is $\frac{1}{b-a}$ that is very important.

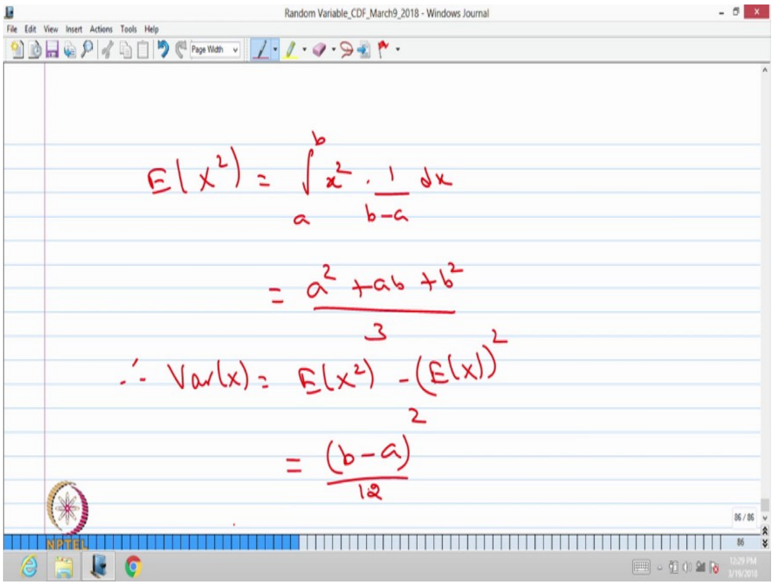
(Refer Slide Time: 06:30)



One can go for finding mean, variance and mgf characteristic function and so on. For this standard continuous type random variable. So, the mean for the random variable, E[X] that is since it is a continuous type it is a $\int_{-\infty}^{\infty} xf(x)dx$. We are doing this calculation with the assumption that the mean exists; the assumption that in absolute sensor this integration is a finite quantity without absolute, we are finding the integration that is the value of an expectation. This is same as $\int_{-\infty}^{a} x.0\,dx+¿\int_{a}^{b} x\frac{1}{b-a}dx+\int_{b}^{\infty} x.0\,dx¿$. The probability density function is 0 between - ∞ to a as well as b to ∞. Therefore, this is nothing but $\frac{1}{b-a}\int_{a}^{b} xdx$.

So, one can simplify and you can get the answer for mean and variance. This is same as $\frac{a+b}{2}$. The mean of continuous uniform distribution is, if the interval is a to b, then addition of the time interval divided by 2 that is going to be the mean. That intuitively also one can say if it is a uniform then adding those endpoints divided by 2 that is a going to be the middle point. So, whenever it is a uniform distribution of the continuous type $\frac{a+b}{2}$ that is going to the mean that is same as average.

(Refer Slide Time: 08:42)

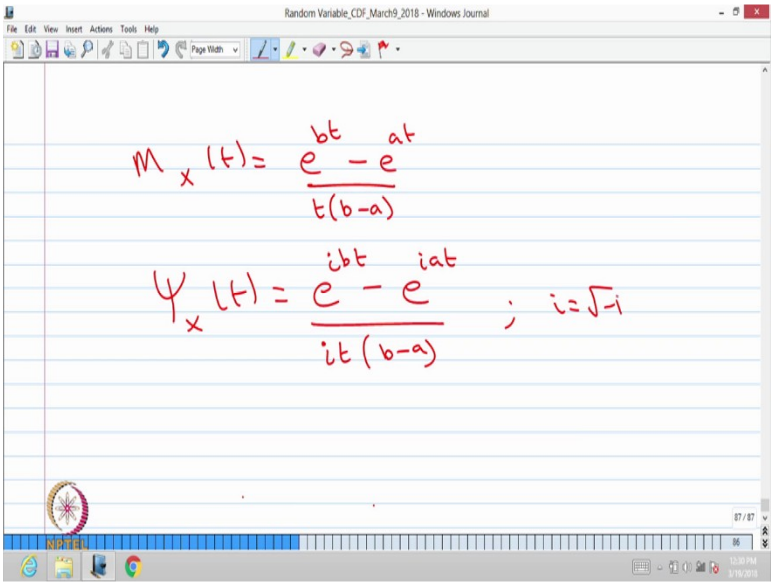


Similarly, one can find the variance, to find out the variance first you should find out what is $E[X^2]$. That is the same way $\int_a^b x^2 \frac{1}{b-a} dx$.

If you simplify this integration you will get $\frac{a^2+ab+b^2}{3}$ therefore, the variance of X is $E[X^2]-E[X]^2$. This is not the only way, you can go for variance of X is equal to $E[X^2]-E[X]^2$. So, you can compute that expectation also or we can find the $E[X^2]$then you can use this formula then you can get the variance. So, this is going to be if you substitute the value of $E[X^2]$that is $\frac{a^2+ab+b^2}{3}$ and E[X] is $\frac{a+b}{2}$ and the whole square. After simplification you can get $\frac{(b-a)^2}{2}$.

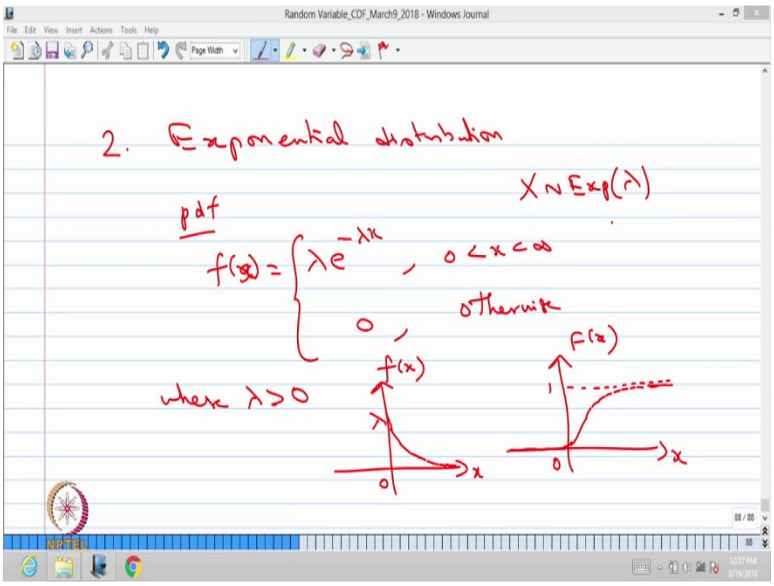
This a very important result. The mean of a continuous uniform distribution between the interval a to b is $\frac{a+b}{2}$ and the variance is $\frac{(b-a)^2}{2}$.

(Refer Slide Time: 10:16)



One can get the moment generating function for continuous uniform distribution also. I am not going for the derivation and directly giving the result $\frac{e^{bt}-e^{at}}{t(b-a)}$.Since, you know the mgf finding the characteristic function is by replace t by it. Therefore, the characteristic function of a continuous uniform distribution is $\frac{e^{ibt}-e^{iat}}{it(b-a)}$, where i is nothing but square root of - 1, complex.

(Refer Slide Time: 11:09)



Now, we will move into the next distribution that is the exponential distribution? So, before that, in notation the continuous uniform distribution is call it as a X $U(a, b)$. So, whenever we give open interval a comma b with a capital U. That means, the random variable X follows a continuous uniform between the interval a to b. Whereas, for discrete uniform distribution we say X $U\{x_1, x_2, \ldots, x_n\}$. If we supply the n points with a capital U that means, that random variable is a discrete uniform distributed.

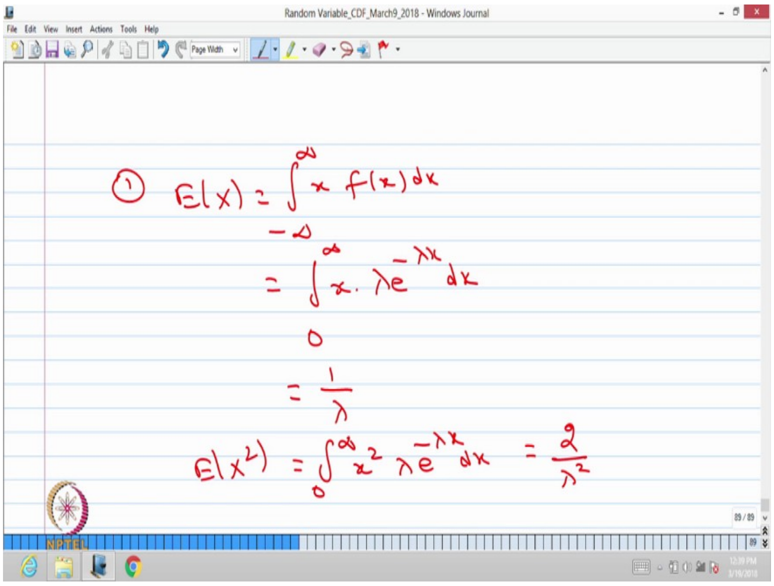
Now, you are moving to the second one that is exponential distribution. A continuous type random variable is said to be exponential distributed random variable when the probability density function of that random variable is of the form $\lambda e^{-\lambda x}$, where x lies between 0 to ∞; otherwise it is 0. Where the λ is strictly great than 0 then only it becomes a probability density function because this is greater is equal to 0 and integration from - ∞ to ∞ is going to be 1, because - ∞ to 0 the probability density function is 0 and integration from 0 to ∞ $\lambda e^{-\lambda x}$ that is going to be 1 when λ > 0.

If you supply the value of λ you are known with the exponential distribution therefore, λ is the parameter. So, we can use the notation X exp¿λ) whenever we write Exp with in bracket λ; that means, that one is exponential distributed random variable with the parameter λ. One can draw the probability density function, the probability density function starts at λ at 0 and it will be keep going down and down and it becomes 0 at ∞. Since, this is the probability density function, area below this curve from 0 to ∞ that is going to be one. When λ is greater than 0 the probability density function will be touching asymptotically 0 at ∞ and the area below that curve that is going to be one between the interval 0 to ∞. And the CDF it is 0 till 0 and keep increasing and it becomes asymptotically 1 at ∞. So, this is CDF.

The same interpretation if the data has a CDF of this form then you can conclude that data follows an exponential distribution. It is a non-linear whereas, a uniform distribution as a slanting line it is a first order in x. Whereas, this one is a non-linear and then probability density function for a uniform distribution is a constant between the interval whereas, here it is a function of x it is $\lambda e^{-\lambda x}$. There are some books they use the word a negative exponential distribution, but here we use the word exponential distribution; that means, the probability density function is $\lambda e^{-\lambda x}$. There are some books they use the parameter is 1/λ instead of λ, whether we use 1/λ or λ does not matter at the end of the day weather we will compute all other moments everything is going to be a function of parameters. So, you should remember whether you write $\lambda e^{-\lambda x}$or if the reciprocal form of λ.

So, in this course I am using consistently $\lambda e^{-\lambda x}$that is the probability density function of exponential distribution with a parameter λ that is a notation X $\exp(\lambda)$. λ.

(Refer Slide Time: 16:19)

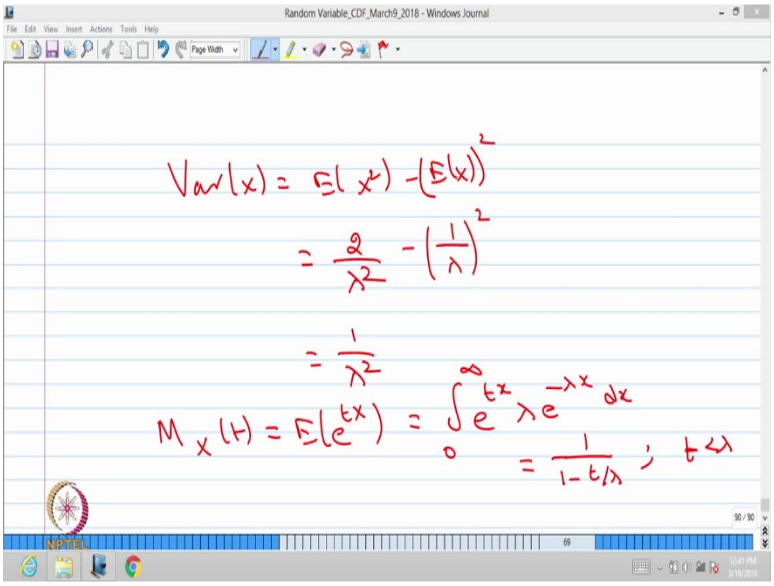


One can get mean provided it exist it is same as the $\int_{-\infty}^{\infty} xf(x)\,dx$. This is same as the since the probability density function is greater than 0 between the interval 0 to ∞. So, you can directly write $\int_0^{\infty} x \lambda e^{-\lambda x} dx$. If you simplify this integration you will get the answer that is 1/λ.

The probability density function is defined when λ > 0. Therefore, expectation of X, mean for the exponential distribution is reciprocal of the parameter. One can get the variance for the variance we can compute the $E[X^2]$ first in the same way that is the $\int_0^{\infty} x^2 \lambda e^{-\lambda x} dx$. If you do the simplification you will get $\frac{2}{\lambda^2}$. Therefore, the variance of X is going to be $E[X^2]-E[X]^2$

that is $\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2$, we got a mean of the random variable is a 1/λ therefore this is $\frac{1}{\lambda^2}$. Therefore, you will get $\frac{1}{\lambda^2}$.

(Refer Slide Time: 17:34)

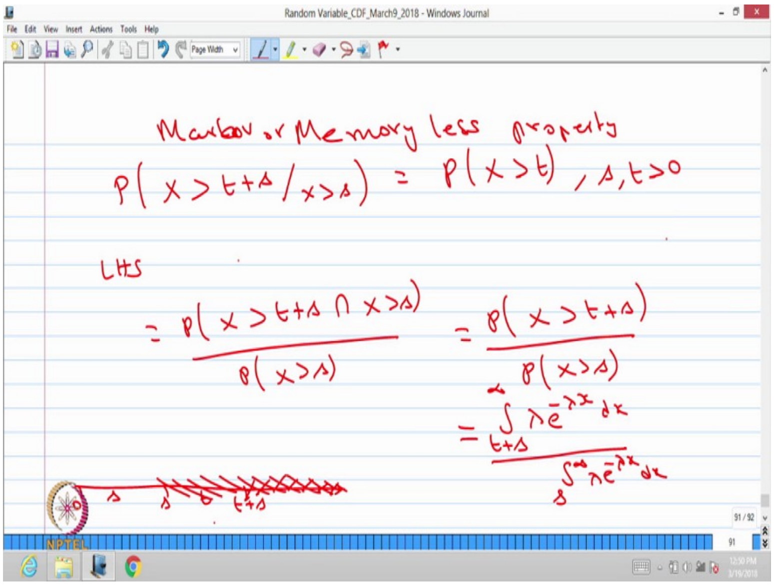


So, it is very important result the mean for exponential distribution with the parameter λ is 1/ λ and the variance of X is $\frac{1}{\lambda^2}$. You can get the mgf of exponential distribution that is $E[e^{tX}]$, that is same as the $\int_0^\infty e^{tx}\lambda e^{-\lambda x}dx$. And integration, the expectation is going to be a finite quality whenever that is going to be less then λ and the value is going to be $\frac{1}{1-t/\lambda}$.

So, this integration is going to give the value whenever the t is going to be less than λ and the value is $1-t/\lambda$. So, the mgf exists between the interval from - ∞ to λ and the value is this much. whereas, λ is strictly greater than 0. Since you know the mgf you can always get the characteristic function by replacing t by it.

Now, we will go to the one important property that is the P{X > t+x}, given x is greater than s. What is that value?

(Refer Slide Time: 19:32)

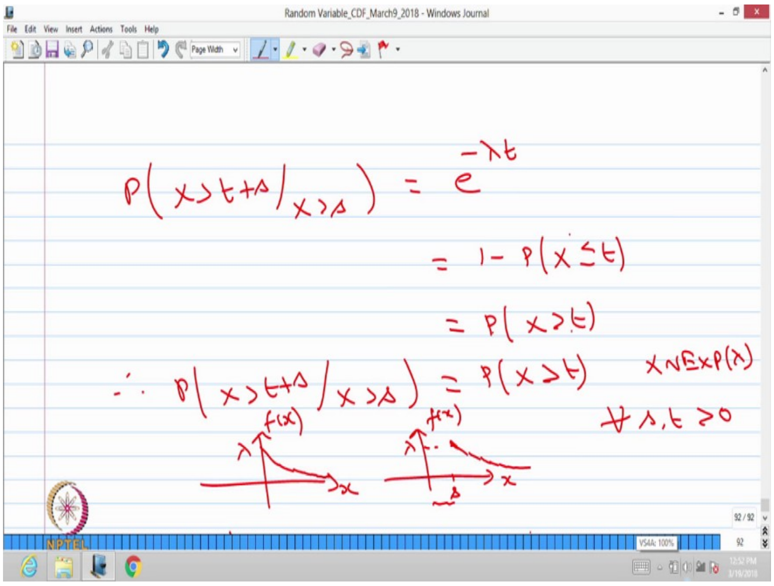


So, before filling up right hand side first we will compute this quantity then we will write down. So, let us start with the left hand side. If you compute this P{X > t + s / X > s} that is nothing but $\frac{P\{X>t+s \cap X>s\}}{P\{X>s\}}$ provided P{X > s} > 0. You can explain this concept in an easy way, just draw a line 0. Suppose you take the length s here. So, this point is s, you take another length t, therefore, this point is t plus s.

So, when I say X > t+s that means, you shade this point, when X > s that means, shade greater than s. Now, you look for what is the common portion of X > t+s with X > s that is nothing but X > t+s. So, therefore, this quantity is a P{X > t+s}, in the denominator it is P{X > s}. Since, X follows exponential distribution P{X > t+s} means nothing but $\int_{t+s}^{\infty} \lambda e^{-\lambda x} dx$.

And the denominator P{X > s} means integration from $\int_{s}^{\infty} \lambda e^{-\lambda x} dx$, either you compute this integration and simplify or you can go for 1 – P{X ≤ t+s} and the denominator also 1 – P{X ≤ s}you can compute that integration then you can do the simplification.

(Refer Slide Time: 22:00)



So, the end of the simplification you will P{X > t + s / X > s} = $e^{-\lambda t}$. The $e^{-\lambda t}$ one can write that is 1 – P{X ≤ t} that is same as P{X > t}. You can directly also can write $e^{-\lambda t}$is same as a P{X > t} or 1 – P{X ≤ t}. That means, the P{X > t + s / X > s} = P{X > t} whenever X follows exponential distribution with the probability of λ. This is true for all s and t greater than 0.

So, the left hand side is a conditional probability right hand side is the P{X > t} and one more observation the left hand side involves s as well as t whereas, the right hand side involves only t which is free from s. That means, the information about the s is disappearing the right hand side whereas, in the left side the conditional probability involves t as well as s. This is possible only when X follows exponential distribution.

That means, now I am concluding the P{X > t + s / X > s} that is same as P{X > t} this is for all s, t greater than 0 whenever X follows exponential distribution. This result we call it as a memory less property. That means, the random variable its going to take the value given it is going to take the value more than s and the probability of getting the value more than t plus s that is same as the probability of getting the value more than t which is a not a function of s it is called the memory less property.

There is another name for this property that is called Markov property. There are two names for this property either you can call it as a memory less property or it is called the Markov property, m a r k o v, because of the information till it is not occurring the s is a disappear therefore it is called the memory less property. Not only this distribution satisfies the memory less property there is a one more distribution also satisfies a memory less property that is a geometric distribution. When X is a geometrical distributed with the parameter p then $P\{X > m + n / X > m\}$ that is same as $P\{X > n\}$ where n and m are positive integers.

So, there are two distribution satisfies the memory less property one is exponential distribution which is of the continuous type, the other one is a discrete type that is geometric distribution. That means, the probability density function of the random variable X that is with the probability density function start from λ and it goes to 0 to ∞. If it does not take the value till s it is going to take the value more than s than the conditional probabilities again it is a probability of X is greater than t. That means, a from this point also the probability density function is going to be of the same form as the version. The probability density function will not change whatever the s you take given probability of X is greater than s the condition probability of X greater than t plus s that is same as probability of X is greater than t.

So, it has the same probability density function at every point in which you does not take the values. That means, this much memory is erased. That means, the interval from 0 to ∞ that information is erased; that means, the memory is erased at every stage. Therefore, it is called a memory less property. If you choose some other distribution in stop exponential distribution finding the conditional probability of left hand side you will get a function of s as well as t.

Therefore, no other distribution satisfies the memory less property, whereas the exponential distribution satisfies the memory less property. Similarly, geometric distribution satisfies the memory less property of discrete type.

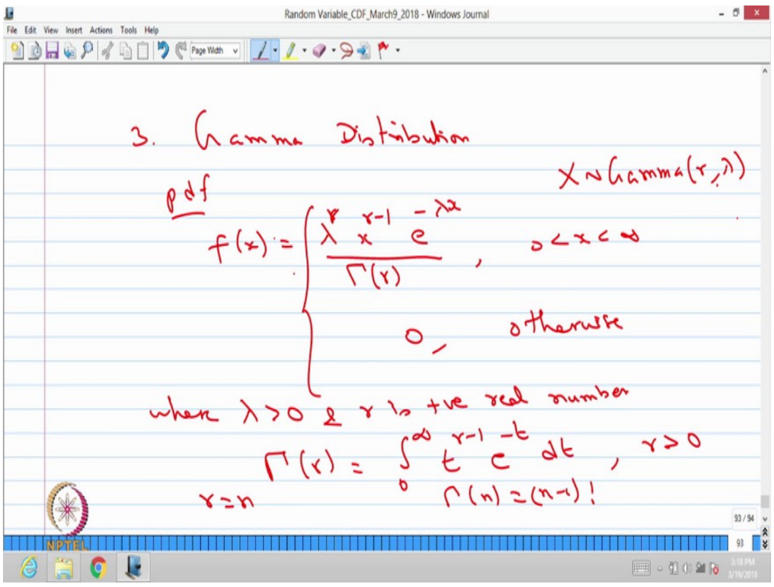
Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 21

So, we have discussed a continuous uniform distribution and exponential distribution as some common continuous type distribution. Now, we are moving into third one that is a Γ distribution, this is also the common continuous distribution.

(Refer Slide Time: 00:15)

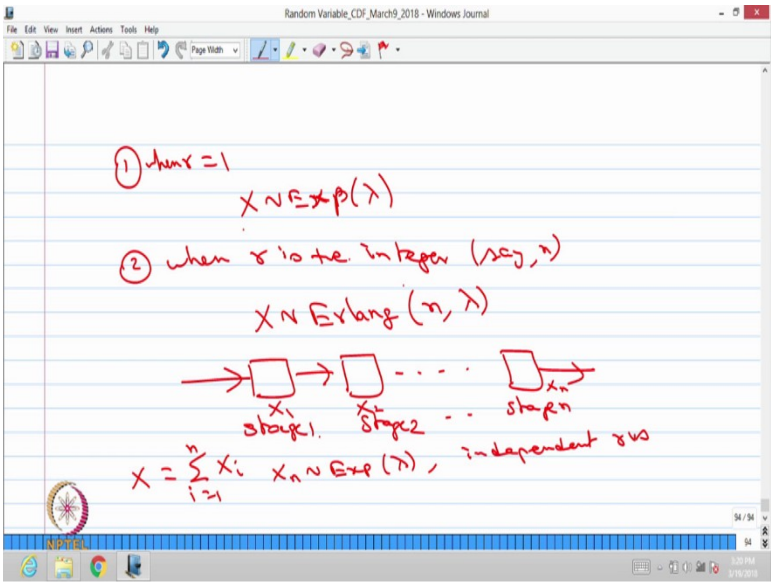


A random variable X is said to be Γ distributed if the probability density function is of the form f(x) the takes a value $\frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$ when x lies between 0 to ∞ otherwise it is 0. Here λ > 0 and r is a positive real number.

Whenever a continuous type random variable whose probability density function of this form we call it as a gamma distributed random variable. I have to define what is a Γ(r) also. Γ(r) = $\int_0^\infty t^{r-1} e^{-t} dt$. So, this is a way the Γ(r) is defined, r can be positive integers also. When r is going to be a positive integer label it as a n then Γ(n) = (n-1)!. In general, r is a real positive number and λ is again a real positive number which is greater than 0.

Then the probability density function of the form, $\frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}$; that means, if you integrate between 0 to ∞ of this f(x) that is going to be same as the denominator is independent of r therefore, you will have a $\frac{\lambda^r}{\Gamma(r)}$can be outside. And if you integrate that quantity is nothing but $\frac{\Gamma(r)}{\lambda^r}$therefore, it cancels out you get the value 1. That means, $\frac{\lambda^r}{\Gamma(r)}$ that is a normalizing constant because that makes the whole integration is 1.

(Refer Slide Time: 03:17)



One more observation, when r = 1 you can see the probability density function, when r is equal to one this f(x) = $\lambda e^{-\lambda x}$, where x is lies between 0 to ∞ otherwise 0.

If you recall this is same as that the probability density function of exponential distribution therefore, when r = 1 this becomes exponential distribution with the parameter λ. That is the notation Exp exponential distribution with the parameter λ. If you know the value of λ and r you know the probability density function of gamma distribution therefore, both r and λ are the parameters. So, in notation we can say X $Gamma(r, \lambda)$.

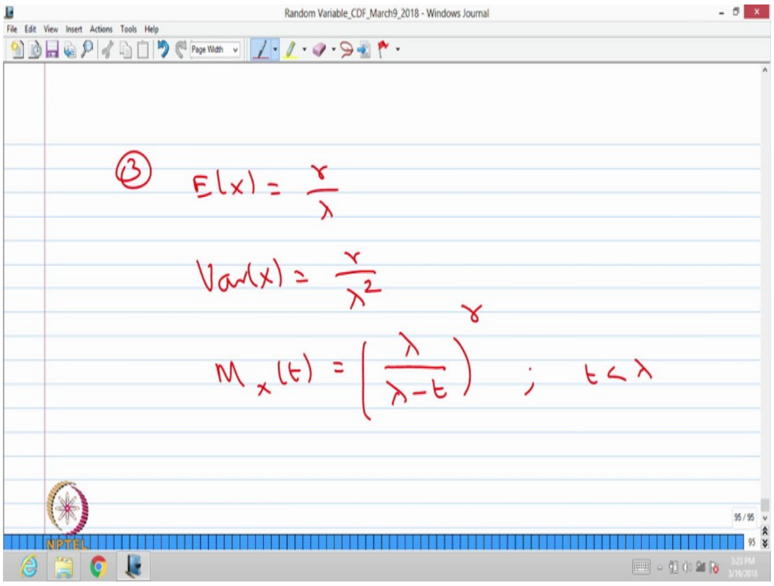
Again, when r is a positive integer say n there is another name for this gamma distribution that is called Erlang distribution, with the parameters n, λ. I am just replacing r by n when r is a positive integer. You can think of a very simple example of distribution the time taken by the system taking different stages.

So, this is a stage one, this is a stage 2, like this is a stage n. Suppose some units spend the time from all the n stages that you denoted by the letter X in which stage one time spending is X_1, time spending in a stage 2 is X_2, and similarly the time spending in the n'th stage is X_n and each one follows an exponential distribution with a parameter λ; And the time taken in each step; each stages are independent. Then the total time spent is $\sum_{i=1}^{n} X_i$this follows an Erlang distribution with a parameters n, λ.

For example, somebody entering into the hospital in which the time spending getting the admission in the hospital that takes an exponential distribution with the parameter λ. Then going to the particular doctor then the time taken spending with the doctor that follows exponential distribution with the parameter λ, the same patient is going to the stores collecting the medicine that also follows exponential distribution.

One followed by the other and all are independent then the total time spending in the hospital by entering into the hospital, spending time in the admission, spending time with the doctor, spending time with getting the medicine, then the total time that is going to be Erlang distribution with the parameters 3, λ. As long as each exponential distribution has a same parameter all are independent and the total time is a sum of all the stages of time then it is going to be considered as the Erlang distribution with the parameters n, λ, or you can call it as a gamma distribution with the parameters n, λ also both are one and the same.

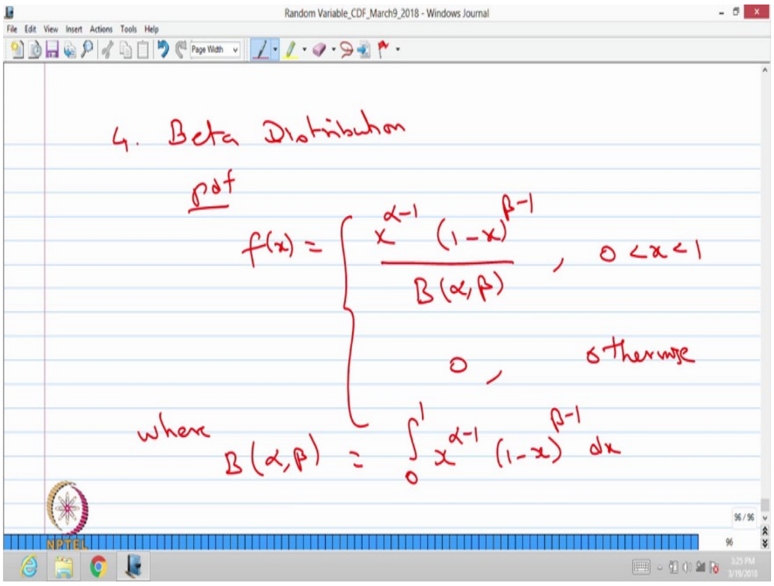
(Refer Slide Time: 07:27)



We can go for finding mean and variance if you find the mean for this gamma distribution the simple calculation will give the result $\frac{r}{\lambda}$. You can verify when r is equal to 1 it becomes an exponential distribution and you know that the mean of exponential distribution is $1/\lambda$. Similarly, if you do the simple calculation you can get the variance of X is $\frac{r}{\lambda^2}$.

So, here also one can verify with the exponential distribution. You can find the mgf of gamma distribution that is $\left(\frac{\lambda}{\lambda - t}\right)^r$ and here the $t < \lambda$. So, this is the mean, variance and mgf of gamma distribution.

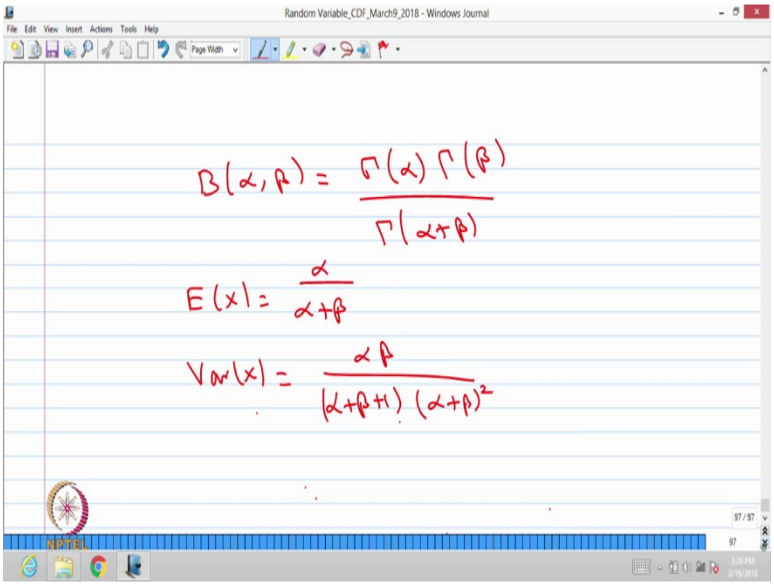
(Refer Slide Time: 08:33)



Now, will move into the next standard distribution that is called the beta distribution; Number 4, Beta distribution. A continuous type random variable X whose probability density function is of the form f(x) is $\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ and this is a valid when x lies between 0 to 1; otherwise it is 0, where the beta function is beta function of α,β; that is nothing but $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$.

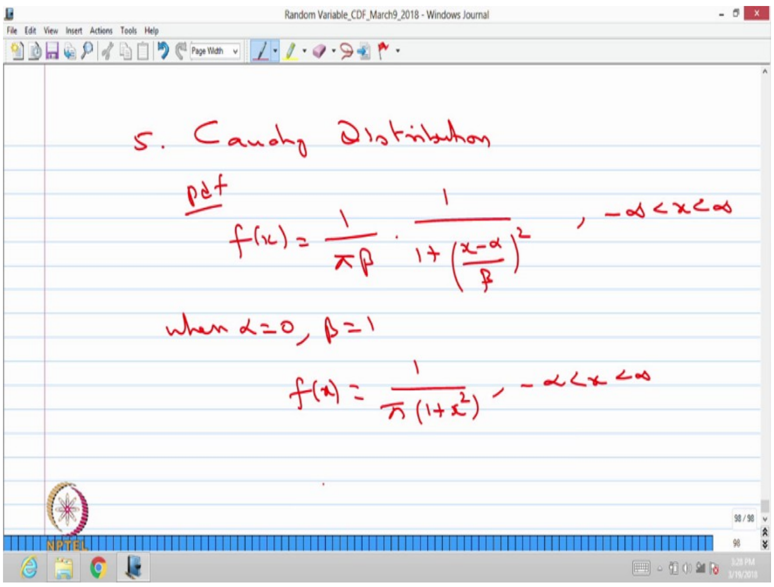
So, if you integrate between the interval 0 to 1 of this function you will get the beta function and the probability density function with the denominator beta function. That means, if you integrate it is going to be 1 that means, the beta function is going to be the normalizing constant for the beta distribution.

(Refer Slide Time: 10:12)



One can write beta function in terms of gamma function also that is $\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. For this random variable also one can get mean that is $\frac{\alpha}{\alpha+\beta}$ and you can get the variance in the same way that is $\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$.

(Refer Slide Time: 10:58)

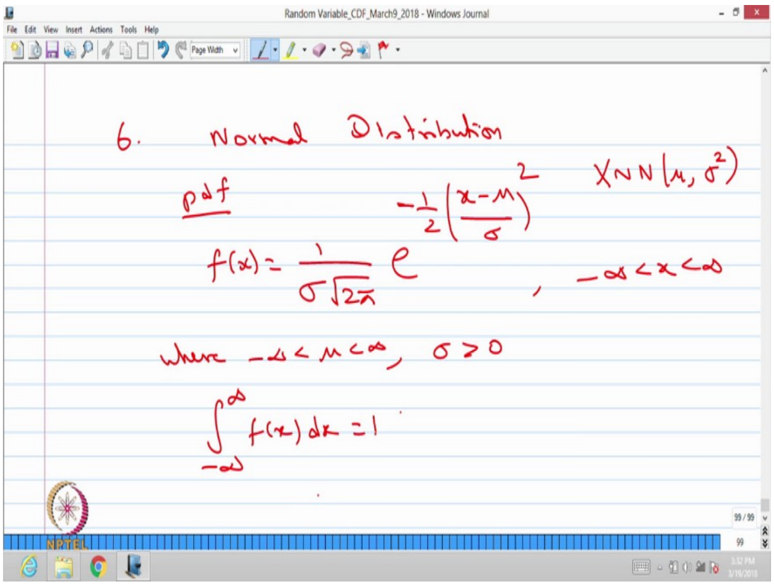


The next standard distribution that is number 5 that is Cauchy distribution; This is also continuous type and the probability density function of the form f(x) that is $\frac{1}{\pi\beta}\cdot\frac{1}{1+\left(\frac{x-\alpha}{\beta}\right)^2}$, where x lies between $-\infty$ to ∞. Sometimes we see the special case that is when α is equal to 0 and β is equal to 1, then the probability density function of the form $\frac{1}{\pi(1+x^2)}$.

So, this is a standard Cauchy distribution the importance of this Cauchy distribution the mean does not exist which we have given the proof while doing the mean for the random variable. Since the mean does not exist further moment does not exist therefore, the mgf does not exist is a very important distribution in which mean does not exist therefore, all the moment of order n does not exist therefore, the mgf also does not exist.

Now, we will move into the next very important distribution in the probability course that is normal distribution.

(Refer Slide Time: 12:48)



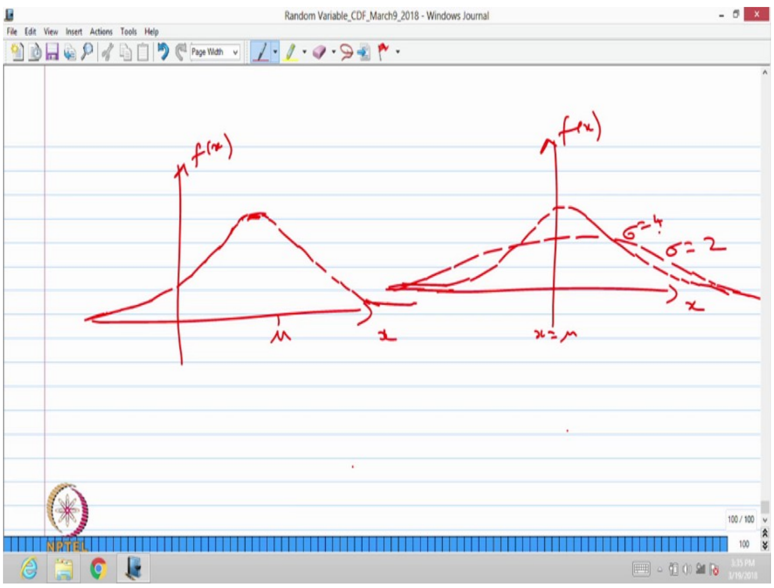
Why it is important? That when you are going to solve many more problems in the probability we will use the normal distribution quite a lot of time and also this a very important distribution because of some important result that is called the central limit theorem. Let me explain what is normal distribution and then will discuss all the moments mgf and so on than finally, we give the central limit theorem also.

Here continuous type random variable is said to be normal distribution whenever the probability density function is of the form $\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, where x lies between - ∞ to ∞. This probability density function involves μ and σ therefore, one can define what is the range of mu, where μ can lies between - ∞ to ∞ and the σ is a positive real number. Once we know the value of μ and σ we are known with the distribution of a normal distribution therefore, μ and σ are called parameters.

So, in notation we use X $N(\mu,\sigma^2)$. Why we write σ^2? That is because of the first parameter is nothing but the mean of the normal distribution and the second parameter σ^2 is nothing but the variance of the normal distribution. Therefore, we write both the parameters μ and σ^2, where σ is the positive square root of σ^2that is called as standard deviation of the normal distribution.

You can verify whether this is going to be a probability density function or not by $\int_{-\infty}^{\infty} f(x)\,dx$ that is equal to; since it is an even function you can use the calculus in particular improper integral evaluate this integration and one can conclude this integration is going to be 1.

(Refer Slide Time: 15:27)

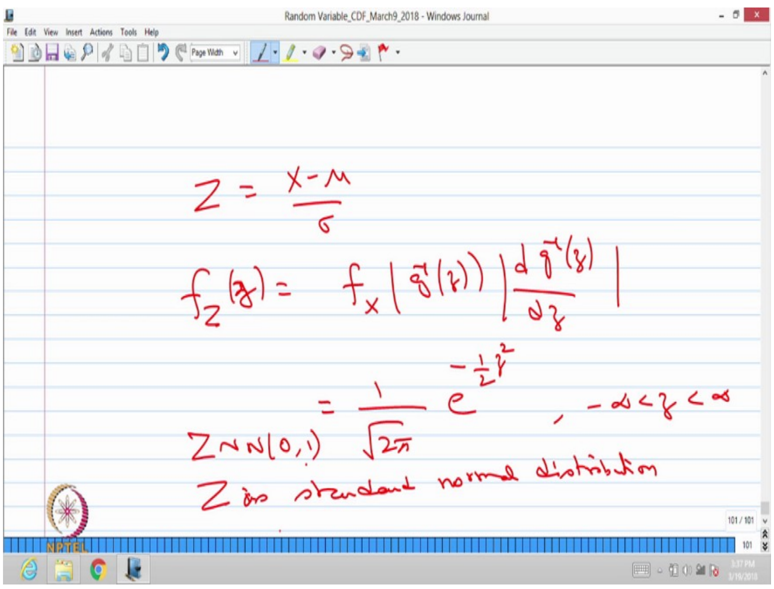


We can draw the probability density function of this. Suppose the μ takes a value here then you can draw the diagram of f(x). So, it is basically asymptotically 0 at - ∞ as well as plus ∞ and it keep increases and it has the maximum value at mu. So, this is going to be the probability density function of normal distribution, that means, the area below this curve is going to be 1. The μ is called a location parameter and the $\sigma^2$is going to be called as a scale parameter. That means, whatever the values of μ that will fix where the peak will come that will fix the location. And the σ square that will fix the spread of the distribution or the variance. So, you can have a different diagram for the different σ value.

So, let me give another diagram the probability density function for x is equal to μ supposed this is going to be the probability density function of. So, this is a peak, x is equal to μ and for example, suppose this is going to be σ is equal to 2, suppose you want to have the same μ and σ has to be the different value suppose you one can draw the σ is equal to 4 with the same μ it is going to be. So, this is going to be σ is equal to 4.

That means, the μ will be same for both the probability density function whereas, the σ will give a spread more σ we will give the more spreader and if the σ is going to be lesser and lesser you will have a less spread and you will have a peak. One can go for making a standard normal distribution in the form of making a transformation of X to another random variable called is Z with the substitution Z ¿ $\frac{X-\mu}{\sigma}$.

(Refer Slide Time: 18:05)



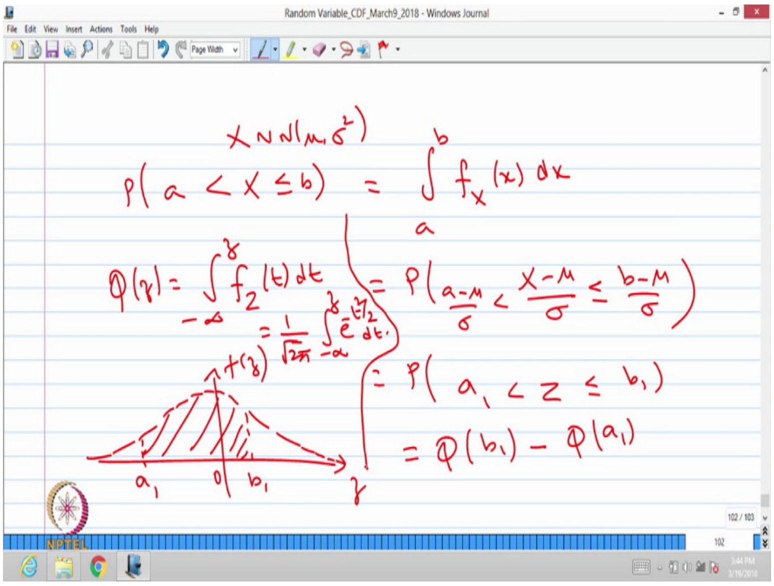
By using a function of a random variable you can get the probability density function of Z as a function of z by using the definition of the probability density function of this $g^{-1}(z)$ then derivative of $g^{-1}(z)$ with respect to z. So, you can use this, you can get the probability density function of standard normal.

You can find the distribution of Z that is going to be $\frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}z^2}$, where z is lies between - ∞ to ∞. The probability density function is $\frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}z^2}$.

Now, we call Z as standard normal distribution. Why it is called a standard normal distribution because this is also a normal distribution with the parameters 0, 1. If you compare the probability density function of the normal distribution with the probability density function of Z you can come to the conclusion $\mu = 0$ and $\sigma = 1$. That means, you do not need to specify the values, the value is always $\mu = 0$ and the $\sigma^2 = 1$. Therefore, we call Z as a standard normal distribution.

So, this is obtained by transforming a normal distribution with the parameters μ and the σ by using $\frac{X-\mu}{\sigma}$ that gives a standard normal distribution. That means, whenever you have a normal distribution you can always do this transmission $\frac{X-\mu}{\sigma}$ gives a standard normal distribution.

(Refer Slide Time: 20:30)



The advantage of this is when you find out the P{a < X ≤ b} when X follows normal distribution with the parameters μ and σ^2. Since it is a continuous type random variable you have to compute the integration from a to b the probability density function of X. This is very difficult to compute because the integration is $e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ form. So, it one cannot compute the closed from expression for this integration, that means, you need a numerical integration to evaluate the $P\{a < X \le b\}$.

So, what we do is first we transfer from this whole probability computation into standard normal distribution form, then we give the table of numerical integration to evaluate the integration therefore, you can get the probability. Let us do the same thing. So, this is same as; this is same as the $P\{a < X \le b\}$ same as $\frac{X-\mu}{\sigma}$ is lies between $\frac{b-\mu}{\sigma}$ and the left side it is $\frac{a-\mu}{\sigma}$ this is same as the probability of.

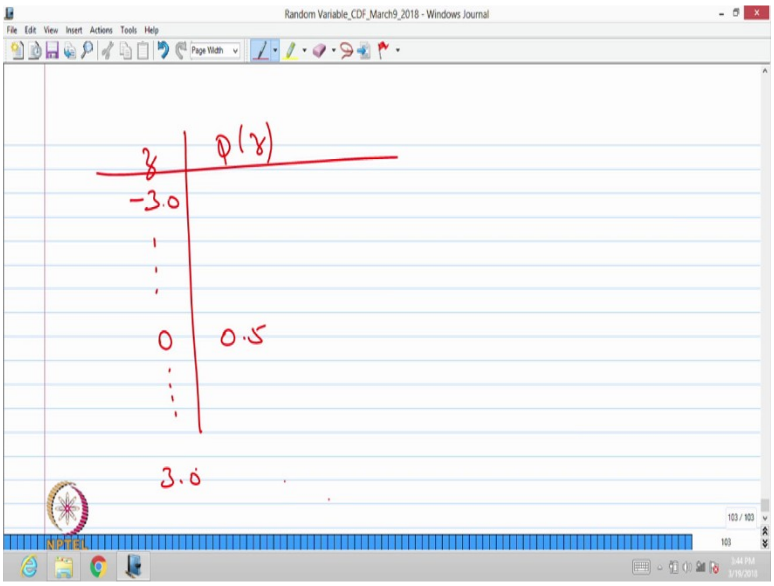
Let me write $\frac{X-\mu}{\sigma}$ as a Z, less than or equal to let me treat $\frac{b-\mu}{\sigma}$ for a known value of μ and σ the whole thing is known therefore, just let me write as the b_1 point and $\frac{a-\mu}{\sigma}$ that I make it as a_1. That means, the probability of X lies between a to b that is same as a probability of Z lies between a_1 to b_1. That is same as in a standard normal distribution if the probability density function is ok; suppose a_1 is somewhere here and suppose b_1 is somewhere here. So, Z lies between a_1 to b_1 that is nothing but this shaded area.

The way I have taken a_1 is in the left and b_1 is in the right of 0 and you know that it is symmetric about z is equal to 0 therefore, the left side area is 0.5 and the right side area is 0.5. Therefore, this is same as for that I am going to do one more notation Φ(z) is nothing but - ∞ to z probability density function of the standard normal distribution. Φ(z) is a CDF at the point z for the standard normal distribution.

Therefore, this becomes Φ(b_1)- Φ(a_1). The probability of a standard normal distributed lies between a_1¿b_1 that is same as Φ(b_1)- Φ(a_1), since it is a continuous type random variable there is no mass at any point. Therefore, this probability computation is valid whether you make it both are closed interval or both are open interval or one is open and one is closed and so on. For all the 4 types the probability lies X lies between a to b that is same as probability Z lies between a_1¿b_1 that is same as Φ(b_1)- Φ(a_1).

Now, Φ(b_1) you can compute numerically, similarly Φ(a_1) you can compute the difference is going to be probability lies between a_1¿b_1.

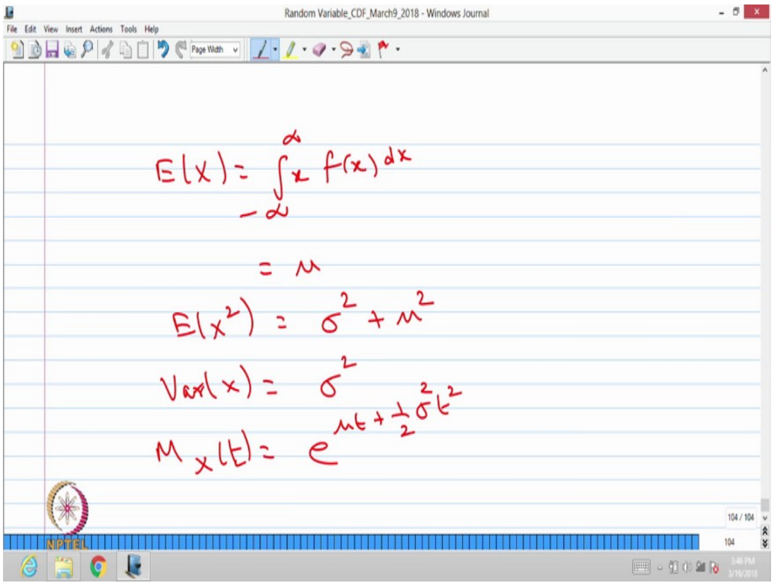
(Refer Slide Time: 24:49)



So, there is a table for different values of z you can have a $\Phi(z)$. It start from, usually it start from - 3 and for 0, $\Phi(0)$ is nothing but integration from - ∞ to 0 and since it is symmetric therefore, this value is 0.5 and it will be keep going till 3.0.

So, as far as the exam is concerned we will supply the $\Phi(z)$ values with the notation $\Phi(z)$ $\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}t^2} dt$. I can write one more line also that is $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-1}{2}t^2} dt$ both are one and the same.

(Refer Slide Time: 25:54)

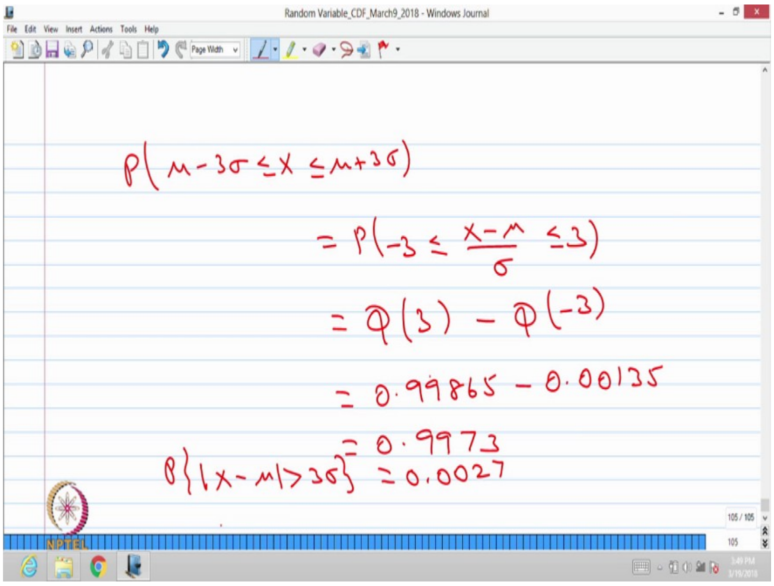


Now, you can go for computing mean and variance of a normal distribution. Mean is $\int_{-\infty}^{\infty} f(x) dx$ where probability density function of this you can simplify by substituting the f(x) value and so on one can get the value μ. And similarly, if you compute $E[X^2]$, E$[X^2]$ you will get a $\sigma^2 + \mu^2$. Therefore, the variance of X is going to be $E[X^2] - E[X]^2$ therefore, you will get σ^2.

One can get the mgf also, mgf of normal distribution, whenever I use the word X that is a normal distribution, whenever I use the word Z; that means, is a standard normal distribution;

So, mgf of normal distribution that is going to be $e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. One can evaluate this is the mgf of normal distribution with the parameters μ and the σ^2.

(Refer Slide Time: 27:17)

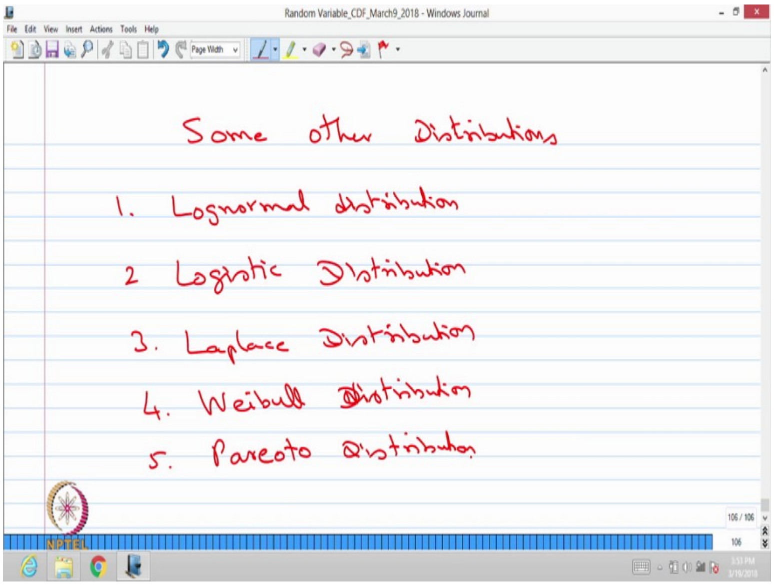


One can get easily the probability of X lies between μ + 3σ; μ - 3σ to μ + 3σ that is same as probability of $\frac{X-\mu}{\sigma}$ lies between - 3 to 3 that is same as Φ(3) – Φ(- 3).

If you compute Φ(3) from the table one can get a 0.99865, and the Φ(-3) is 0.00135 therefore, you will get 0.9973. That means, when X is a normal distribution and the probability of a μ - 3σ to μ+3σ is 99.73 percentage accumulate in that range that is equivalent of same the $P\{\vee X-\mu\vee ¿3\sigma\}$ that probability is 0.0027. So, away from μ - 3σ to μ + 3σ that probability is 0.0027 or within the interval X lies between μ - 3σ to μ + 3σ that is probability is 0.9973 is an important observation.

Like that one can compute what is the probability of X lies between μ - 2σ to μ + 2σ. Similarly one can compute probability of X lies between μ - σ to μ + σ. These are all the standard results for the normal distribution. The main use of this normal distribution is a central limit theorem that I will explain after I introduce a multidimensional random variable and so on, so that I will discuss at the end of the probability part.

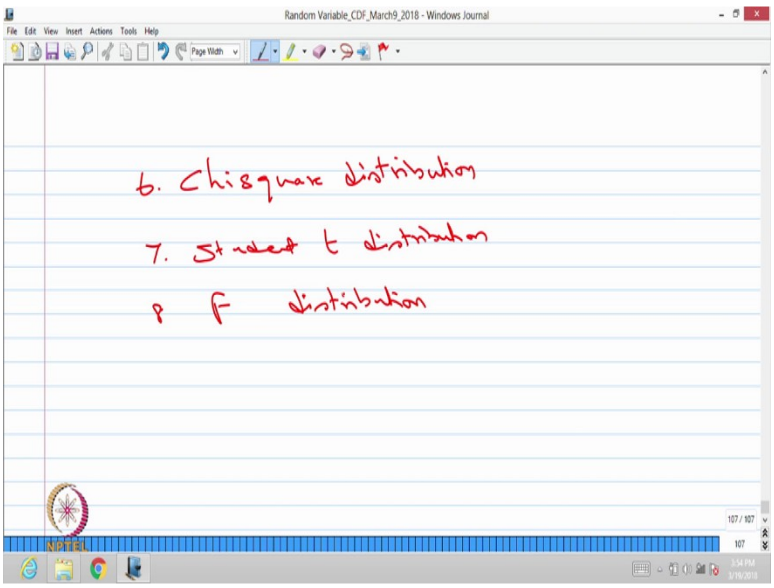
(Refer Slide Time: 29:44)



Some other distributions the first one which I am not going to discuss in detail I am just going to give the name of the distribution that is Lognormal distribution. This distribution is the function of a normal distribution in which the range is between 0 to ∞. Therefore, it is of the important.

The second one that is logistic distribution, this is also an important distribution which is in the form of exponential function; Third one Laplace distribution, and the fourth one Weibull distribution, and fifth Pareoto distribution.

(Refer Slide Time: 31:11)



Then we have a chi square distribution, then we have a student t distribution, then we have F distribution. These three distributions that is a chi square distribution, student t distribution and F distribution, these three distributions are very important distribution for statistical inference. So, whenever you do the course on statistics then we need this distribution to discuss the statistical inference. So, we are not going to discuss in detailed about these distributions as for as this course is concern.

Like that there are many more standard distributions which is or not interest to us therefore, now I am stopping here.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 22

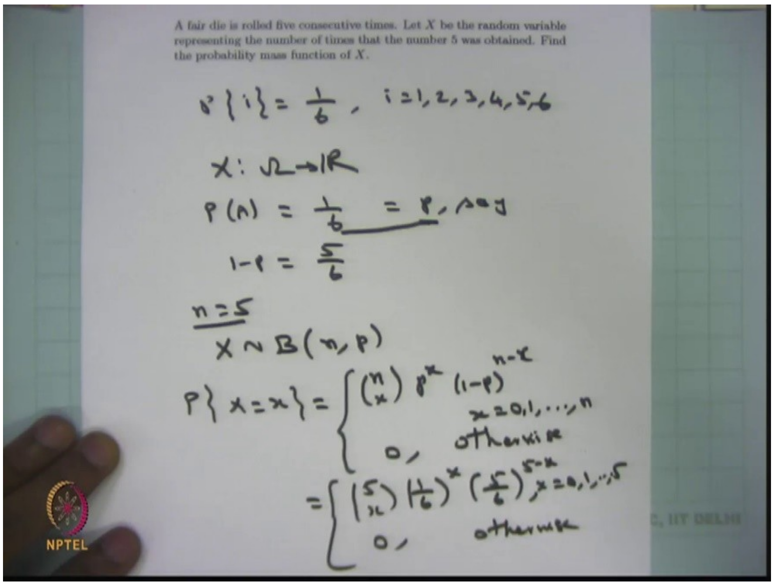
In this module we have discuss the various Standards Distributions, both the discrete type as well as the continuous type in last two lectures. That is in the first lecture we have discussed various discrete type common distributions, starting from constant random variable, then Bernoulli distribution, then binomial distribution, geometric distribution, then discrete uniform distribution, then we discuss the Poisson distribution. So, these are all the standard discrete type distribution which we have discussed in the lecture 1.

In the lecture 2 we have discussed standard or common distributions of continuous type random variables. In that lecture we have discussed continuous uniform distribution between the intervals, then we discuss the exponential distribution, then we discuss the gamma distribution, beta distribution, cauchy distribution, and we have given some list of distributions.

In particular we have discussed the normal distribution that is very important distribution of a continuous type random variable which is a common distribution. So, we have discussed normal distribution also. From the normal distribution how one can get the standard normal distribution then we have solved 1 or 2 problems in the normal distributions also.

Now, in this lecture we are going to give few problems then one can identify what is the correct distribution attached with those problems, then using those common distributions one can get the solution. That means, we are going to use these common distributions to get the solution of the given problem.

(Refer Slide Time: 02:00)



The first problem is a fair die is rolled 5 consecutive times. Let X be the random variable representing the number of times that number 5 was obtained. The question is to find the probability mass function of the random variable X. Here the fair die means it is unbiased; that means, the probability of occurrence of each possible outcome foreseen and since there are 6 possibilities therefore, the probability of each possible outcome is going to be 1/6.

So, a fair die is rolled 5 consecutive times therefore, the probability of each possible outcome that is equal to 1/6, when i takes the value 1, 2, 3, 4, 5 and 6. So, this is going to be the collection of all possible outcomes and the probability is going to be 1/6 because a fair die is rolled consecutively 5 times. So, now, the random variable X is defined from **Ω** to R where **Ω** is a collection of all possible outcomes based on this random experiment that random experiment is a fair die is rolled 5 consecutive times. Therefore, let X be the random variable representing the number of times the number 5 was obtained.

So, you can make an event A that is nothing but the number 5 was obtained, the event A is the number 5 is obtained. The probability of event A is going to be 1/6. So, that can be treated as the probability of success for each roll. Like that we are making 5 independent roles; that means, the probability of success is 1/6 and the probability of failure is 5/6 in each Bernoulli trial, here when I say Bernoulli trial the getting the number 5 with the probability 1/6 that is going to be the probability of success and the probability of a failure is not getting the number 5 that is with the probability 5/6.

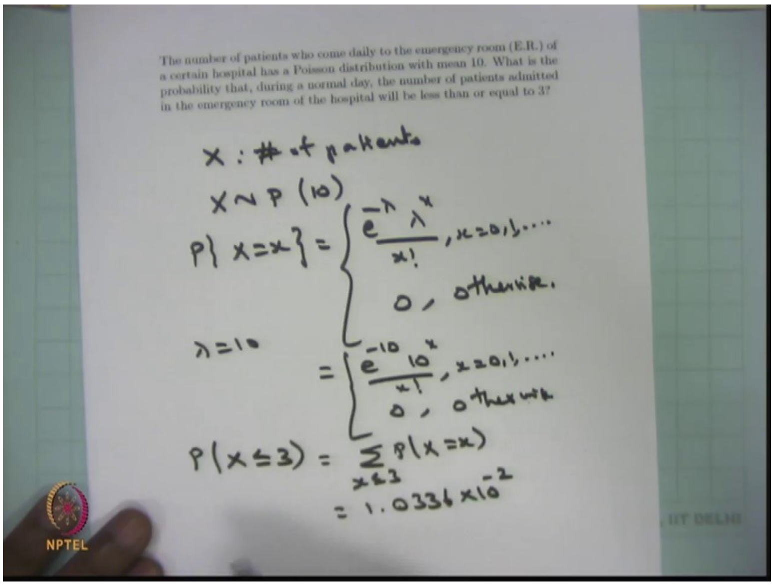
So, therefore, each Bernoulli trial with the probability p that is 1/6 and the failure probability is 5/6. Like that we have n independent Bernoulli trials here n is going to be 5. So, the random variable say is representing the number of times the number 5 was obtained. That means, it is same as 5 independent Bernoulli trials and the X represents the total number of n independent Bernoulli trials gives the values.

Therefore, we can conclude X follows binomial distribution with the parameters n, p here n is 5 and p is 1/6. In this problem the X follows a binomial distribution with a parameters n and p where n is 5 because 5 consecutive times we are rolling the dice all are independent and the probability of success in each role getting the number 5 that is 1/6.

Now, the question is a find the probability mass function of X. You know that since it is binomial distribution immediately you can write the probability mass function is of the form ${}^{n}C_{x}\, p^{x}(1-p)^{n-x}$, where x takes a value 0, 1 and so on till n; 0 otherwise. This is a probability mass function of binomial distribution. So, in this problem this is going to be ${}^{5}C_{x}\left(\frac{1}{6}\right)^{x}\left(1-\frac{1}{6}\right)^{5-x}$, when x takes a value 0, 1 and so on till 5; 0 otherwise.

In this problem I have stopped it finding the probability mass function. Once you know the probability mass function suppose you want to find out the $P\{X \leq 3\}$ or if you want to find out the $P\{X > 4\}$. So, all those things nothing but the probability of some events; so, once you know the probability mass function you can get the probabilities. Suppose the question is a find the mean, variance because since it is a binomial distribution you can use the relation of mean and variance and so on and even you can find the further moments for this problem.

(Refer Slide Time: 07:22)



Now, we will move into the second problem. The second problem is the number of patients who come daily to the emergency room of a certain hospital has a Poisson distribution with the mean time. What is the probability that during a normal day the number of patients admitted in the emergency room of the hospital will be less than or equal to 3?

This is a very typical situation in which the number of people or number of customers or number of units entering into the system. Any system sometimes we can make the assumption of Poisson distribution because it comes in a very rare event and the possible values are 0, 1, 2 and so on countably infinite in that case it is good to make the assumption of at that follows Poisson distribution.

So, therefore, in this problem it is already made the assumption it follows a Poisson distribution with the mean time. So, the question is what is the probability that the number of patients entering into the hospital will be less than or equal to 3. Since already made the assumption, it follows the poison distribution. So, the question is immediately you can make out you can create a random variable X is nothing but the number of patients, who come daily to the emergency room.

Since we made already the assumption X follows Poisson distribution with the mean time usually, we write the parameter. If you recall the Poisson distribution suppose the Poisson distribution has the parameter λ the mean is all λ even the variance is also going to be λ.

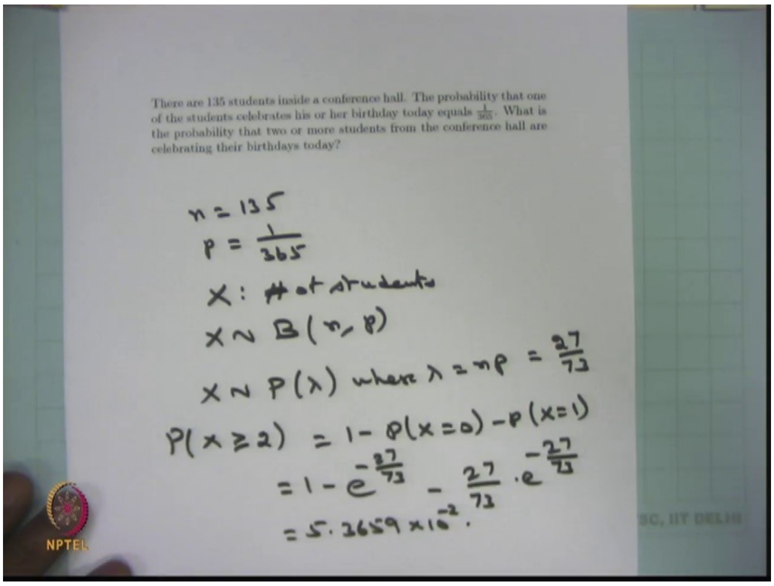
So, here the information is given the Poisson distribution with the meantime that means, it is a λ is 10. Therefore, you can immediately write down the probability mass function of the Poisson distribution that is $e^{-\lambda}\frac{\lambda^x}{x!}$, when x takes the value 0, 1, 2 and so on; otherwise 0. So, this is a probability mass function. So, in this problem λ is 10. Therefore, this is $e^{-10}\frac{10^x}{x!}$ when x takes a value 0, 1,…, so on, otherwise 0.

So, the question is what is the probability that the during a normal day in the number of patients admitted in the emergency room of the hospital will be less than or equal to 3. That means, you have to convert the given problem into the form of that is X less than or equal to 3 this is the required property. This is same as summation of probability of X takes value small x, when small x is going to be less than or equal to 3; that means, it is P{X = 0}+P{X = 1}+P{X = 2}+P{X = 3}. You are substitute X is equal to 0, 1, 2, 3 and get the probability mass at those add all the values if you simplify you will get the answer it is a 1.0336¿ 10^{-2}.

Even this problem can be asked in find out the probability that no customer or no patient admitted in a normal day; that means, that is a P{X = 0} or you can ask what is the probability that always some patients admitted in the emergency room on a normal day; That means, that is 1 – P{X = 0}. Or there is a possibility of question what is the variance of X.

Since you know the poison distribution, the mean and variance are same, the variance is also going to be meantime. So, like that many more problems can be created once the situation is given and some information is provided you can find other measures in nice way. We will move into the third problem.

(Refer Slide Time: 12:16)



The third problem is a there are 135 students inside a conference hall. The probability that one of the students celebrates his or her birthday today equals 1/365 what is the probability that two or more students from the conference hall are celebrating their birthdays today.

In this problem we made some assumption the year is not a leap year therefore, we made the probability that the one of the students celebrate his or her birthday today equals to 1/365; that means, we made the assumption the year has 365 days. The question is what is the probability that two or more students from the conference hall celebrating their birthday today.

Two or more that is equivalent of saying negation in the form of 1 minus no and 1; that means, at the probability that two or more students from the conference hall celebrating their birthday today that is same as 1 minus probability that no one is celebrating minus only 1 person celebrating because a summation of all the probabilities is 1. So, this question either you can solve it what is the probability that 2 people having as a birthday today and 3, 4 and so on add all the values or we can go for 1 minus of no one celebrating minus only 1 person celebrating.

So, let us find out the probability. The clue is there are 135 students and the probability of a success is 1/135. Suppose I make it n is equal to 135 and p is equal to 1/365. I can match this problem with the number of students celebrating the birthday that is number of students celebrating their birthdays that is going to be X. Then I can conclude X follows binomial distribution with the parameters n, p, where n is a 135 and the p is a 1/365. By doing this I can get the $P\{X \geq 2\}$ by applying the binomial distribution probability mass function and so.

Here we should observe the n = 135 and small p is; n is sort of very large and p is almost close to 0. Even though we say the p is open interval 0 to 1, here the p probability of success that is very small therefore, it is good to do by approximating this binomial distribution with the Poisson distribution, that is a X follows Poisson distribution with the parameter λ. Now, the question is what is λ? Where λ = np that means, you multiply 135 with 1/365 that is going to be the λ value.

So, one can simplify and get the value λ. So, if you do the simplification that λ is going to be $\frac{27}{73}$. So, once you get the λ that's a parameter for the Poisson distribution; that means, you know the probability mass function of the Poisson distribution therefore, the required probability is the probability that two or more students from this conference hall are celebrating their birthday today, that is nothing but $X \geq 2$. That is same as either one way find out the probability of X equal to 2, probability of X is equal to 3 and so on sum it up or the other way is $1 - P\{X = 0\} - P\{X = 1\}$.

So, this is easier than finding out the other side because the summation of probabilities 1 therefore, this is same as this. So, you can get a $P\{X = 0\}$ because you know the probability mass function. So, that is 1 minus probability of mass function for the Poisson is $e^{-\lambda}\frac{\lambda^0}{0!}$ as x = 0.

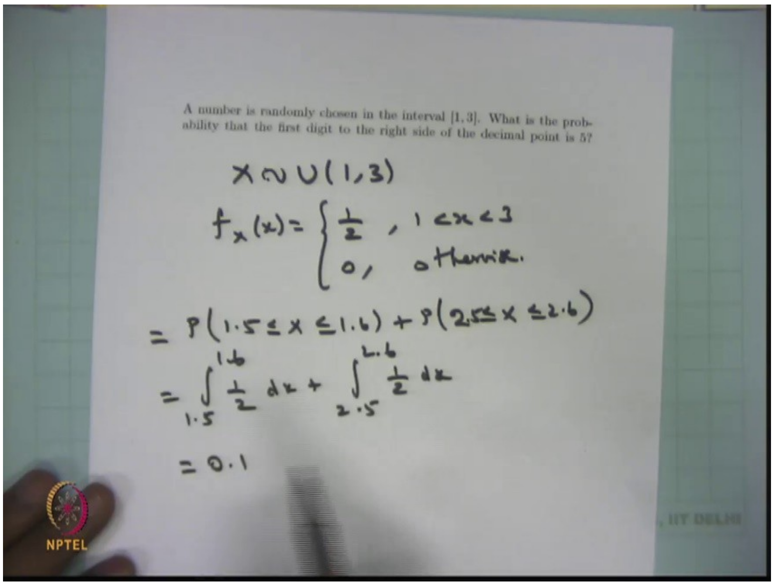
So, it is going to be $e^{\frac{-27}{73}}$ when you substitute $P\{X = 1\}$ you will get $\frac{27}{73}e^{\frac{-27}{73}}$. If you simplify you will get the answer that is 5.3659 times 10^{-2}. So, this is an easy problem you can create a many more problems with the same first two statements. Since, the n is large and p is almost 0 we are going for binomial to the Poisson distribution.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 23

So, till now we have discussed the problems on common distributions of discrete type. We have discussed one problem in binomial distribution and other two problems in the Poisson distribution. Now, we are moving into the problems in common distributions of continuous type random variables.

(Refer Slide Time: 00:28)



So, let us start with the first problem. A number is randomly chosen in the interval 1 to 3. What is the probability that the first digit to the right side of the decimal point is 5? Note that whenever we choose a number randomly; that means, the possible sample points are equiprobable; whenever we use a word a number is randomly chosen; that means, it is a real number.

The real number is a randomly chosen for example, in the interval [0, 1] that is basically a random number generation. That means, each possible real number is equiprobable. Whenever it is equiprobable or equi likely of that events; that means, the underlying distribution is uniform distribution. Since we are randomly choosing a real number therefore, it is going to be of the common distribution in particular it is of the uniform distribution of continuous type.

So, in this problem the real number is randomly chosen in the interval [1, 3], the random number generation the default random number generation that is between the interval [0, 1]. But here in the interval [1, 3] therefore, we can conclude X is a random variable which denotes getting the real number between the interval [1, 3] that is going to be a random variable.

So, since X is a random variable which denotes number obtaining in the interval 1 to 3 this follows uniform distribution between the interval [1, 3]; between the interval [1, 3]. That means, the probability density function of this random variable is; since this random number is randomly chosen; that means, it is equiprobable therefore, the density is going to be constant then only it is going to be equiprobable.

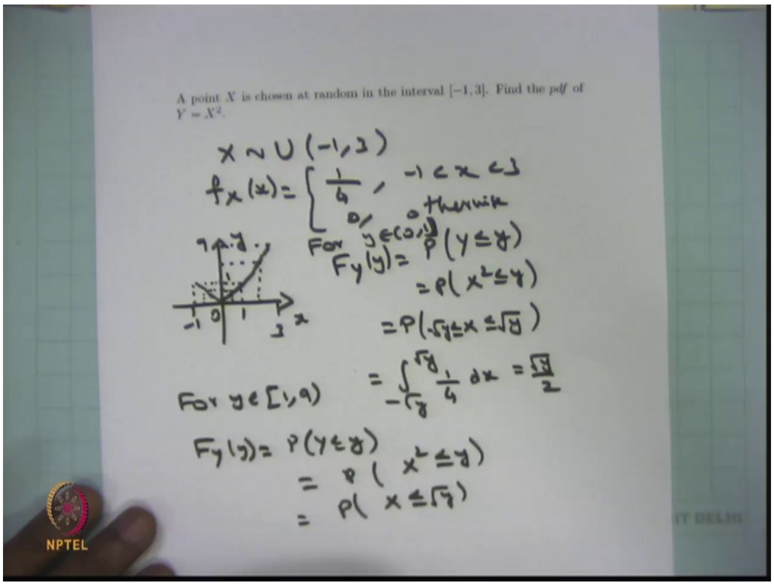
Since the interval is [1, 3]. So, the length of the interval is 2 therefore, the probability density function in the constant that is 1/2. Between x when x lies between [1, 3]; 0 otherwise. Since the random number is randomly chosen; that means, it is uniform. The interval length is 2 therefore, it is 1/2 between the interval [1, 3].

So, this is a probability density function of a uniform distribution between the interval [1, 3]. So, now, the question is the probability that first digit to the right of a decimal point is 5; that means, this is possible when X lies between 1.5 to 1.6 as well as the X lies between 2.5 to 2.6. The required property is $P\{1.5 \leq X < 1.6\} + P\{2.5 \leq X < 2.6\}$ that is same as.

Since it is a continuous type random variable $\int_{1.5}^{1.6} \frac{1}{2} dx + \int_{2.5}^{2.6} \frac{1}{2} dx$, the probability density function is 1/2. You simplify get the answer that is 0.1; that means, the probability that the first digit to the right side of the decimal 0.5 is 0.1.

The probability can be always a can be represented in the form of proportion also so; that means, you can go for proportion or percentage; that means, it is a 10% that it is going to be having first digit to the right of the decimal point is 5, like this we can create many more problems with this setup.

(Refer Slide Time: 05:17)



We will move into the second problem; in this problem a point X is chosen at random in the interval, find the probability density function of $Y = X^2$. That means, we can use the concept of the previous problem, that is a point X is chosen at random, a point means it is a real number randomly chosen between the [-1, 3], whenever it is random insets a uniform distribution. Therefore, we can conclude X follows uniform distribution between the [-1, 3] from the previous problem in the same way one can conclude X follows continuous uniform distribution between the interval [-1, 3]. Therefore, the probability density function is going to be 1 divided by length of the interval that is 4 when x lies between - 1 to 3; 0 otherwise.

We will keep the probability density function of X because of the question is find the probability density function of $Y = X^2$. That means, with the help of the distribution of X here to find out the probability density function of Y. There are two ways you can use the CDF method to get the CDF of Y then by differentiating, you can get the probability density function of Y, or you can use the continuous type random variable result finding the probability density function, if it satisfies the function is a monotonic and differentiable and get the probability density function using the probability density function of X.

But here the interval is minus when x takes a value, - 1 to 3, $Y = X^2$; that means, you just have to draw the parabola. So, between - 1 to 0 it is a decreasing, 0 to 3 it is increasing. Therefore, you cannot apply the theorem of continuous type random variable satisfying the monotonic function and so on, you cannot apply. Therefore, we will go for finding the probability density function of Y by finding the CDF of Y first, then we go for the probability density function of Y.

That is by seeing the diagram you can make out $Y = X^2$where x takes a value - 1 to 3 therefore, y is going to take the value from 0 to 9, the y going to take the value from 0 to 9. Therefore, the CDF is going to be 0 and 1 here to find out what is the CDF between the interval 0 to 9, from 9 onwards at the CDF is going to be 1.

Because of x takes a value from -1 to 3 and $y = x^2$therefore, the CDF of Y till 0 it is 0, from 9 onwards it is going to be 1. So, the question is a what is a CDF between 0 to 9. One more observation you see the diagram when x takes a value from -1 to 1 y takes a value 0 to 1 whereas x takes a value from 1 to 3 y takes a value from 1 to 9. Therefore, the CDF is going to be different form from 0 to 1 for the random variable Y, whereas 1 to 9 the CDF is going to be of the different form.

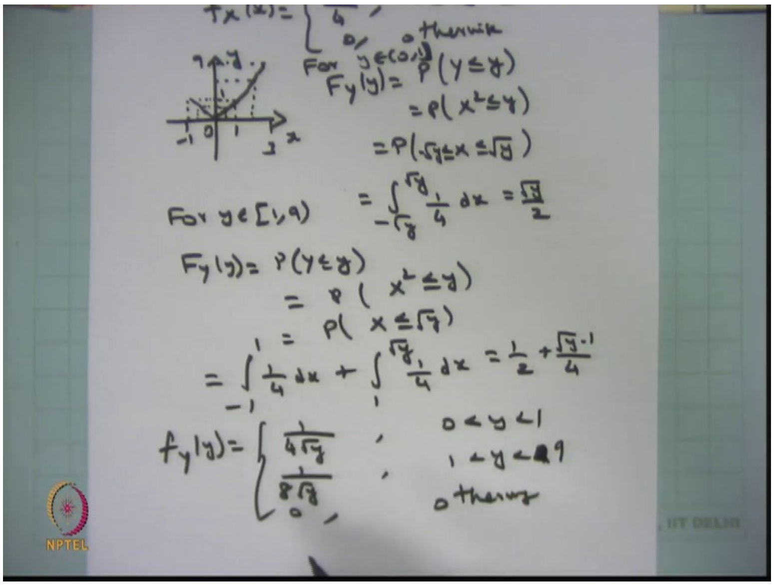
Let us first write how the CDF going to be calculate. The CDF of the random variable Y that is nothing, but the P{Y ≤ y}, that is same as $P\{X^2 \le y\}$. That is same as$P\{-\sqrt{y} \le X \le \sqrt{y}\}$. Here I making the assumption y > 0, as I said the CDF till 0 that is going to be 0.

So, our calculation goes for the CDF between the interval 0 to 9. This is same as since X is a uniformly distributed between the interval [-1, 3], this is same as $\int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{4} dx$ and the probability density function of the X is 1/4. You can simplify and you can get the answers $\frac{\sqrt{y}}{2}$.

Now, will go for what is the CDF between. So, this is a CDF for Y lies between 0 to 1; 0 to open interval 1. I am splitting the interval 0 to 9 in the form of 0 to 1, then I will go for 1 to 9. So, when y belonging to 1 to 9, now the CDF of Y because only 0 to 1 you have the P{-$\sqrt{y} \le X \le \sqrt{y}$} because so, for every point you have two inverse images. Therefore, you will get a -$\sqrt{y}$ to $\sqrt{y}$ when y lies between 0 to 1 fine.

So, now we are going for a y belonging to 1 to 9. So, the CDF of Y that is same as P{Y ≤ y}, that is same as now for every y when y is lies between 1 to 9, you have only one inverse not the two inverses $P\{X^2 \le y\}$, that is same as the $P\{X \le \sqrt{y}\}$.

(Refer Slide Time: 12:05)



That is same as the $P\{X \leq \sqrt{y}\}$when y is lies between 1 to 9; that means, it is a from -∞ to -1 the probability density function of X is 0. Therefore, it is $\int_{-1}^{1} \frac{1}{4} dx + \int_{1}^{\sqrt{y}} \frac{1}{4} dx$, the probability density function is 1/4.

You see the difference when y lies between 0 to 1 the P$\{X^2 \leq y\}$is same as $\{-\sqrt{y} \leq X \leq \sqrt{y}\}$ because it has two inverses between the interval 0 to 1. And y takes a value 0 to 1, X has two inverse images. Therefore, it is -$\sqrt{y}$ ¿ $\sqrt{y}$. When y takes a value 1 to 9, the P$\{X^2 \leq y\}$ that is same as $P\{X \leq \sqrt{y}\}$ that is same as -∞ to -1 the probability density function is 0 + $\int_{-1}^{1} \frac{1}{4} dx + \int_{1}^{\sqrt{y}} \frac{1}{4} dx \cdot$

So, you do the simplification, you can get the answer that is $\frac{1}{2} + \frac{\sqrt{y}-1}{4}$. Therefore, one can write the CDF of the random variable that is combining -∞ to 0 that is 0, from 0 to 1 that is $\frac{\sqrt{y}}{2} \wedge$ ¿between the interval 1 to 9 it is $\frac{1}{2} + \frac{\sqrt{y}-1}{4}$, from 9 onwards the CDF is going to be 1. By differentiating the CDF of Y you can get the probability density function of Y that is the question. The question is a find the probability density function of Y ¿ X^2.

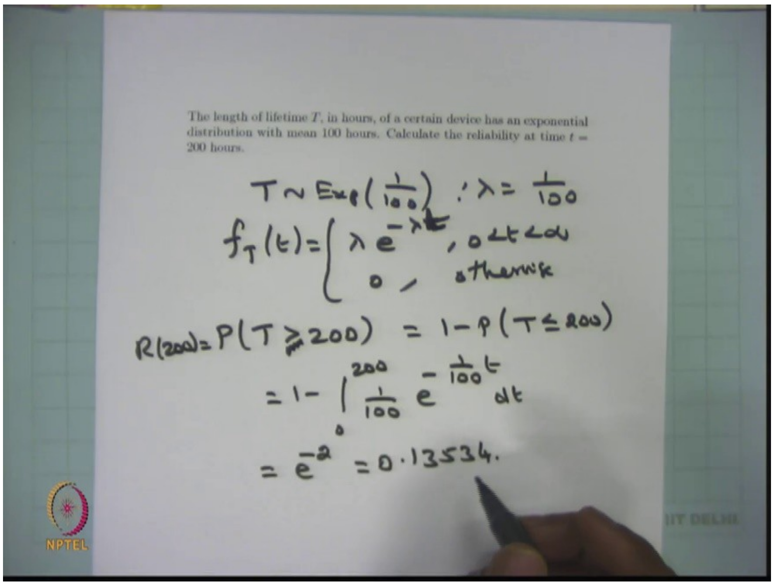
If the question is find the distribution of Y, you can leave it with the CDF. Since the question is find the probability density function, you have to differentiate the CDF to get the probability density function of Y. By differentiating the CDF you will get when y is lies between 0 to 1, you will get $\frac{1}{4\sqrt{y}}$ by differentiating between the interval 0 to 1, you will get $\frac{1}{4\sqrt{y}}$.

When y is lies between 1 to 9, you will get a $\frac{1}{8\sqrt{y}}$; otherwise 0. So, this is a probability density function, you can verify also by $\int_0^1 \frac{1}{4\sqrt{y}} + \int_1^9 \frac{1}{8\sqrt{y}}$ you will get the answer 1 therefore, this is the probability density function of Y. It is a very important problem, because of this interval [-1, 3], the CDF is changing between 0 to 1 and 1 to 9.

Suppose this problem would have a point x chosen at random in the interval [-1, 1], in that case the interval of y between 0 to 1 you have to inverse images you can apply the remarks of the theorem, which I have send it in the one dimensional random variable. You can get the probability density function of -1 to 0, one density then 0 to 1 another density you can sum it of you can get the probability density function of Y or you can apply the CDF method to get the answer.

If the question is at the point is chosen at random in the interval [1, 3], then the interval is going to be has only one inverse then also the problem is going to be the different form. So, in these because of -1 to 3, in some portion it has a two inverse images in some portion it has only one inverse therefore, the you are partitioning in the interval in to piece by piece finding the CDF separately, then you are combining everything.

(Refer Slide Time: 16:50)



We will move to the next problem, the problem is a length of lifetime that is denoted by capital T, unity is in hours of a certain device as an exponential distribution with the mean 100 hours. Calculate the reliability at time T is equal to 200 hours.

This is a very typical problem in reliability analysis most of the time we make the assumption the lifetime of an electrical component follows exponential distribution. In this problem also they made already the assumption that the lifetime or the time in which this electrical certain device is going to work, that follows exponential distribution, why it is exponential? Because the exponential distribution range is from 0 to ∞ and most of the time you never know when it is going to pay. So, it is always a time that is great now equal to 0. Therefore, the lifetime most of the time we make the assumption, that follows exponential distribution.

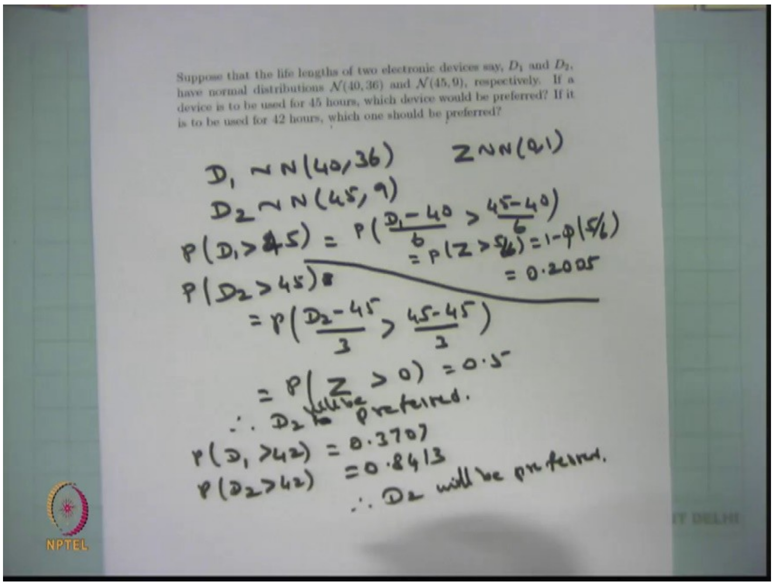
Based on the statistical history, they find the parameter value or estimate the parameter values. So, here already the parameter values are given, that is mean 100 hours; that means, the lifetime that is T that is a random variable that follows an exponential distribution with the parameter, we always write a parameter λ in that case the mean is going to be $1/\lambda$. So, here the mean is applied that is 100 hours. Therefore, the parameter is going to be 1/100, because the unit of the random variable T that is in hours, and the parameter is 100 hours. mean and the parameter relationship is reciprocal. Therefore, that is exponentially distributed with a parameter 1/100. Therefore, you can immediately write down what is the probability density function of a this, that is a it is $\lambda e^{-\lambda t}$, where t is lies between 0 to infinity; 0 otherwise.

In this problem the λ is 1/100 by substituting and λ = 1/100, you will get the probability density function of random variable T in this problem. So, the question is, calculate the reliability at time T is equal to 200 hours; that means, a you should know what is the definition of reliability. Reliability is nothing, but the probability that the system is working till this time or equivalent of saying the probability that the component missing after this time that is called the reliability. So, at least it works this which time; that means, a here the lifetime T is this much time working that is a life.

So, therefore, the reliability at a time T is equal to 200 that is same as the in notation it is a reliability at 200, that is same as the probability that T is going to have a life more than 200 hours. The reliability at 200 that is same as probability that the lifetime of this device is going to have more than or equal to 200, that is same as sorry not more than or equal to that is going to be greater than 200. It is a continuous type random variable. So, whether you right greater than or equal to does not matter, but the reliability is defined probability that the lifetime is going to be more than that time.

So, this is same as 1 – P{T ≤ 200} either you find out the P{T ≥ 200}; that means, you integrate the between the interval 200 to infinity the probability density function or you find out probability of T is less than or equal to 200 by integrating 0 to 200 the probability density function then substitute here both are all the same. So, this is same as 1 - $\int_{0}^{200} \lambda e^{-\lambda t} dt$. You simplify he will get the e^{-2} that is going to be the answer. So, numerically it is going to be 0.13534. So, this is the reliability of a device at time T is equal to 200, like this you can create many more problems with this setup and you can solve.

(Refer Slide Time: 22:08)



Now, will move into the next problem suppose that the life lengths of two electronic devices say D_1 and D_2, have normal distributions. I am using the notation capital N means it is a normal distribution, the first parameter 40 is a mean and the second parameter 36 that is variance of the first normal distributed random variable. The second one that is also normal distributed random variable with the mean 45 and variance 9. If a device is to be used for 45 hours, which devices would be preferred; if it is to be used for 42 hours, which want to be preferred.

So, there are two questions, in the earlier problem we made a lifetime follows exponential distribution. Here we made a life time follows a normal distribution with the mean 40 and 45 respectively, variance 36 and 9 respectively for two different devices. So, these are all the assumptions we make the assumption the lifetime follows normal distribution and so on and we get the results. So, what we can do we can write the problem that is a D_1 is a random variable that is a life time length of first device, which follows a normal distribution with the mean 40 and the variance 36. Similarly, the D_2 is a second random variable denotes a life length of the second electronic device which is normal distribution with the mean 45 and variance 9.

So, the question is if the device is to be used for 45 hours which devices would be preferred. So, what you will do? We will try to find out what is the probability of if D_1 is going to be greater than 45, what is a value and similarly you will find out what is the probability of D_2 is going to be greater than 45. So, whichever has the more probability, you will go for preferring that if a device is used for 45 hours. So, whichever has the more probability of working more than 45 hours we will prefer that electronic device. Let us find out the P{$D_1 > 45$} that is same as probability of whenever you have a problem in the normal distribution, first you have to convert into the standard normal distribution then use the table to get the numerical value of the probabilities. So, D_1 minus mean is 40 divided by standard deviation that is 6. Here also you have to do the same thing $\frac{45-40}{6}$ which is same as.

Now, the D_1 is normal distributed, normal distribution minus their mean divided by the standard deviation, that becomes a standard normal. So, I use a notation Z for standard normal. Z is a standard normal distributed; that means, a mean is 0 and variance is 1. So, probability of Z is greater than 45 minus 40. So, it is 5. So, it is 5. So, you find out the P{Z > $\frac{5}{6}$} by using the table that is a 1 – φ(5/6). I have already defined what is the meaning of φ(x); that means, the integration from minus infinity to x, the probability density function of standard normal distribution that is going to be φ(x). So, you get the value from the table. So, this value is going to be 0.2005.

Now, we will compute the P{$D_2 > 45$}. So, that is going to be the same way that is a probability of a D_2 minus there mean. So, the mean of second device is 45 and variance is 9. Therefore, the standard deviation is 3 greater than $\frac{45-45}{3}$. $\frac{D_2-45}{3}$ that becomes Z standard normal distribution greater than 45 - 45 that is 0. Probability of a standard normal distribution greater than 0, you know that it is symmetric about Z is equal to 0 therefore, the whole area is divided into 50-50 percentages the whole area is 1.

So, the P{Z > 0} that is a from 0 to infinity, that is going to be 0.5. Now we got the result P{D_1 > 45}, that is 0.2005 and P{D_2> 45} is 0.5; that means, the second device the probability of a second device can work for more than 45 hours is more than that of first complement therefore, D_2 is preferred. Second question if it is used for 42 hours which one should be preferred. Now, we will go for the similar exercise for P{$D_1 > 42$}, if you do the simplification for the problem you may get the answer 0.3707, similarly P{$D_2 > 42$} that is going to be 0.8413.

Again, the probability of D_2 greater than 42 is more compare into the that of D_1. So, again D_2 will be preferred. D_2 will be preferred in the first case as well as in the second case. You see the problem both are normal distributed which has the different mean different variance. Therefore, suppose the means are same and the variance are going to be different, then you can conclude something else. Suppose the variance are same the means are different then also you can have the different issue. But here the means as well as the variance are different therefore, unless otherwise you compute the probability you cannot conclude.

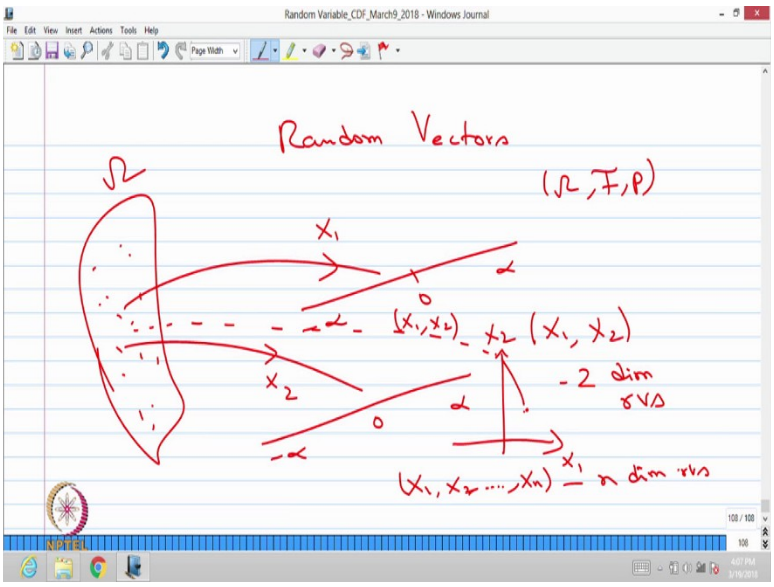
So, we are concluding which one is better by getting the probabilities of both the scenarios. So, in this lecture we have discussed the six problems, three from discrete type and three from the continuous type and some more problems you can see it from the assignment sheet or the problem sheet. And when you solve the different problem then you will come to know how to use the different standard distributions and getting the answer.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 05
Higher Dimensional Distributions
Lecture - 24

We are moving into the 5th model; let me recall first model we discussed the basics of probability. And the second model we discussed the random variable and the third model we discussed the movements and inequalities and the fourth model we discussed the standard distributions both the discrete and continuous type.

(Refer Slide Time: 00:53)



Now, we move in to the 5th model that is random vectors. In most of the random experiment you may have Ω in which you may be able to create many random variables simultaneously. And also sometimes the interest will be finding the some measures together or given one set of measures takes this value, what could be the distribution of other random variable? Or sometimes you may interest to find the distribution of more than one random variable at the same time. So, till now we have discussed only one random variable and then later we have discussed a function of a random variable; that means, again that is another one random variable. There is a possibility you may need to study more than one random variable at the same time or simultaneously or together, in that case you need a random vector.

When we use the word random vector; that means, it is n dimensional random variable. That means, we put n random variables together that we call it as n dimensional random variables or random vectors of size n. That means, each one is random variable and we are studying n random variables together or jointly. Many real world problems you may need to study the distributions of a more than one random variable jointly therefore, we need a random vector. That means, whatever we have understood the concept of a random variable and their distributions and movements; the same concept has to be extended from one dimensional to multi-dimensional.

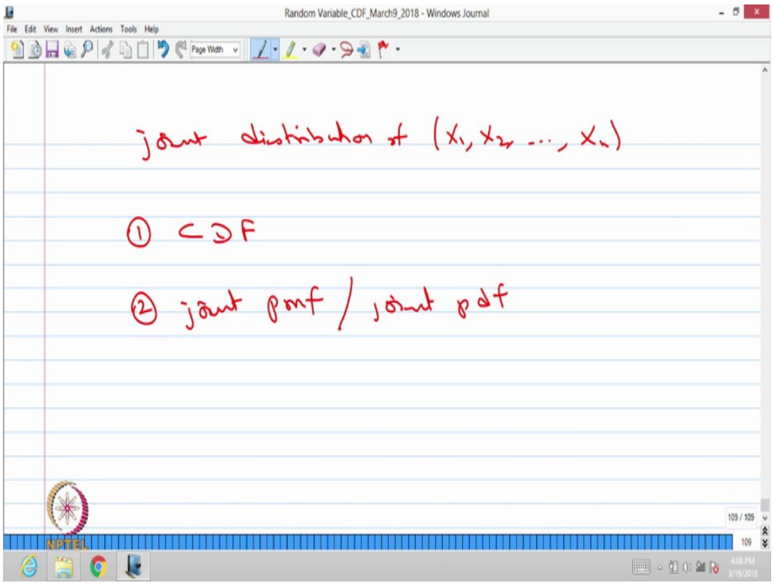
So, whatever the calculus we have used, calculus means finding the integration or finding the derivative and so on; whatever we have done the calculus part with the one-dimensional variable or one variable, now we have n dimensional random variable. So, the corresponding calculus of several variables has to be used. Let me explain it how the random vector coming into the picture. The **Ω** consists of many samples, it could be finite or countable infinite or uncountable many. The random variable one random variable is defined as X_1.

Another random variable defined from the **Ω** that is X_2 like that there are many more random variables defined in the **Ω** that is a real valued function from **Ω** to R satisfying the condition that is $X^{-1}(-\infty, x]$belong to F then only the real value function is a random variable. That means, you have a probability space (Ω, F, P), **Ω** is a collection of all possible outcomes and F is the sigma field on **Ω** and P is the set function satisfying the three conditions the Kolmogorov axiomatic condition. Therefore, P is the probability; so, this is the probability space.

In this probability space X_1 is defined, the same probability space X_2 is defined like that many more random variables are defined in the same **Ω**. Now we are going to create a random vector by making few random variables together or jointly. That means, if you go for X_1, X_2 together then this is called two dimensional random variables or it is a random vector of size two, X_1 and X_2. That means, on the same **Ω**, (X_1, X_2) is defined such that the possible values are going to be in the two dimensional plane.

This is X_1 and this is X_2. Earlier the X_1 is defined from **Ω** to R now the (X_1, X_2) is defined from **Ω** to R × R. So, again each one X_1 is random variable X_2 is a random variable together we call it as a random vector of size 2 or two dimensional random variable. The same way one can define n dimensional random variable X_1, X_2 and so on till X_n; each one is a random variable therefore, this is a n dimensional random variable, in short rv's random variables.

(Refer Slide Time: 06:45)



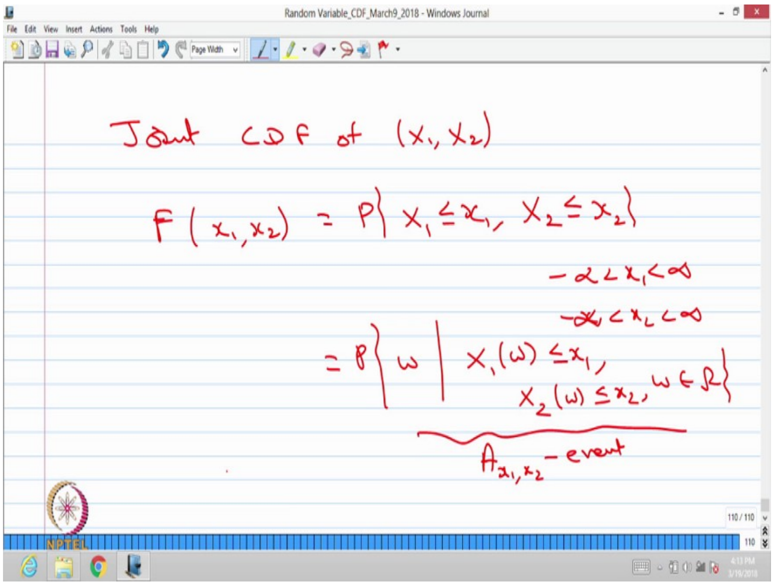
That means, we can go for finding what is the joint distribution of n dimensional random variable $(X_1, X_2, \ldots, X_n)$.

Earlier we used a word distribution of X_1, now we are studying the distribution of n random variable together or jointly. Therefore, it is called a joint distribution of n dimensional random variables. So first we can discuss; what is the CDF Cumulative Distribution Function; if it is a discrete type random variable one can discuss; what is a joint probability mass function; if each random variables of the continuous type then one can study what is a joint probability density function of n dimensional random variable.

Initially, we study all the random variables of the discrete type or all the random variables are of the continuous type. Later we will see one random variable of the discrete type and the other one is of the continuous type and so on that we will do it in the later.

So, our interest is to find the joint distribution of n dimensional random variables in which each one is a random variable; that means, that satisfies the definition of random variable. Therefore, this is a n dimensional random variable. first let us go for how to give the definition of joint CDF.

(Refer Slide Time: 08:31)



Joint CDF of instead of going for the n dimensional random variable first we will restrict to the two dimensional random variables. Once we discuss the two dimensional random variables CDF, then it is easy to visualize or easy to understand for more than two dimensional random variable. It is (X_1, X_2). The joint CDF can be written as the two functions (x_1, x_2) is nothing but the $P\{X_1 \le x_1, X_2 \le x_2\}$.

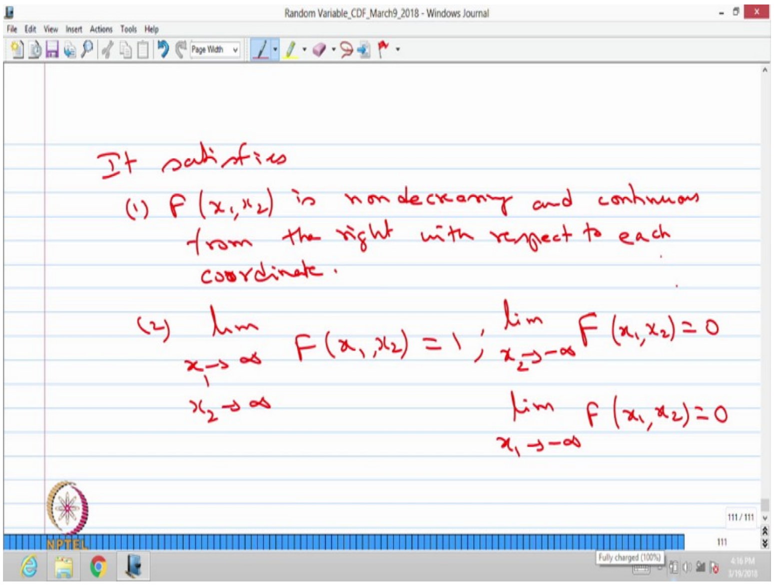
Both $x_1 \wedge x_2$ lies between -∞ to ∞. Similar way we have defined the CDF of one dimensional random variable, now we are defining the CDF of two dimensional random variable with the 2 variables (x_1, x_2) is nothing but the $P\{X_1 \le x_1, X_2 \le x_2\}$.

That is nothing but $P\{w \vee X_1(w) \le x_1, X_2|w| \le x_2\}$. That means, this value can lie between minus infinity to small x_1. And the w's belonging to **Ω**. That means, we are collecting a few possible outcomes satisfying the condition under the operation X_1 it should give the value less than or equal to x_1 under the operation of X_2 it gives values less than or equal to small x_2.

So, we are collecting those possible outcomes then finding the probability and that probability of these possible outcomes satisfying this condition that is going to be the CDF at the point (x_1, x_2) where x_1 *can* lie between -∞ to ∞, x_2 can lie between -∞ to ∞. The collection of w such that this condition we can label this as the A_{x_1, x_2} this is an event because whenever you collect a few possible outcomes; that is nothing but the event.

So, the event A_{x_1,x_2} that is an event; that means, the probability of event by using a Kolmogorov axiomatic definition, probability of any event or P of any event always greater or equal to 0 and P(Ω) = 1 and P$\left(\cup A_i\right) = \sum_i P(A_i)$ as long as A_i's are mutually disjoint events. With the same logic the P of any event getting from the different values of $x_1 \wedge x_2$; this is also going to satisfy the Kolmogorov axiomatic definition. Therefore, one can make what are all the conditions or what are all the properties is going to be satisfied by this capital F.

(Refer Slide Time: 12:43)

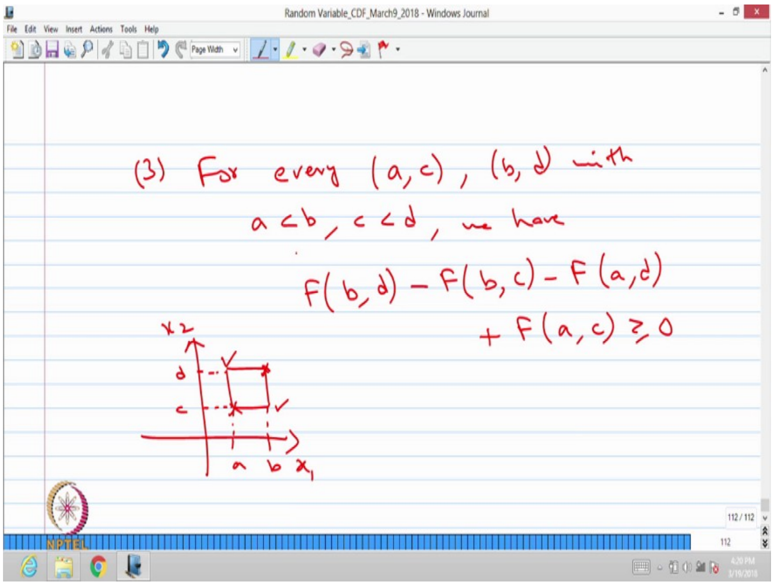


So, let me write down; it satisfies, it satisfies the first condition the $F(x_1, x_2)$ that is non decreasing and continuous from the right.

That means it is a right continuous with respect to each coordinate, each coordinate that is $x_1 \wedge x_2$ it is a non-decreasing as well as a continuous from right with respect to each coordinate that is a first point. Second point the $\lim\limits_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F(x_1, x_2) = 1$. $\lim\limits_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F(x_1, x_2)$ is always 1.

And the $\lim\limits_{x_2 \to -\infty} F(x_1, x_2)$ = 0 and $\lim\limits_{x_1 \to -\infty} F(x_1, x_2)$ = 0. When both becomes positive infinity, it becomes 1 when either one of them is minus infinity and $\lim\limits_{x_2 \to -\infty} F(x_1, x_2)$ or $\lim\limits_{x_1 \to -\infty} F(x_1, x_2)$ is going to be 0.

(Refer Slide Time: 15:11)

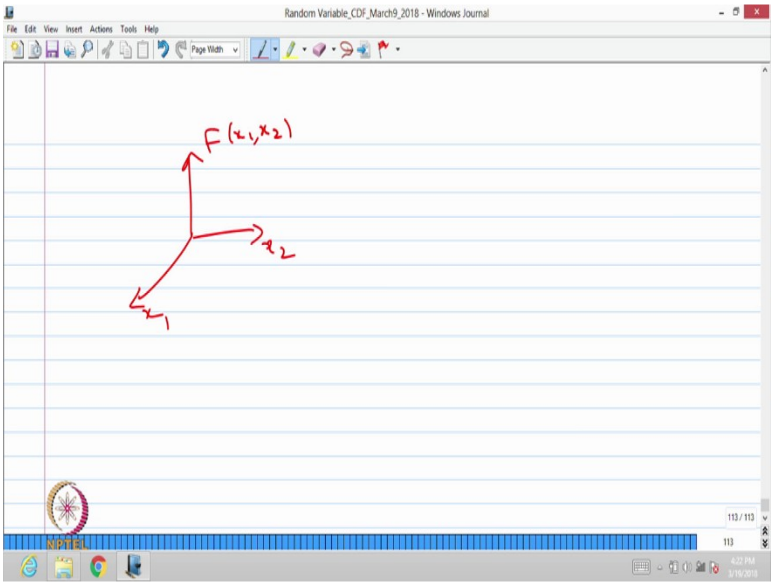


The third point for every (a,c) and (b,d) with a < b and c < d; we have I can just draw the diagram first then I can go for it that is easy. So, I can take as small box in which this is going to be a and this is going to be b and this is going to be c and this is going to be d. I can go for after drawing the diagram, now I can go for F(b, d) – F(b, c) -F(a, d)+ F(a, c) ≥ 0.

So, the cross is going to have a positive symbol and a tick mark has a negative symbol; that means, F evaluated at (b, d) - F evaluated at (a, d) and (b, c) with the minus sign then plus F evaluated at (a, c) that value has to be greater or equal to 0. Whenever you have two dimensional random variable whose CDF always satisfies these three conditions. The third condition is a very important condition in the sense even you may have a real valued function with two variables satisfying these 2 conditions may not be a CDF of one dimensional random variable unless otherwise it satisfies a third condition.

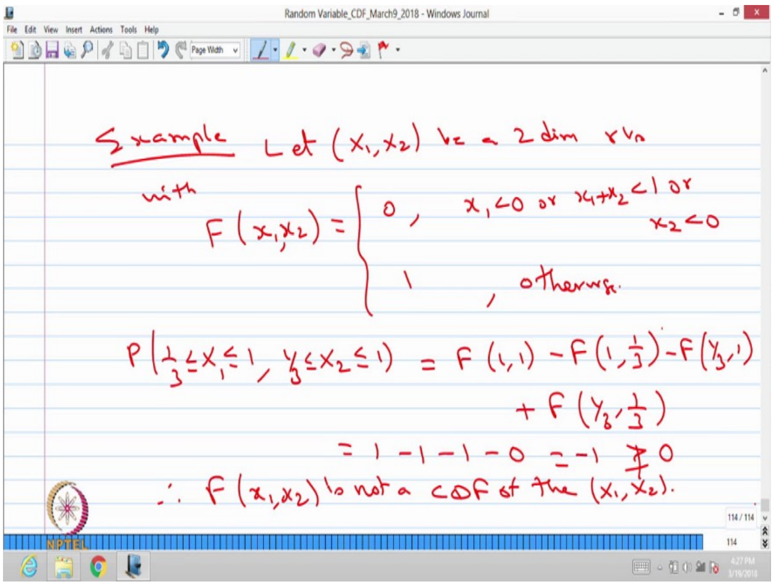
But if you have two dimensional random variable, you will have always unique CDF with the two variables that satisfies all these three conditions. CDF of two dimensional random variables can be represented in the form of graphical.

(Refer Slide Time: 18:01)



So, one can visualize the CDF in the form of x axis x_1, x_2 and this is going to be CDF of (x_1, x_2). So, this is possible only for two dimensional X_1 is one random variable, X_2 is another random variable. So, z axis that is a CDF of (x_1, x_2) whereas, you cannot visualize you cannot make a graphic representation of more than two dimensional random variable.

(Refer Slide Time: 18:44)



Let us go for one simple example in which we can conclude whether this is going to be CDF or not. As example let (X_1, X_2) be a two dimensional random variables with CDF F(x_1, x_2) that is either 0 when x_1 < 0 or $x_1 + x_2$ < 1 or $x_2 <$ ¿ 0. It takes a value 1 otherwise. Verify whether the capital F is going to be the CDF of two dimensional random variable. That means, this is real valued function with 2 variables.

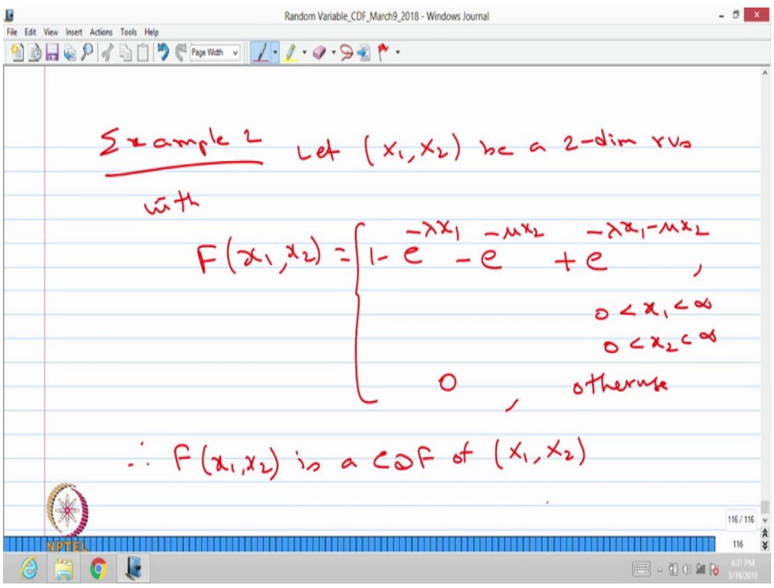
Whether this satisfies the three properties which we have given; if all these 3 properties are satisfied then you can conclude this will be the CDF of two dimensional random variable. You can easily verify the first 2 properties the function is a non decreasing as well as continuous from the right. Similarly, you can easily verify the $\lim\limits_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F(x_1, x_2)$ that value is 1. Either x_1 is $-\infty$ or x_2 is $-\infty$ the limit is going to be 0, that also can be easily verified.

We have to verify the third condition that is for any 2 points; the CDF of the difference at 2 points with a positive sign, 2 points with the negative sign has to be greater or equal to 0. For example, you do it with the $P\{\frac{1}{3} \leq X_1 \leq 1, \frac{1}{3} \leq X_2 \leq 1\}$.

If you compute this probability that is nothing but F(1, 1) – F(1, 1/3) – F(1/3,1) + F(1/3, 1/3). If you substitute the value F(1, 1) that is 1, F(1, 1/3) that is again 1, F(1/3, 1) again 1, F(1/3, 1/3) that is 0. And this value is going to be minus 1 which is not greater or equal to 0. So, for any arbitary points this third property has to be satisfied.

So, since the third property is not satisfied; you can conclude this F is not a CDF of the random vector (X_1, X_2) or the two dimensional random variables (X_1, X_2).

(Refer Slide Time: 22:58)

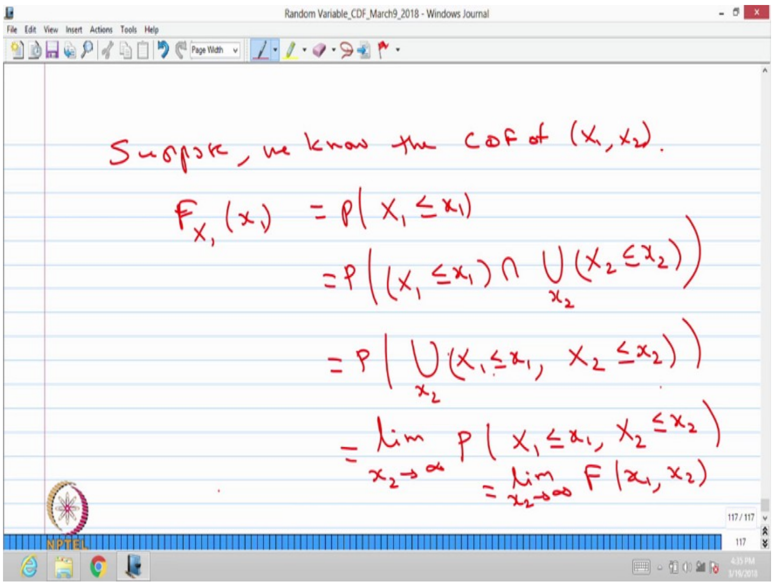


We will give some example in which that is going to be CDF. So, example 2 let (X_1, X_2) be a two dimensional random variables with the function F(x_1, x_2) is $1 - e^{-\lambda x_1} - e^{-\mu x_2} + e^{-\lambda x_1 - \mu x_2}$.

So, this is going to be the value when both x_1 and x_2 lies between 0 to infinity; otherwise it is 0. We can verify whether this is going to be the CDF of two-dimensional random variables. By seeing this function, you can easily say when x_1 and x_2 is the positive infinity, it becomes 1; either x_1 or x_2 is going to be minus infinity that is going to be 0. And it is a non-decreasing function and continuous from the right. Therefore, the properties 1 and 2 are easily satisfied.

For different values one can able to verify the third property also will be satisfied therefore, hope one can conclude this is going to be the CDF of two-dimensional random variables. So, I am not giving the proof of the third property. But that can be verified we can conclude this is going to be the CDF of two-dimensional random variables.

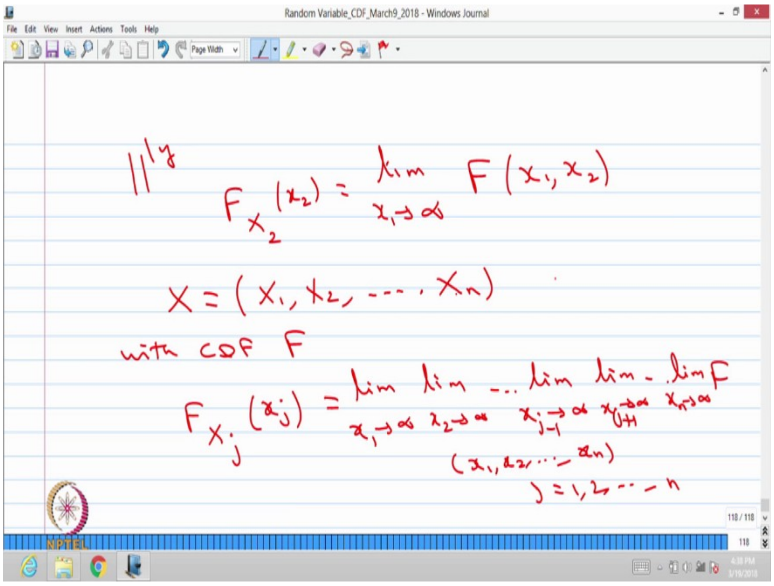
(Refer Slide Time: 25:08)



Suppose we know the CDF of (X_1, X_2); one can able to find the CDF of any one random variable, that is one can find the CDF of random variable X_1 as a function of $x_1$that is nothing but P$\{X_1 \le x_1\}$. That is same as probability of $X_1 \le x_1$ which intersect ¿ $x_2 (X_2 \le x_2)$. That is same as the probability of union of all possible values of small x_2, $X_1 \le x_1$ and $X_2 \le x_2$.

That is same as the P of union is nothing but $\lim_{x_2 \to \infty} P\{X_1 \le x_1, X_2 \le x_2\}$. That is same as $\lim_{x_2 \to \infty} F(x_1, x_2)$. That means, if you know the CDF of (x_1, x_2) by taking limit of the other variable tends to plus infinity that will give the CDF of the other random variable.

(Refer Slide Time: 27:41)



Similarly, the CDF of the random variable X_2 as a function of x_2 that is nothing but $\lim_{x_1 \to \infty} F(x_1, x_2)$; so, this is valid for two dimensional random variable. So, this say concept can be extended to n dimensional random variable. That means, suppose I denote X as n dimensional random vector; instead of again and again writing n dimensional random variable, I can write X with the CDF F; that means, F also has n variables. Then we can find the CDF of any one random variable suppose I want j^{th} random variable.

Suppose, I want to find out the CDF of j^{th} random variable that is nothing but $\lim_{x_1 \to \infty} \lim_{x_2 \to \infty} \ldots \lim_{x_{j-1} \to \infty} \lim_{x_{j+1} \to \infty} \ldots \lim_{x_n \to \infty} F$, I assume that j is in between 1 to n, which has elements just I will write down in the next line $(x_1, x_2, \ldots, x_n)$.

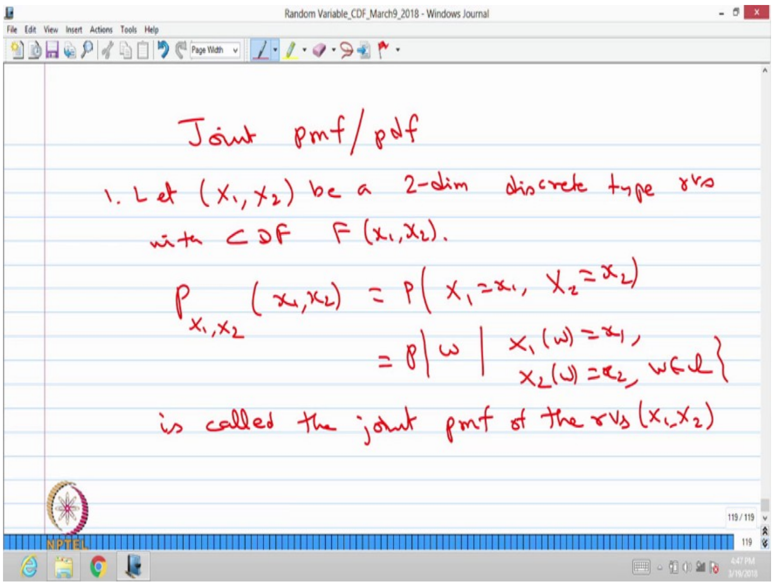
So, this is the way one can get the CDF of one-dimensional random variable from the CDF of n dimensional random variable, where j can be 1, 2 and so on till n. In this exercise, I have not said whether the random variable is of the discrete type or the continuous type or mixed type. So, based on each random variable or of the discrete continuous or mixed; one can discuss the joint probability mass function or joint probability density function and so on. So, at present we do not know; what is a type of each random variable.

Therefore, we stopped it at the CDF of the n dimensional random variable. Once we know the type of each random variable, then one can go for studying the probability mass function jointly, probability density function jointly that is going to be the second lecture.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 25

(Refer Slide Time: 00:04)



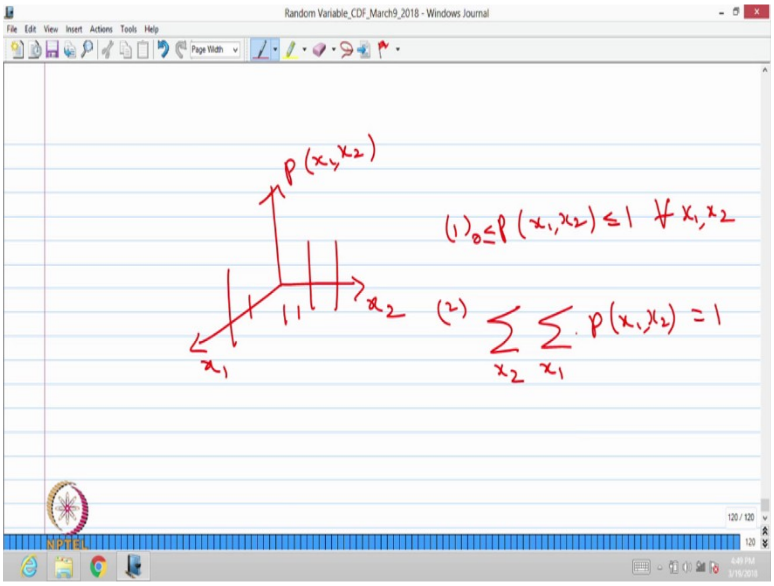
Now, we will move into joint probability mass function, and joint probability density function of n dimensional random variables, or random vector of size n.

Let me start with the joint probability mass function. Let me start with the 2 dimensional that is easy. Let (X_1, X_2) be a two-dimensional discrete type random variable; that means, X_1 is a discrete type random variable, as well as a X_2 is also discrete type random variable with the CDF capital F(x_1, x_2).

One can define the probability mass function in together; that is, P(x_1, x_2). That is with the variable (x_1, x_2); that means, the P($X_1 = x_1, X_2 = x_2$). Where x_1 is the images of X_1 or ranges of X_1, and $x_2$is the ranges of the random variable X_2 or the images of X_2. Put together that is P($X_1 = x_1, X_2 = x_2$). This is nothing but the P{w|$X_1(w) = x_1, X_2(w) = x_2$}. and w belonging to Ω; that means, this is the event collection of possible outcomes satisfying this event.

So, P of this event that is the probability of event satisfying this condition. So, this function is called the joint probability mass function of the random variable (X_1, X_2). This is the probability mass at the point $(x_1, x_2$¿.

(Refer Slide Time: 03:08)

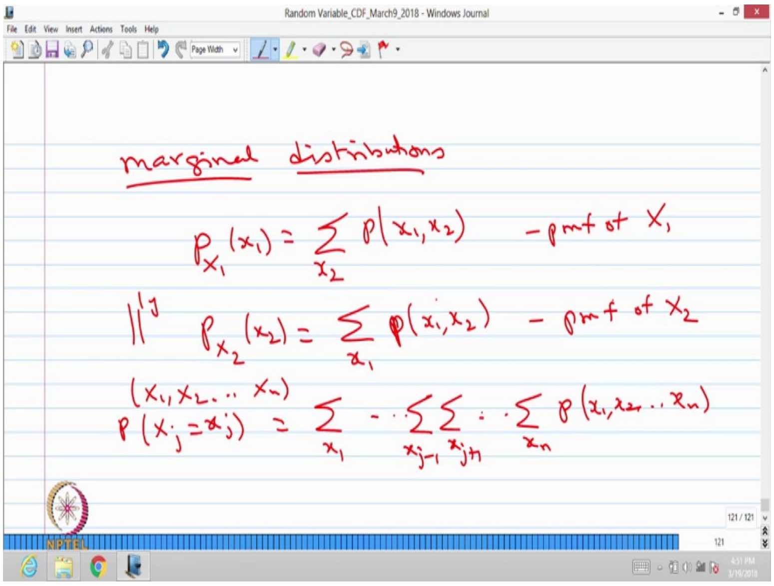


You can go for graphical representation of joint probability mass function $(x_1, x_2$¿. So, this is the probability mass function of $(x_1, x_2$¿; that means, at some point in the two dimensional plane $x_1 x_2$ plane, whatever the smaller heights; whatever the heights, that is going to be the probability mass function at the point $(x_1, x_2$¿.

Both are discrete type random variables; therefore, this can be represented in the three dimension plane x_1 is the one axis coordinate, and x_2 is another coordinate, and height is z axis is the probability at the point $(x_1, x_2$¿. The joint probability mass function satisfies two properties this is always going to be lies between, it always lies between 0 to 1 for every x_1, x_2.

The second condition if you make a $\sum_{x_2}\sum_{x_1} P(x_1, x_2)$ that is going to be 1; that means, if you add all the heights over the $x_1 x_2$ plane that addition is going to be 1; that means, wherever there is a mass it has to be greater than 0 if you had all the masses that is going to be 1. From the joint probability mass function, one can get the probability mass function of $x_1 \wedge x_2$, they are called marginal distributions.

(Refer Slide Time: 05:08)

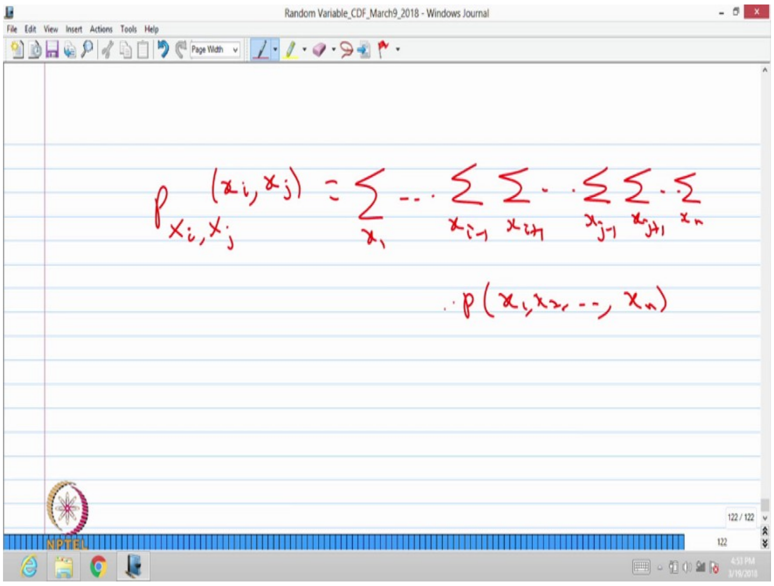


That means if I want to find out the probability mass function of X_1 is $\sum_{x_2} P(x_1, x_2)$, I can get the probability mass function of X_1. We can verify whether this is going to be the probability mass function, in this summation this value always going to be greater or equal to 0, lies between 0 to 1, and if you make a summation over x_1 that is going to be double summation over x_1 and x_2. That is going to be one therefore; this is the probability mass function of the random variable X_1.

Similarly, one can find the probability mass function of X_2 is $\sum_{x_1} P(x_1, x_2)$ so, this is the probability mass function of X_2. The way we have done, we can go for n dimensional random variable, then we can get the probability mass function of any one random variable by summing over the joint probability mass function of X_1¿$X_n$except j^{th}variable X_j.

We can get the marginal from the joint distribution; from the joint probability mass function of n dimensional.

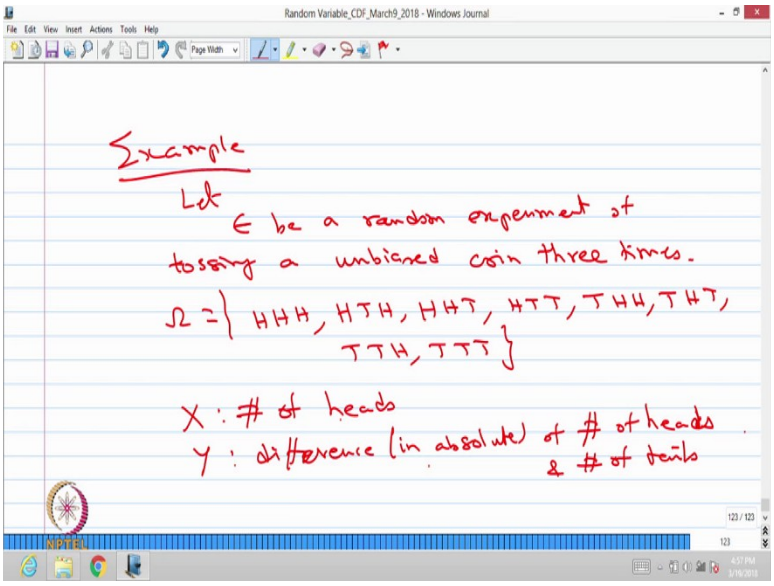
(Refer Slide Time: 07:24)



I can find the joint distribution of X_i and X_j by $\sum_{x_1} .. \sum_{x_{i-1}} \sum_{x_{i+1}} \ldots \sum_{x_{j-1}} \sum_{x_{j+1}} \ldots \sum_{x_n} P(x_1, x_2, \ldots, x_n)$. That means, by n- 2 summations without x_i and x_j, one can get the joint probability mass function of X_i, X_j. That means, always from n dimension random variable either CDF, or if they are discrete type random variable, you can get the lesser distributions of jointly by summing it over the other variables.

So, by doing again and again you can get the marginal distribution of one random variable. So that means, from n random variables you can get the joint distribution of n - 1, then n - 2 and so on finally, you can get the marginal distribution of any random variable. Let us go for one simple example how one can visualize the two dimensional discrete type random variable as an example.

(Refer Slide Time: 09:22)



Let E be a random experiment of tossing an unbiased coin three times. The random experiment is tossing an unbiased coin three times.

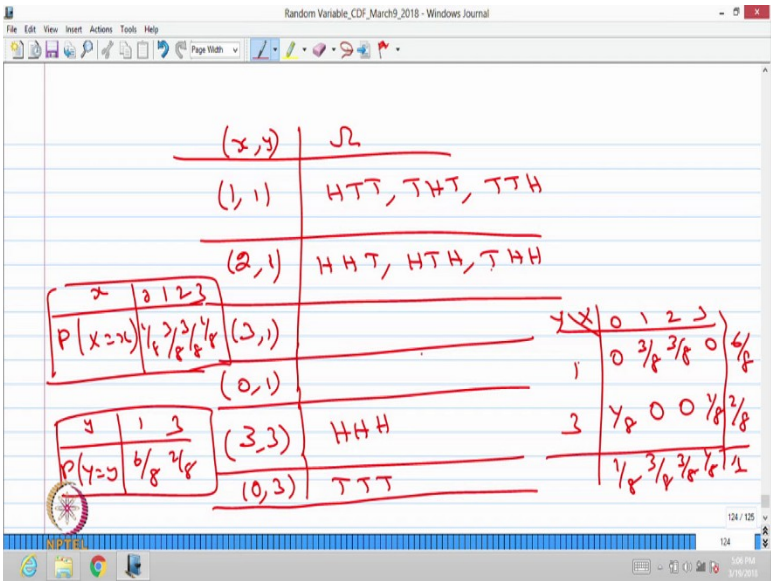
Therefore, the Ω is going to be the collection of all possible outcomes, i.e., I use the notation H for getting head, T for tail. So, since we are tossing an unbiased coin three times, therefore, you will have a 2^3; so you have 8 possibilities. So, head head head; head tail head or head head tail, and head tail tail, then tail head head, tail head tail, tail tail head, then last tail tail tail. So, these are all the 8 possibilities, or 8 possible outcomes of this random experiment of tossing an unbiased coin three times.

Now, I am going to define two random variables in this random experiment. And our interest is to find out the joint distribution of these two random variables, first let me define first random variable X as a number of heads in tossing a unbiased coin three times. The random variable Y is nothing but difference in absolute of number of heads and number of tails.

You see it very carefully, the random variable X is number of heads; whereas, the random variable Y is difference in absolute of number of heads and the number of tails. Therefore, you should know what are all the possible values of X, what are all the possible values of Y, then you can conclude what type of the random variable X and Y. Then you can go for finding out the distribution based on whether it is a discrete or continuous. The way the X define the number of heads, and the random experiment is a tossing a coin unbiased coin three times. Therefore, there is a possibility you will get no times head or one times head or 2 times head or 3 times head.

Therefore the possible values of X; that is 0, 1 or 2 or 3, whereas, the Y is the difference in absolute of number of heads and number of tails, therefore, the possible values of Y is going to be 1 and 3 because of a difference in absolute.

(Refer Slide Time: 13:47)



Therefore, you can go for make out the table of different values of (x, y), and what is the collection of possible outcomes which is going to give the values of (x, y).

For example, suppose you go for x takes a value 1; y takes a value 1; that means, number of heads is 1, and the difference in absolute with a number of heads and tails that is also 1. That means, the possible outcomes from the **Ω**; that is head tail tail or tail head tail or tail tail head. All these 3 possibilities give the value of (x, y)is (1, 1). Similarly, you can go for what are all the possible outcomes which gives the values (2, 1), that is going to be the number of heads is going to be 2, and the difference in absolute with the number of heads and tail that is going to be 1. Therefore, it is going to be head head tail, head tail head, tail head head. The next one you can go for finding (3, 1).

If you go for (3, 1), you will get no possible outcomes. Similarly, if you go for (0, 1), there also you would not get any possible outcomes. If you go for (3, 3) number of heads is 3, and difference in absolute heads with a tail that is also 3; that is possible with the head head head.

Similarly, you can go for (0, 3); number of heads is 0, and the difference in absolute that is going to be 3 that is possible with the tail tail tail. You see that there are totally 8 possible outcomes. So, one we have 3; other we have 3, and other we have one and one; so, that total is going to be 8.

Therefore, now we can go for finding out the joint probability mass function of (X, Y) using this box. That is when x takes a value, when x takes a value 0, 1, 2 or 3, and y takes a value 1 or 3. We can make a table x takes a value 0; y takes a value one that is nothing. Therefore, the probability is 0, when x takes a value 1, y takes a value one that is 3 possibilities. It is an unbiased coin therefore, the probabilities going to be 3/8.

When x takes a value 2 and y takes a value 1 there are 3 possibilities, therefore, this is going to be the 3/8. When x takes a value 3, y takes a value one and nothing, therefore, no possible outcomes therefore, empty set probability of empty set is 0. Similarly, (0, 3), that only one possibility so, 1/8. (1, 3), there is no possibility therefore, it is 0 and (2, 3), there is no possibility therefore, it is 0 and (3, 3) is only one possibility, therefore 1/8.

If you add all the values 3/8 + 3/8 + 1/8 + 1/8, that is going to be 1. If you make a row sum or column sum, you will get the marginal distribution, and if you had those values again you will get the one. So, it is 0 + 1/8 + 1/8 + 3/8 + 3/8, if you add up all these values it is going to be 1. Similarly, if you add 3/8 + 3/8, that is 6/8, 1/8 + 0 + 1/8 that is 2/8.

So, if you add 6/8 + 2/8 you are getting 1. That means, the probability mass function of X takes a value small x, that is going to be for different values of x it is 0, 1, 2 and 3. So, it is going to be for x takes a value 0, that is 1/8; for 1, 3/8; for 2, 3/8 and for 3, 1/8. So, this is going to be a probability mass function of X. Similarly, one can make a probability mass function of Y. So, different values of y are going to be 1 and 3. So, for one it is 6/8; for 3 it is 2/8 so, this is a marginal distribution of Y.

So, from this page one can get the joint probability mass function of X comma Y, from the joint distribution, you can always get the marginal distribution of X and Y separately. Or you can find out the marginal distribution from the random variable X itself, you do not need finding the joint distribution then the marginal of X.

You can find the way you have defined X, you can directly get the probability mass function of X. But here what am saying is if you know the joint distribution, you can always get the marginal distribution the other way or the converse is not drawn; that means, from the marginal one cannot get the joint always, whereas, from the joint distribution you can always get the marginal.

Therefore, here we get the probability mass function of X and Y from the joint probability mass function of (X, Y), this easiest example. Since both the random variables are of the discrete type, we are able to give the joint probability mass function of the random variable (X, Y). So, with this example let me complete the joint probability mass function. In the next class we will go for when both the random variables are of the continuous type, then one can define the joint probability density function that we will do it in the next class.

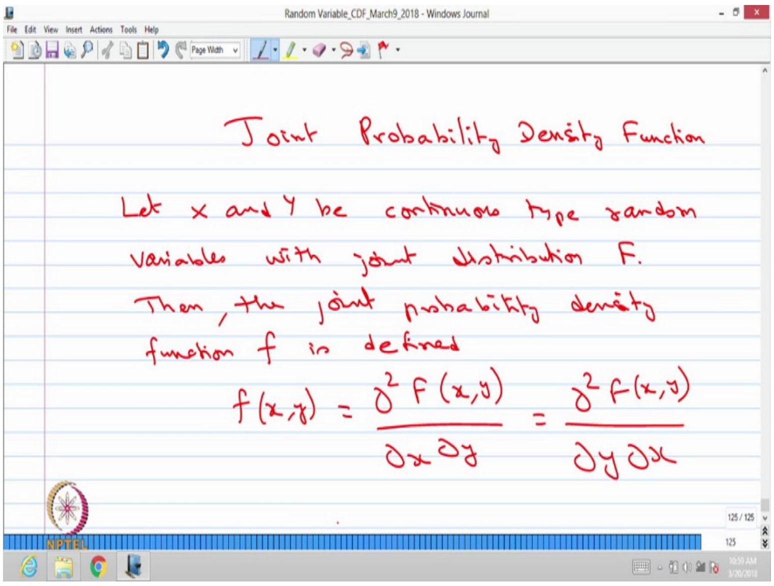
Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 26

In the module of random vectors, we started with the 2 and high dimensional random variables, and we discussed the joint distributions, first we discussed the CDF of two dimensional random variable then we have generalized this in to the n dimensional random variable, how the CDF look like with the several variables.

Then we start discussing joint probability mass function; whenever the underlined random variables are of the discrete type; that means, we have n dimensional random vector with each random variable is going to be of the discrete type.

(Refer Slide Time: 00:46)

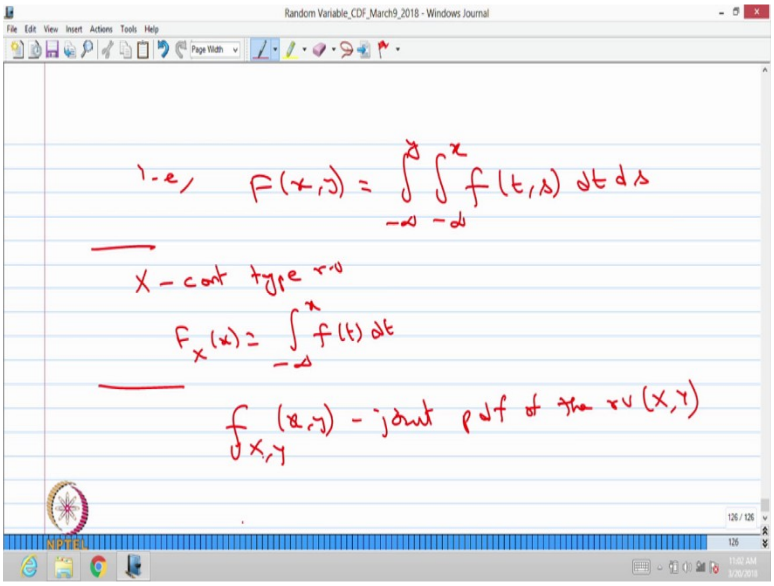


Now, we are coming in to joint probability density function. Let me start with the two dimensional random variable.

Let X and Y be continuous type random variables with joint distribution, that is F. That is basically a function of x, y. Then the joint probability density function, $f(x,y)=\frac{\partial^2 F(x,y)}{\partial x \partial y}=\frac{\partial^2 F(x,y)}{\partial y \partial x}$.

Since we have a continuous type random variable, the CDF is continuous function in both x and y. Since it is a continuous function in both x and y by taking $\frac{\partial^2 F(x,y)}{\partial x \partial y}$, that is going to be the probability density function of X, Y.

(Refer Slide Time: 03:11)

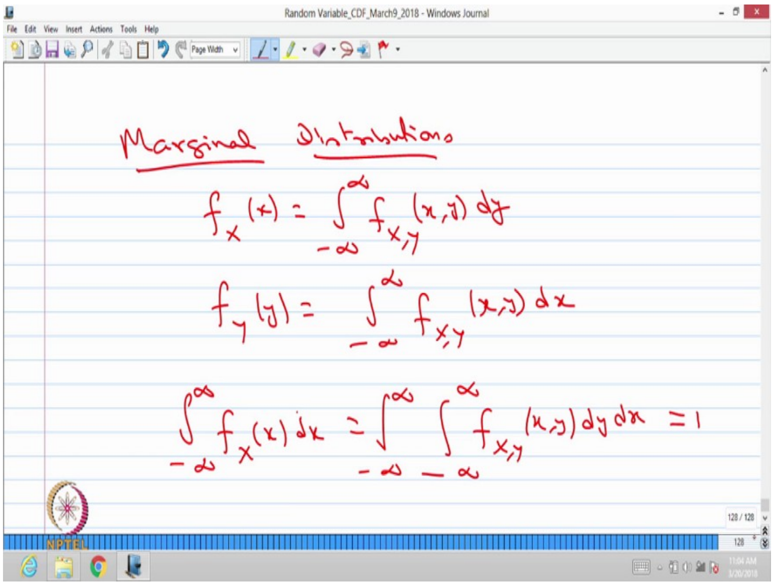


That is same as the CDF of two dimensional continuous type random variable can be written in the form of doubled integration with respect to x and y; that means, since both the random variables are of the continuous type, the CDF of two dimensional continuous type random variable is a continuous function in both x and y, one can write CDF in the form $\int_{-\infty}^{y}\int_{-\infty}^{y} f(t,s)\, dtds\cdot$

Here the integrant is the joint probability density function. It is similar to one dimensional random variable, in which the X is of the continuous type random variable then the CDF of X can be written in the form $\int_{-\infty}^{x} f(t)\, dt$and that is a probability density function of the random variable X. The same way we are writing a for a two dimensional continuous type random variable therefore, the probability density function is nothing but the $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ since it is a continuous function in both X and Y, whether you change the order of partial derivative does not matter it is going to be the c.

So, once you know the joint probability density function of X, Y, one can find the probability density function of one random variable. That means, from the joint probability density function so, here the X, Y small f so, this is the joint probability density function of the random variable X, Y. From the joint probability density function one can always get the marginal distribution of X, Y that is marginal distributions.

(Refer Slide Time: 05:48)



From the joint probability density function, one can get the marginal distribution of one random variable by $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$.

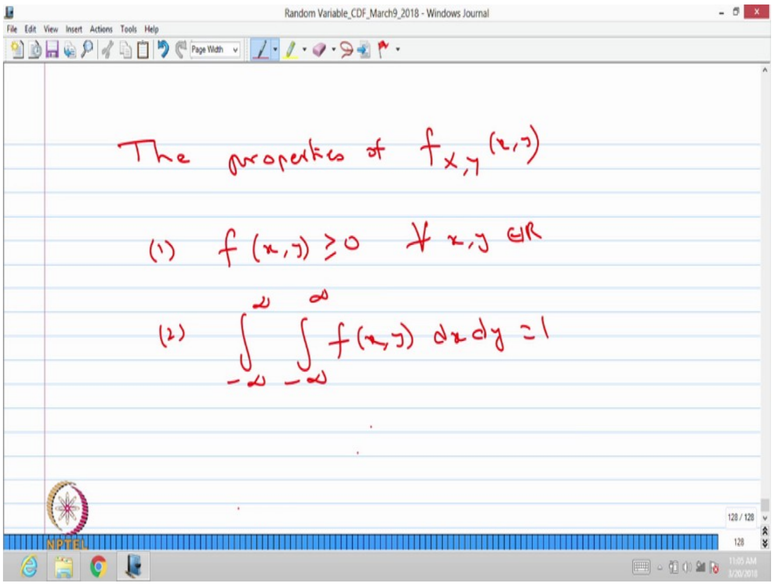
Similarly, we can get the probability density function of the random variable Y by $\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx$. The same thing we have done in for discrete type random variable. So, this is the continuous type random variable, therefore, the marginal distribution of X and Y can be obtained from the $f_{X,Y}(x,y)$.

One can verify how this is going to be the probability density function of X, and how this is going to be the probability density function of Y, because this function you have F(x) ≥ 0.

And if you integrate $\int_{-\infty}^{\infty} f_X(x)\,dx$, that is going to be $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy\,dx$.

Therefore, we know that $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dydx$, this is going to be 1.

(Refer Slide Time: 07:50)



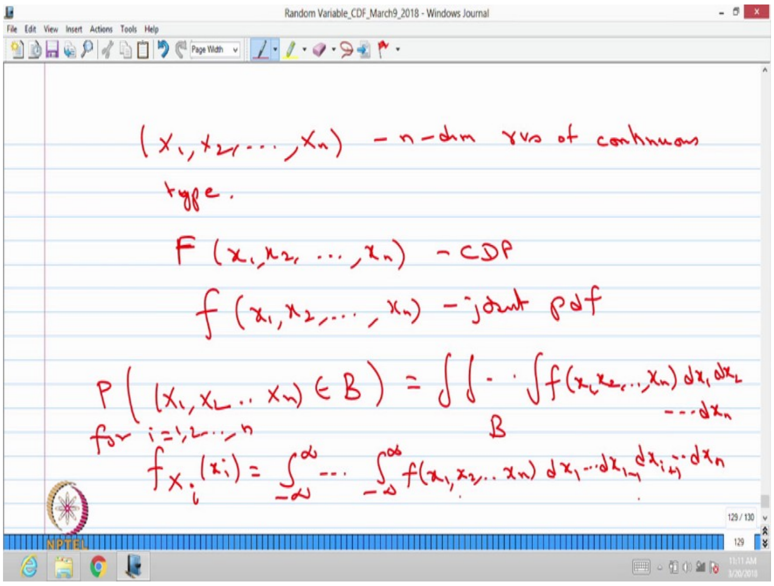
Therefore it is the properties of the joint probability density function. The properties of joint probability density function is, this is always going to be greater or equal to 0, for all x, y belonging to real and the second condition, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dydx$ is always going to be 1. That means, if you have any real valued function with the two variables satisfying these two conditions may be a joint probability density function of some two dimensional random variables.

If you have a two dimensional random variables of continuous type, then which has the joint probability density function that satisfies these two properties. The first property gives the value is going to be always greater than equal to 1, and has a double integration is going to be 1 that is nothing but the volume below the surface of F(x, y) that is going to be 1.

The way we say for the single random variable of continuous type the probability density function is greater or equal to 0, then the integration 1, means the area below the curve that is going to be 1. The same also here for a two dimensional random variable, the volume below the surface that surface is F(x, y), that is going to be 1. So, only one can visualize or make a graphical representation for a single dimensional random variable and two dimensional random variable not more dimensions. Therefore, we started the explaining two dimensional random variable with the graphical representation not for any more dimension.

The same concept can be extended for n dimensional random variable.

(Refer Slide Time: 10:06)



That means this is n dimensional random variables of continuous type; means each random variable is a continuous type random variable, therefore, we get n dimensional random variables of the continuous type; that means, it has the CDF with the n variables. This is the CDF of the n dimensional random variable. We have the joint probability density function with the n variables x_1, x_2 and so on.

This is the joint probability density function. Whenever you need to find out probability of $x_1, x_2, \ldots, x_n$ belonging to some Boral set some capital B, which is in the R^n the B is belonging to R^nthen that is nothing but finding out the probability is nothing but the n dimensional integration over the Boral set, B of the joint probability density function for n dimensional random variable, if you know the each random variable of the continuous type then you have a CDF.

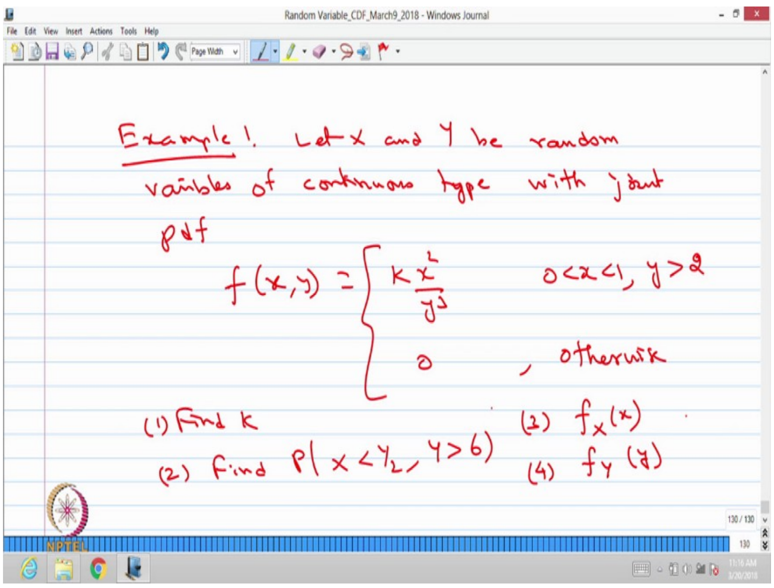
Similarly you have a joint probability density function, and you can always find probability of all the n dimensional random variable takes a values in the Boral set, where Boral set is from the R^nthat is nothing but n dimensional integration over a B of joint probability density function with respect to all the variables.

The way we got the probability belonging to the B, one can find the probability density function of any one random variable for i is equal 1 to n, you can always find the marginal distribution of any one random variable from the joint distribution of the n dimensional by

$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f\left(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n\right) d x_1 \dots d x_{i-1} d_{x+1} \dots d x_n,$ you can always get the marginal distribution of any one random variable from the joint distribution of n dimensional random variable which is of the continuous type. Now we will move into one simple example how one can describe the two dimensional random variable of the continuous type.

Then how one can discuss the marginal distribution and finding out the probability and so on.

(Refer Slide Time: 14:10)



Let us start with the easy example, the example is let X and Y be random variables of continuous type with the joint probability density function is given by f(x, y) that takes a value, $\frac{k x^2}{y^3}$, whenever x takes a value from 0 to 1, and y takes a value greater than 2, otherwise it is going to be 0. So, you can think of as some surface with the three dimension plane, x is one coordinate, y is another coordinate, and z that is the joint probability density function.

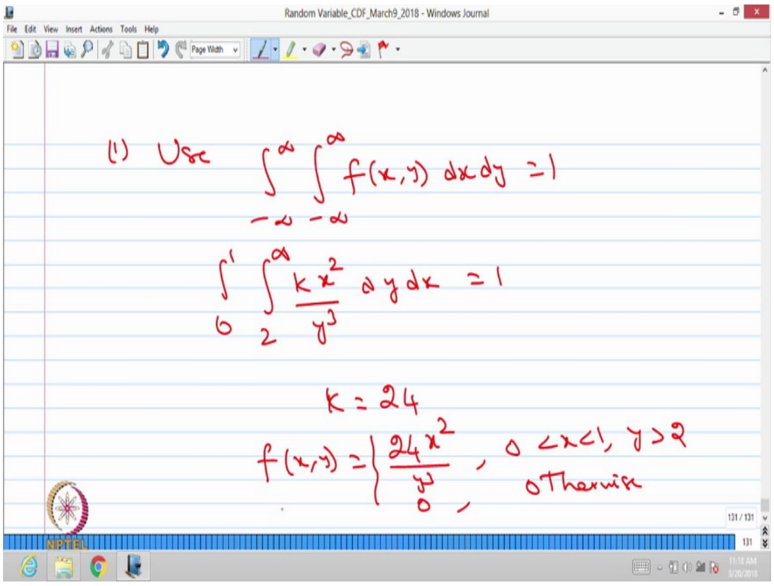
So that means, you have a surface which is greater than 0, the value is going to be greater than or equal to 0 in between x lies between 0 to 1, and y is from 2 to infinity in that the function is going to be greater than 0 that is $\frac{k x^2}{y^3}$; otherwise it is 0, Therefore, the volume below that it is going to be 1.

Now, the question is since I have made $\frac{k x^2}{y^3}$. First you have to find out what is a k in which the given function with two variables is going to be joint probability density function, after that you have to find out some more results so, the first question is find k. So, that this is going to be joint probability density function of two dimensional random variables of continuous type.

Second question find $P\{X < ½, Y > 6\}$, that is the second question is similar to identifying the probability of X lies between X belonging to some Boral set. The third question find out the probability density function of the random variable X. The 4th question is find out the probability density function of the random variable Y.

So, these are all the 4 questions for this problem it is an easy problem. So, let us go for finding out what is value of k in which it is going to be joint probability density function. We know that the property of joint probability density function is always going to be greater or equal to 0, and double integration has to be 1.

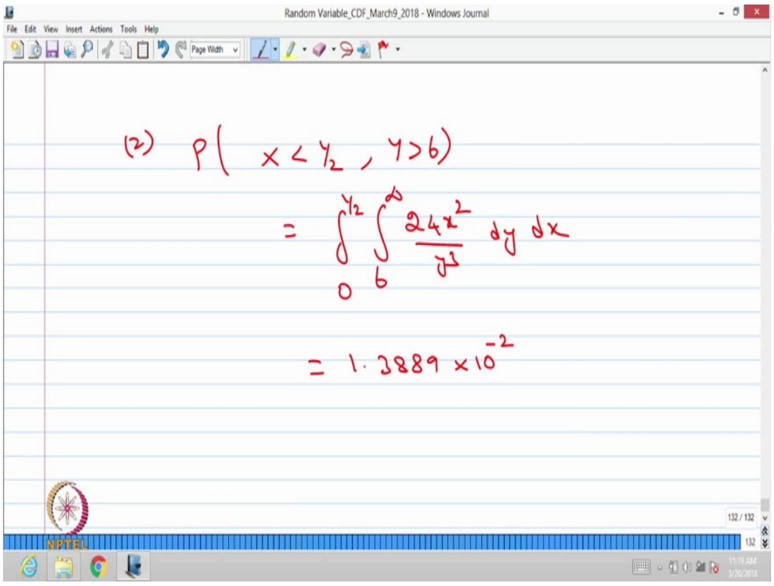
(Refer Slide Time: 17:53)



So, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy$ is going to be 1, but this joint probability density function is greater than 0 between some interval. So, we can make out $\int_{0}^{1}\int_{2}^{\infty} \frac{kx^2}{y^3}\,dy\,dx$ that is equal to 1; $\int_{0}^{1}\int_{2}^{\infty} \frac{kx^2}{y^3}\,dydx$ has to be 1, you do the simple calculation over this integration. You can come to the conclusion the k value has to be 24. Therefore, the joint probability density function is $\frac{24x^2}{y^3}$, when x lies between 0 to 1, y is greater than 2, otherwise is 0.

So, we have answered the first question, second one find out the P{X < ½, Y > 6}.

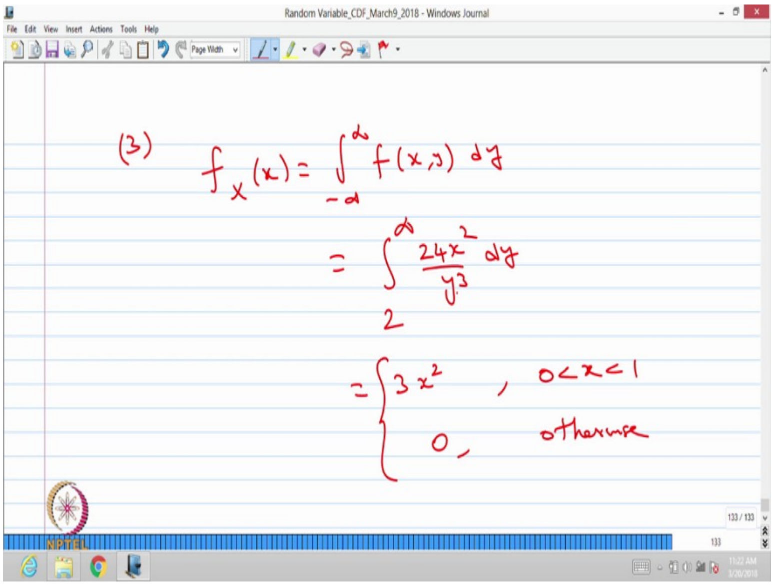
(Refer Slide Time: 19:13)



So, this is nothing but the $\int_0^{1/2}\int_6^{\infty}\frac{24x^2}{y^3}dydx$. Here also if you do integration and do the simplification, you can get the answer it is 1.3889 $\times 10^{-2}$, I am not spending time on the integration.

You can do the integration and you can get the answer, that is am using the concept of the probability of; am using this concept of probability of n random variable, belonging to the Boral set is nothing but integration of n dimensional random variable over the Boral set, the same concept is used to compute the P{X < ½, Y > 6}.The same way am going to use the finding out the marginal distribution from the joint distribution for the next question. So, the next question is find out the marginal distribution of X. Similarly, marginal distribution of Y.

(Refer Slide Time: 21:06)

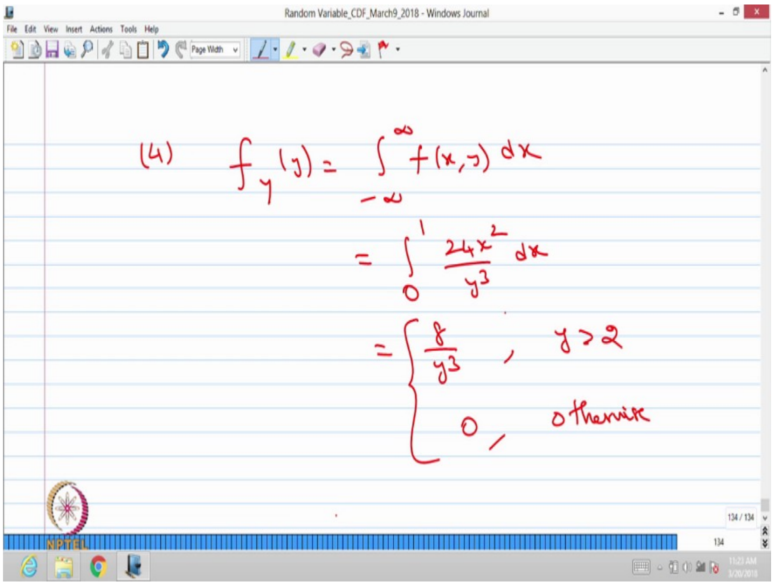


So, third question the probability density function of X is nothing but $\int_{-\infty}^{\infty} f(x, y) dy$.

And we know that the probability density function is greater than 0, when y is lies between 2 to infinity therefore, this is nothing but $\int_{2}^{\infty} \frac{24x^2}{y^3} dy$.And if you do the little simplification, you can get the answer that is $3x^2$. And if you recall the joint probability density function is lies between 0 to 1 for x and for y it is 2 to infinity, therefore, this probability density function is going to be $3x^2$ when x is lies between 0 to 1, otherwise it is 0. This is a very easy problem in which the interval of x does not involve y.

Sometimes you may have a complication of the interval of x is a function of y, or interval of y may be a function of x also. So, you have to use the calculus of several variable and the integration concepts correctly to get the marginal distribution of X.

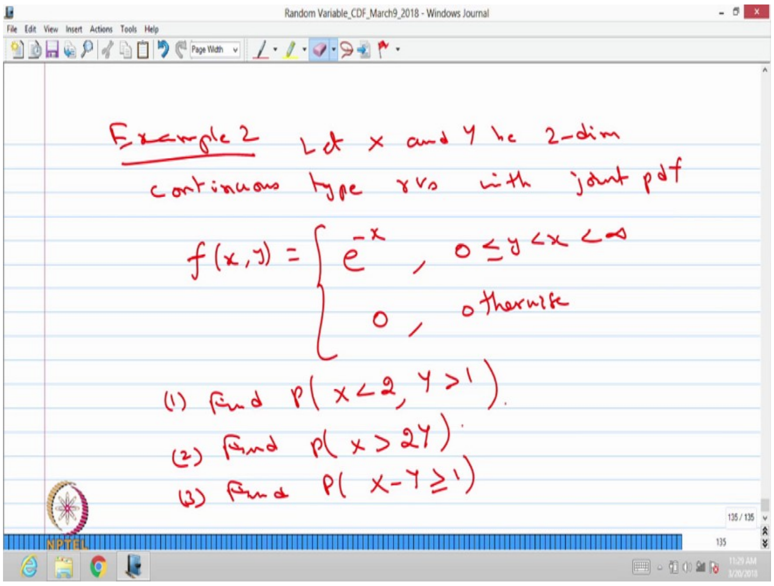
(Refer Slide Time: 22:56)



Similarly, we can go for to find the probability density function of Y that is nothing but $\int_{-\infty}^{\infty} f(x,y)\,dx$.So, here the joint probability density function is greater than 0, when x lies between 0 to 1. So, $\int_0^1 \frac{24x^2}{y^3}\,dx$.. That is same as if you do the little simplification, you will get $\frac{8}{y^3}$ and if you recall the range of y is 2 to infinity. Therefore, y > 2; the probability density function of Y is $\frac{8}{y^3}$, 0 otherwise.

Otherwise means the probability density function is 0 between -∞ to 2, from 2 to ∞ the value is $\frac{8}{y^3}$. We will go for second problem example 2.

(Refer Slide Time: 24:15)

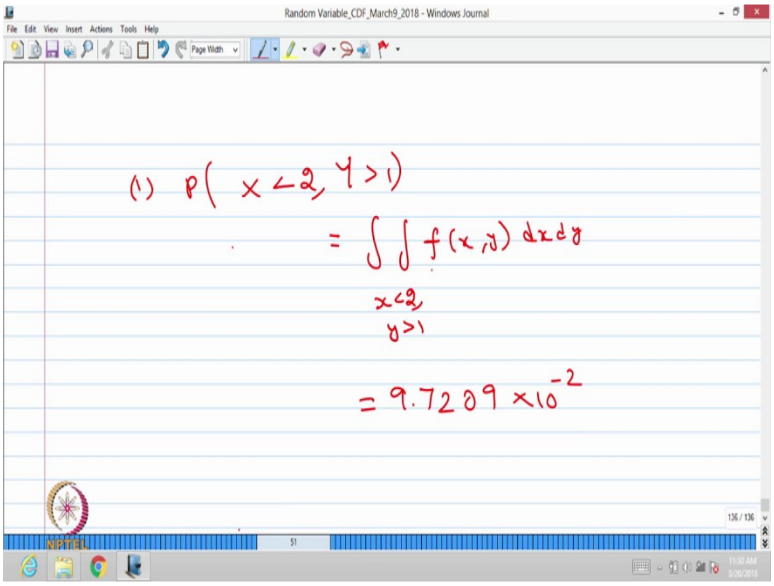


In this problem let me discuss X and Y be two dimensional continuous type random variables with joint probability density function is given by small f. Whenever I use a small f that is a joint probability density function if it is a capital F; that means, it is a CDF cumulated distribution function. So, here the joint probability density function it takes a value e^{-x}, when $0 \le y < x < \infty$.

The range of y and range of x is a function of other variable, 0 otherwise. That means, if you visualize the graphical representation of this joint probability density function, x axis y axis, and z axis joint probability density function. So, between the y is lies between 0 to x, you can make equal also when y is between the 0 to x; that means, y is equal to 0 and y is equal to x. So, in that when x lies between y to infinity you have e^{-x} surface; that means, at the volume below e^{-x}between this region in the xy plane that volume is going to be 1 you know e^{-x} how it goes.

So, as x tends to infinity it asymptotically touches 0, therefore, the surface e^{-x}goes down over the region of $0 \le y < x < \infty$, and the volume below that it is going to be 1. The question is here, find P{X < 2, Y > 1}. And the second question find the P{X > 2Y}, and the third question find the P{X - Y ≥ 1}. It is a very simple problem, the joint probability density function is given finding out the probability of a different Boral sets in R^2.

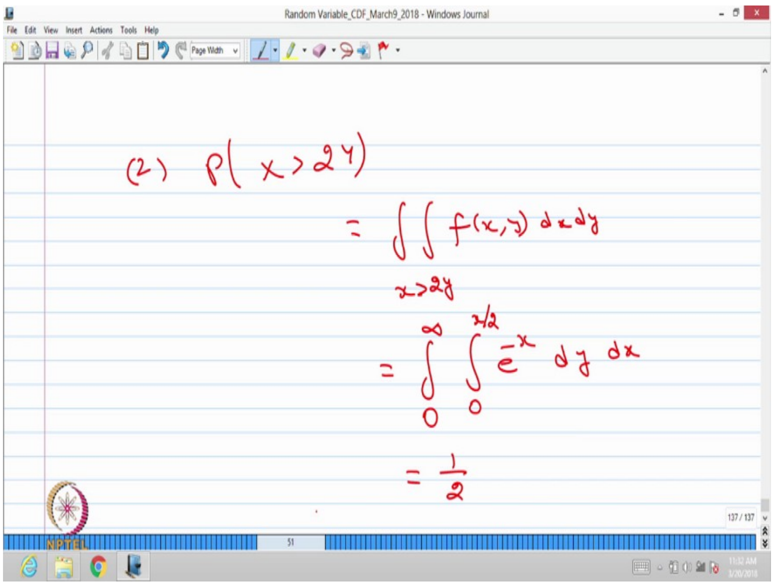
(Refer Slide Time: 27:40)



Will go for finding the first one find the P{X < 2, Y > 1}.

That is nothing but the $\iint\limits_{\substack{x<2 \\ y>1}} f(x,y)\,dxdy$. You can always get this integration and we can get the final answer. So, the final answer is 7.7209 times 10^{-2}. I am not evaluating this integration, the integration is over x < 2 and y > 1.We will go for the second problem.

(Refer Slide Time: 28:47)

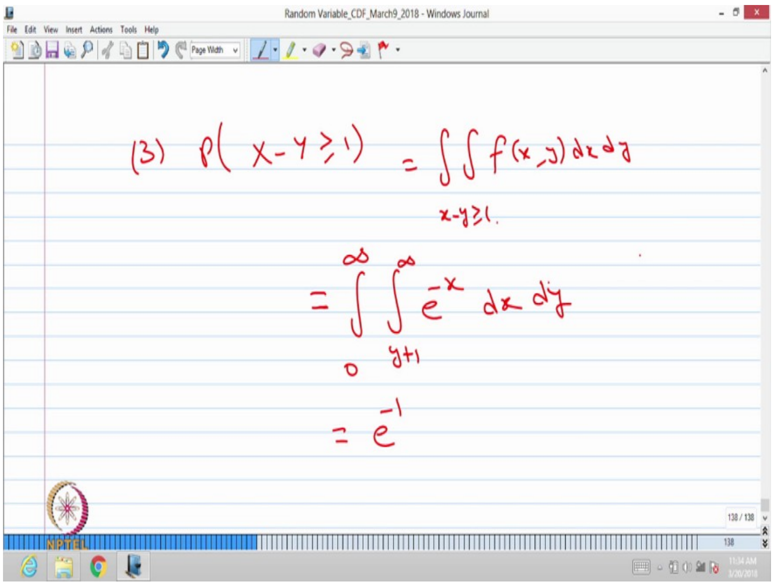


Find the P{X > 2Y}, this is also in the similar way. $\iint\limits_{x>2y} f(x, y)\, dxdy$ That is same as

$\int\limits_0^\infty \int\limits_0^{x/2} e^{-x} dydx$, because x > 2 y.

So, the range of y is 0 to x/2; range of x, that is from 0 to infinity. That means, the joint probability density function integration over x > 2 y, where our joint probability density function is greater than 0, when 0 ≤ y < x < ∞, therefore, you will get the integration with respect to y is 0 to x/2 and integration with respect to x is 0 to ∞. If you do the simplification you can get the answer 1/2. So, this is the P{X > 2Y}.

(Refer Slide Time: 30:31)



The third problem, the P{X -Y ≥ 1}; again the same concept; that is $\iint\limits_{x-y\geq 1} f(x, y)\, dxdy$. That is same as the $\int\limits_0^\infty \int\limits_{y+1}^\infty e^{-x} dx\, dy$.

The x range is from y + 1 to infinity, and y range is from 0 to infinity. If you do the simplification you will get e^{-1}. So, this is way one can find the probability of any Boral set by integrating the joint probability density function over the range.

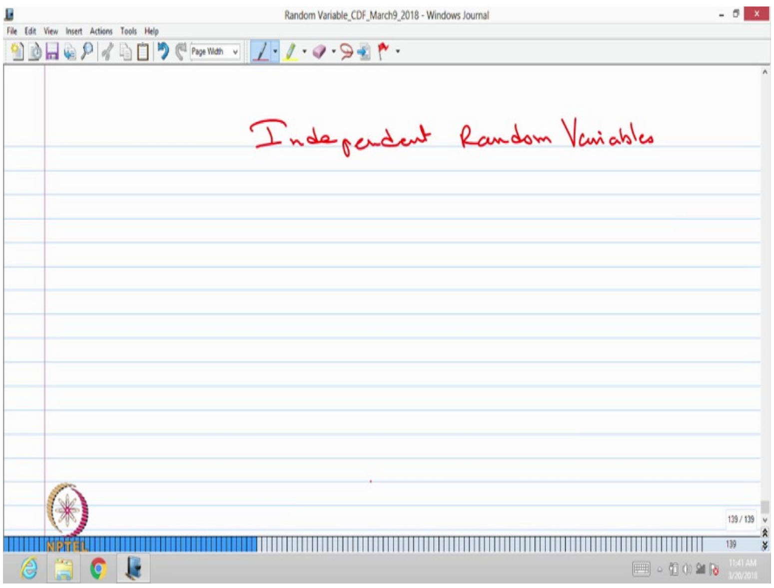
Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 27

In this lecture we are going to discuss the independent random variables. When we discuss many random variables, sometimes some random variable they may have a relations within themselves or sometimes may not. So, it is important to study whether these random variables are having some dependency or not. So, this dependency can be studied by using nice mathematical way through the CDF; that is a joint CDF. If the joint CDF of two dimensional random variable or n dimensional random variable satisfy some conditions then we can conclude those random variables are independent.

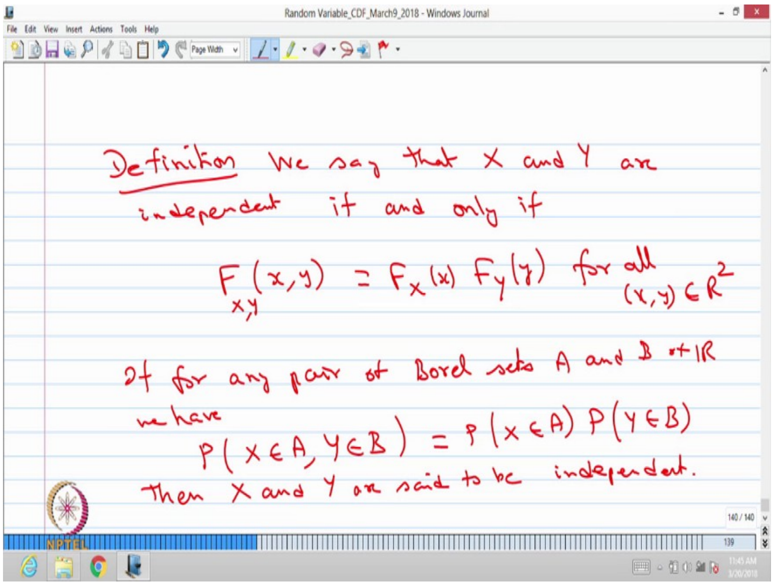
Then the next question comes why do you need to study the independence of random variables. If the random variables are independent, then some of the prediction or some of the sum of finding the probabilities of those random variables will be easy when those random variables are independent. So, let me start with the concept called independent random variables.

(Refer Slide Time: 01:21)



Let me start with the definition of independent random variables, then few more properties when these random variables are independent at the end I will give one or 2 examples for the conceptual understanding.

(Refer Slide Time: 01:54)



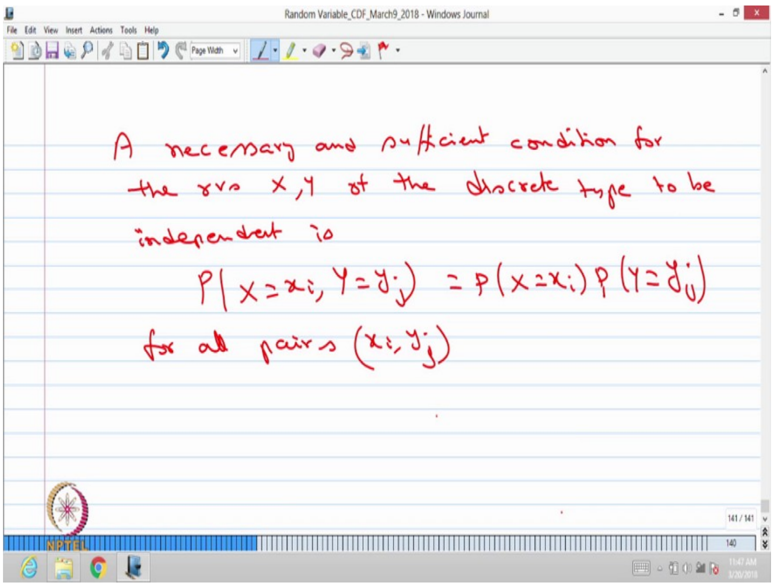
The definition, let me start with the two dimensional random variable, then the same concept can be extended to the n dimensional random variable. So, that way it is easy to explain the concept. We say that the random variable X and the random variable Y are independent if and only if the CDF of two dimensional random variable that is same as the CDF of one dimensional random variable with the product. Whenever I write the suffix; that means, the CDF is corresponding to that random variables. So, $F_{X,Y}$; that means, it is the CDF of two dimensional random variable.

F_X; That means it is the CDF of the random variable X, F_Y; that means, CDF of the random variable Y. If the $F_X(x) F_Y(y) = F_{X,Y}(x,y)$ for all x, y in R^2. Then we can conclude these are independent random variables. This is if and only if condition; that means, if two random variables are independent this condition will be satisfied, if this condition is satisfied then we can conclude both the random variables are independent. It is immaterial of the random variables are of the discrete type or continuous type or mixed type because the CDF is always exist whatever be the type of random variable. Here we are saying the two random variables are independent if and only if the condition on the CDFs.

That means, if I have a for any pair of Boral sets. Suppose I keep the Boral set A and another Boral set B of real line then we have if two random variables are independent, then the $P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$. That means, if two random variables are independent if and only if condition is satisfied.

Whenever two random variables are independent, then we can always get the probability of X belonging to Boral set and Y belonging to another Boral set that is same as X belonging to the one Boral set multiplied by probability of Y belonging to the other Boral set. This is if this condition is satisfied then X and Y are said to be independent whereas, the first condition is if and only if condition we are not saying the random variable is a discrete type or continuous type. Now, I am going to make a condition on whether the random variables are of the discrete type or continuous type and how this if and only if condition changes.

(Refer Slide Time: 06:27)

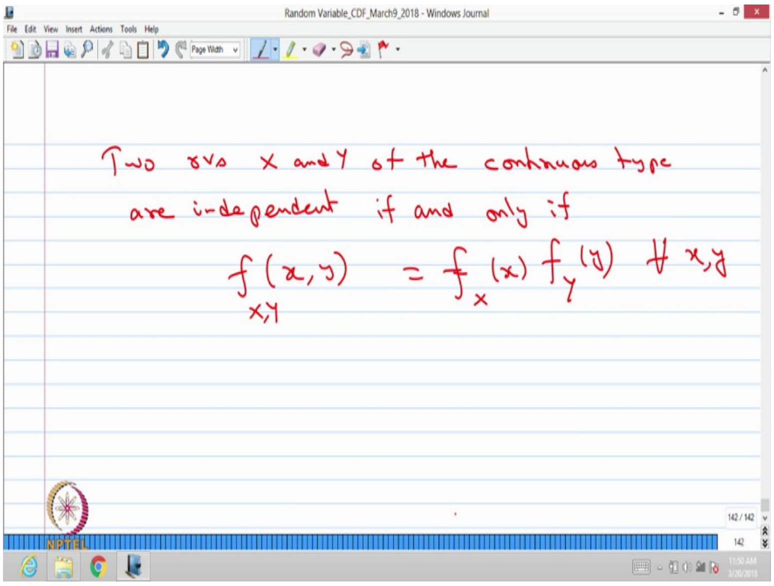


That is a necessary and sufficient condition for that random variables X and Y of the discrete type to be independent if P{X ¿ x_i,Y ¿ y_j} = P{X ¿ x_i}P{Y ¿ y_j} this is for all pairs (x_i, y_j).

That means, this is if and only if condition if both the random variables are of the discrete type, then the independent condition of CDF can be replaced by the independent condition on the joint probability mass function or joint probability mass function is same as product of probability mass functions of X and Y. This is also if and only if condition; that means, if two random variables are of the discrete type are independent then this condition will be satisfied we call this condition as independent condition.

Similarly, if this condition is satisfied for all pairs, then both the random variables are of the discrete type and they are independent.

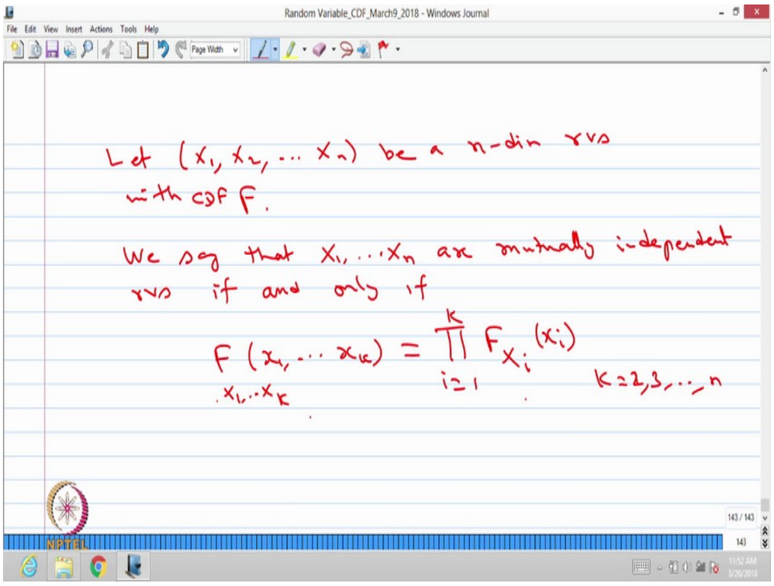
(Refer Slide Time: 08:40)



The similar results is for the continuous type also. two random variables X and Y of the continuous type are independent, if and only if you have to replace the condition of the CDF by the probability density function. So, the joint probability density function when I write suffix X, Y; that means, it is $f_{X,Y}(x,y)=f_X(x)f_Y(y)$ this is for all x, y. So, if this condition is satisfied then two random variables are of the continuous type or independent.

If two random variables are of the continuous type or independent, then this condition will be satisfied for all x, y. Therefore, the condition for independent random variables either in the level of a CDF or if it is a discrete random variable in the form of probability mass function if the random variables are of the continuous type, then it is a probability density function. So, all are all 3 are going to be if and only if condition not the only one side it is in the both side. So, even though we have explained through the two dimensional random variables this can be extended to the multi dimensional random variable also.

(Refer Slide Time: 10:49)



That means, let $(X_1, X_2, \ldots, X_n)$ be a n dimensional random variables with the CDF F is a function of n variables. We say that the random variable $X_1, X_2, \ldots, X_n$ are mutually independent random variables if and only if you take any fewer random variables CDF k, for k is equal to 2, 3 and so, on till n, that is same as product of i is equal to 1 to k the CDF of those random variables CDF. This means if you take any two random variables the CDF of those two random variables is same as a product of a CDF of only those two random variables.

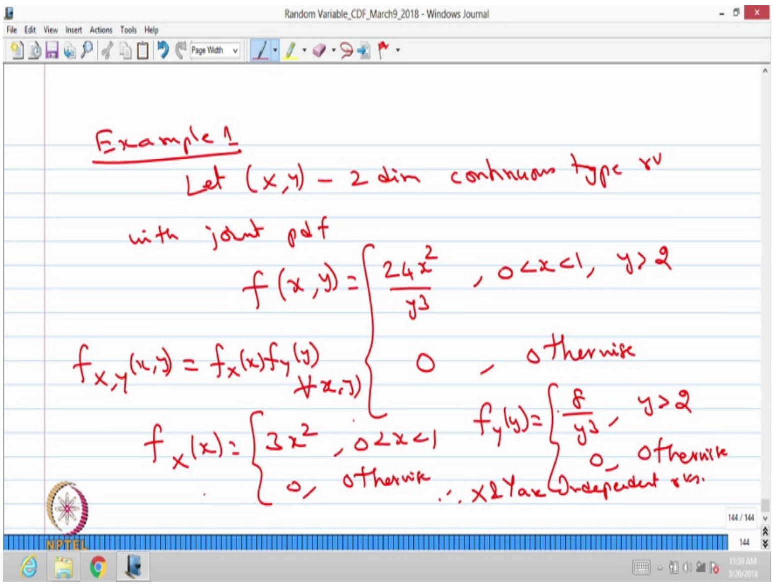
If I take any 5 random variables, when n is greater than 5 then CDF of 5 random variables is same as product of those 5 random variables CDF, then we conclude they are mutually independent. It is same as the mutually independent events. If you have an n events and once you say that they are pairwise dependent. That means, any two events satisfies the independent concept or independent condition then they are called pairwise. If it means mutually; if they are mutually independent that means, whatever be the collection of events you take the independent condition is satisfied then we conclude their mutually independent events.

The same thing here if you have n random variables whatever be the number of random variables you take it from those n random variables, that satisfies independent condition then it start from any two till all the random variable then we conclude they are mutually independent. Whenever we say more than two random variables are independent; that means, by default they are mutually independent whenever we say more than two random variables are independent; that means, by default they are mutually independent.

Many time we would not write again and again mutually independent word when we discuss more than two random variables, but whenever we have more than two random variables when we use the word independent random variable. That means, they are mutually dependent; that means, the independent condition is satisfied for all the forms of collection of random variables, satisfying the independent condition.

Now, we will move into some problem of explaining how the independent random variable is playing a role.

(Refer Slide Time: 14:41)



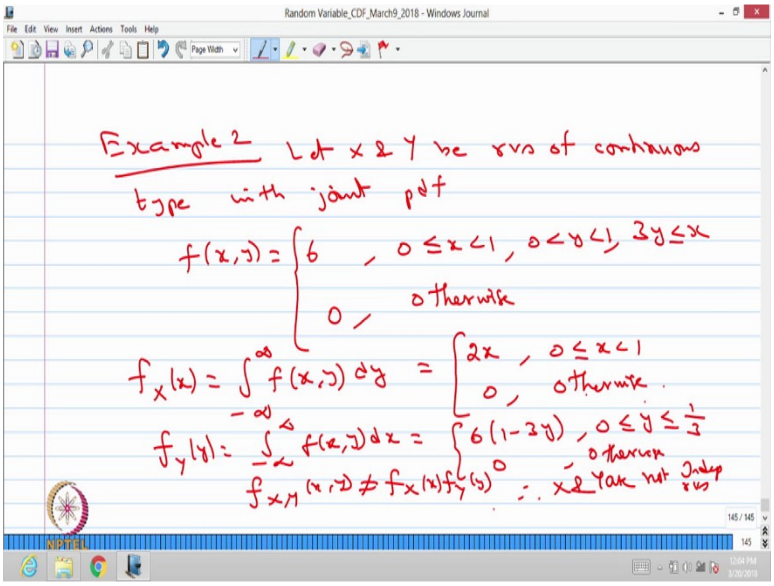
The first example that is same as the example which we have considered earlier, that is let X, Y be a continuous type random variable two dimensional continuous type random variable with joint probability density function is of the form f(x, y), that is after we find the value of k we got a $\frac{24x^2}{y^3}$, when x is lies between 0 to 1 and y is greater than 2 otherwise it is 0. So, this problem just now we have discussed when two random variables are of the continuous type for that we found the k, that k value was 24 and we found the probability of x between some interval.

Here, we will verify whether these two random variables are independent or not. So, this is the joint probability density function already in the same example we got the probability density function of X, that is $3x^2$ when x is lies between 0 to 1; otherwise 0. Similarly, we got the probability density function of Y that is $\frac{8}{y^3}$ when y is greater than 2; 0 otherwise.

Easily it can be verified in this example the joint probability density function is same as product of probability function of random variables X, Y. Because it is $3x^2$ the other one is $\frac{8}{y^3}$, not only that the $3x^2$ is range between 0 to 1 and $\frac{8}{y^3}$ the range is y is greater than 2, which is same as if you make a product that is $\frac{24x^2}{y^3}$ and the range of x is 0 to 1 and the range of y is 2 to infinity that is same as the joint probability density function of $\frac{24x^2}{y^3}$ when x lies between 0 to 1 and y is greater than 2.

So, their interval matches and the value matches; 0 otherwise matches therefore, for all x, y the joint probability density function is immersed the product of density functions of X and Y therefore, these two random variables independent. Therefore, X and Y are independent random variables. Each one is of the continuous type we can check it from the CDF also that is if and only if condition for the CDF. But, since the joint probability density function is given you can find the probability density function of X and Y, then you can verify the independent conditions on probability density function that is satisfied. Therefore, both the random variables are independent.

(Refer Slide Time: 18:40)



We will go for one more example, example 2 again we have a continuous type let X and Y be a random variables of continuous type with joint probability density function that is f(x, y), which takes a value 6 when x is lies between 0 to 1 and y is lies between 0 to 1 as well as 3 y ≤x. So, the joint probability density function is greater than 0, that is 6 when x is lies between 0 to 1, y is lies between 0 to 1 and 3y ≤ x; 0 otherwise.

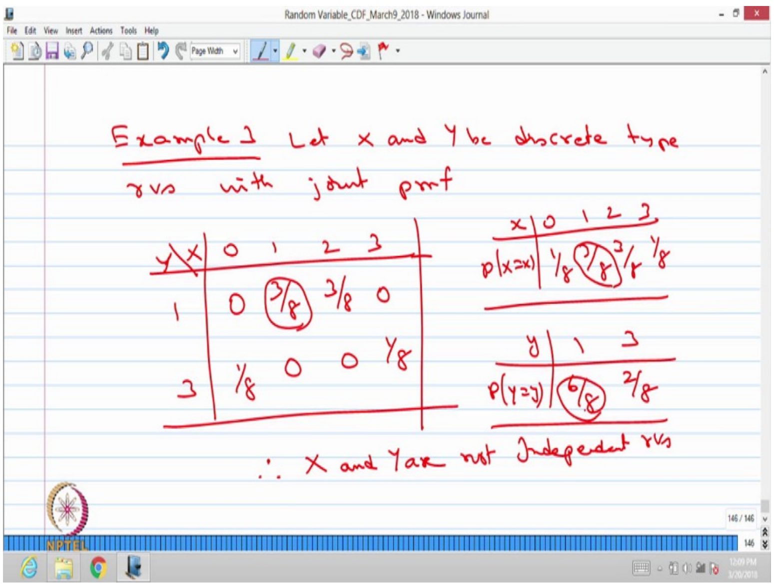
Before we proceed the problem, we can always verify whether this is a correct joint probability density function; that means, if you integrate double integration over x and y this has to be 1. So, one can verify this is going to be double integration is one. Therefore, this is the correct joint probability density function. From these; one can find the marginal distribution of X by $\int_{-\infty}^{\infty} f(x,y)dy$. If you do the little simplification you can get the answer that is 2x when x is lies between 0 to 1; otherwise it is 0.

You can verify this result also whether this is a correct probability density function of X by integrating 0 to 1 for x you will get the value 1 and it is greater than or equal to 0 therefore, this is the probability density function of X. Similarly you can compute the probability density function of Y by $\int_{-\infty}^{\infty} f(x,y)dx$. Here also am skipping the integration part one can get the answer that is a 6(1 – 3y), when y is lies between 0 to 1/3; 0 otherwise. That means, within this interval 0 to 1/3 the probability density function is greater than 0, 6(1 – 3y); otherwise 0.

By seeing the probability density function of X probability density function of Y this product is not going to be the joint; that means, the f(x, y) is not equal to the product of probability density function of X and Y. For all x, y if this condition is satisfied equal to then you can conclude they are independent.

But since by seeing this you can say it is a 2x times 6(1 – 3y) whereas, the joint probability density function is 6 obviously you can say they are not equal. Therefore, X and Y are not independent random variables. So, I have given the first example in which they are independent random variable by finding the marginal distribution whereas, in this example by finding the marginal distribution of X and Y, we are concluding condition is not satisfied, independent condition is not satisfied therefore they are not independent random variables.

(Refer Slide Time: 23:27)



We will go to the one more example that is example 3, because already we discussed two problems of the continuous type we will see one problem of the discrete type also. That is let X and Y be a discrete type.

Let X and Y be discrete type random variables with joint probability mass function is given by. If you recall this is same as the problem which we have discussed in the last class, the possible values of X is 0, 1, 2, 3 and the possible values of Y is 1 and 3 where X denotes number of heads obtained when we tossing a unbiased coin 3 times. And Y is the difference in absolute of a number of heads and number of tails obtained. Therefore, the possible values of Y is 1 and 3 and the possible values of X is 0, 1, 2 and 3, and in that problem we have got the joint probability mass function that is 0, 3/8, 3/8, 0 then 1/8, 0, 0 and 1/8 and in that problem we have got the marginal distribution of X and Y also if you recall.

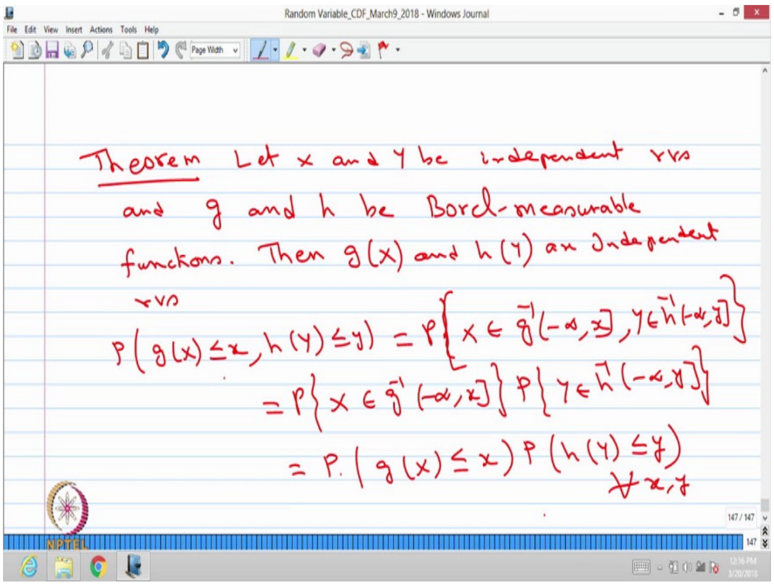
So, for possible values of x that is 0, 1, 2 and 3 and the $P\{X = x\}$ that is going to be 1/8, 3/8, 3/8, 1/8. So, this is a probability mass function of X and similarly probability mass function of Y that is 1 and 3. So, 1 it takes a value 6/8, and 3 takes a value 2/8. So, this is the probability mass function of Y. Now you can verify whether these two random variables are independent suppose x takes a value 1, y takes a value one that probabilities 3/8.

Here x takes a value 1 and y takes a value 1 that is this much. So, if you make product of $3/8*6/8 \neq 3/8$. Even if one pair it is not satisfied then you cannot conclude it is independent random variable. If all the pairs the joint probability mass function is same as product of probability mass functions of X and Y, then only you can conclude their independent. Since any one pair does not satisfy then you can immediately conclude both the random variables are not independent.

This will be a set of obvious because the random variable Y is defined as difference in absolute with the number of heads and number of tails whereas, the random variable X is defined number of heads; that means, the Y itself is a function of X; that means, Y is dependent on X. Therefore, there is a dependency between the random variable Y and X in the definition itself. Therefore, from there itself you can conclude they are not independent random variable, but that we have concluded from the distributions also.

Lecture – 28

(Refer Slide Time: 00:05)



When two random variables are independent, I am going to give it as one important result as a theorem. Let X and Y be independent random variables, and g and h be Boral measurable functions, then we will conclude after we get the results. So, let me keep the then word as it is, let us go for finding out what is the probability of $P\{g(X) \le x, h(Y) \le y\}$.

Let us go for finding out function of a random variable with X in the form of g(X), function of a random variable with the random variable Y, In the form of h(Y). We will try to find out the joint CDF of are CDF of the random variable g(X) and h(Y). Since X is a random variable g is the Boral measurable function therefore, g(X) is a random variable so, you can think of g(X) is some other random variable. Similarly, Y is a random variable h is a Boral measurable function therefore, h(Y) is also a random variable.

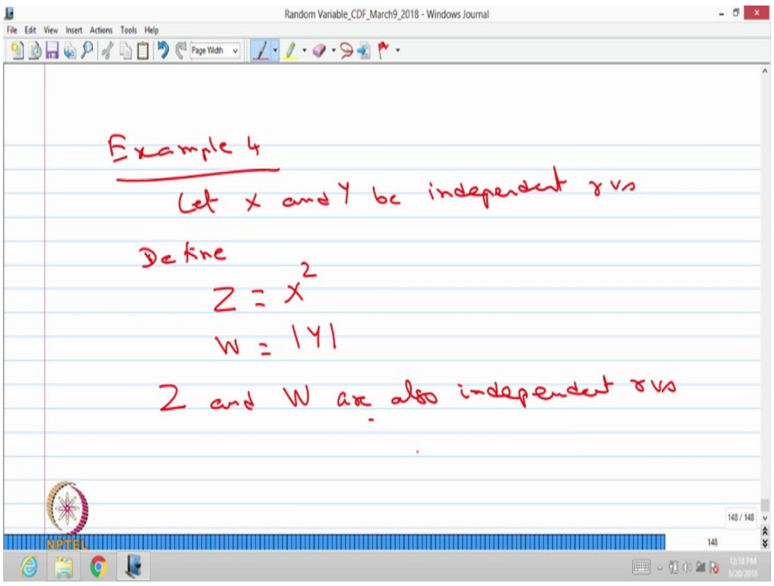
So, you can think of some other random variable; that means, we can treat g(X) is another random variable, third random variable other than X and Y, h(Y) is a 4th random variable. We are trying to find out what is the joint CDF of these two random variables, whether that satisfies the independent condition.

If they are going to be satisfies the independent condition, then you can conclude g(X) and h(Y) are also independent random variable. That is P{X $\in g^{-1}¿\}$ and since h(Y) $\leq$ y this also can be written $P\{Y \in h^{-1}¿\}$. That is same as the probability of; the way we write X belonging to g inverse, Y belonging to h inverse.

We know that the random variable X and Y are independent, since X and Y are independent. X belonging to some Boral set, Y belonging to some Boral set. The probability of that is same as probability of X belonging to the Boral set, that is P{X $\in g^{-1}¿\}$. Similarly, the product of P{$Y \in h^{-1}¿\}$ because X and Y are independent, that is same as the P{g(X) $\leq$ x}P{h(Y) $\leq$ y}. And this is valid for all x and y.

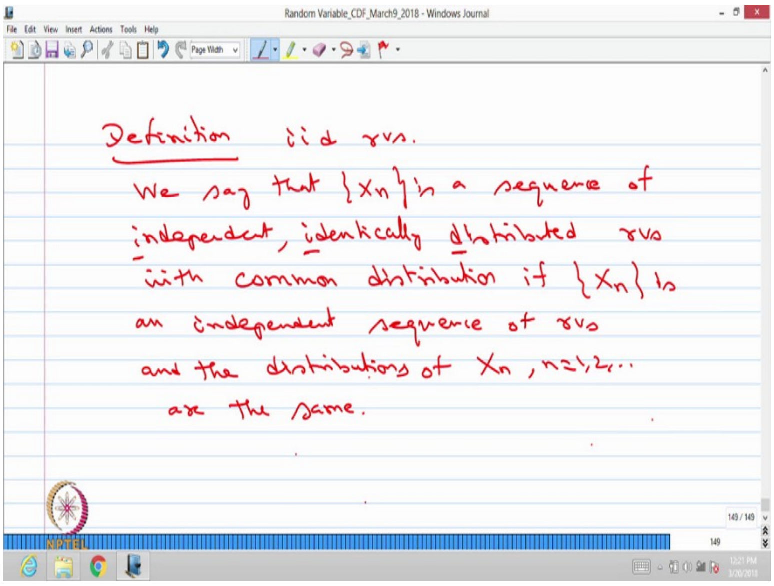
This is valid for all x, y ; that means, the probability of joint CDF is same as P{g(X) $\leq$ x}P{h(Y) $\leq$ y}. Since this condition is satisfied, we can conclude the random variables g(X) and h(Y) are independent random variables. This is a very important result whenever you have independent random variable, if you create a Boral measurable functions on those independent random variables, that is also going to be an independent random variables. As an example we can think of already we discussed three examples.

(Refer Slide Time: 05:34)



So, the 4th example is a let X and Y be independent random variables. Define Z = X^2, W = |Y|; that means, I am creating a g(X) as X^2, and h(Y) = |Y| and since these two are Boral measurable functions g(X) and h(Y) form, we can conclude Z and W are also independent random variables. Since X and Y are in independent random variable, Z is a Boral measurable function of X, W is a Boral measurable function of Y, therefore, Z and W are also independent random variables.

(Refer Slide Time: 07:04)

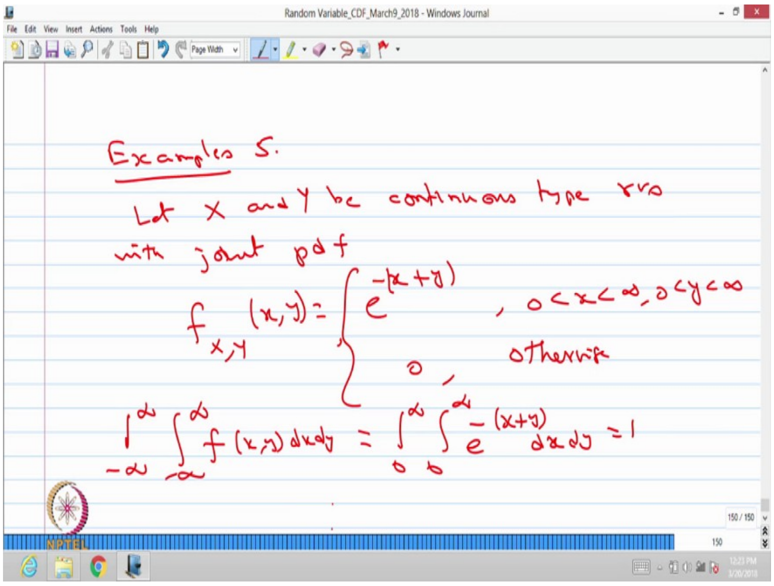


The next concept which we are going to discuss as the form of definition. That is iid random variables; we say that X_n is a sequence of independent identically distributed random variables with common distribution, if the sequence X_n is independent sequence of random variables and the distribution of each random variable X_n are the same.

Whenever we say the sequence of random variables are iid; that means, the first i comes from here the second i comes here the d comes from, whenever we say the collection of random variable or sequence of random variables are iid; that means, they are mutually independent as well as the distributions are same.

Whenever the distributions of few random variables are same, as well as all those random variables are mutual independent then we can conclude that collection of random variables are the iid random variables.

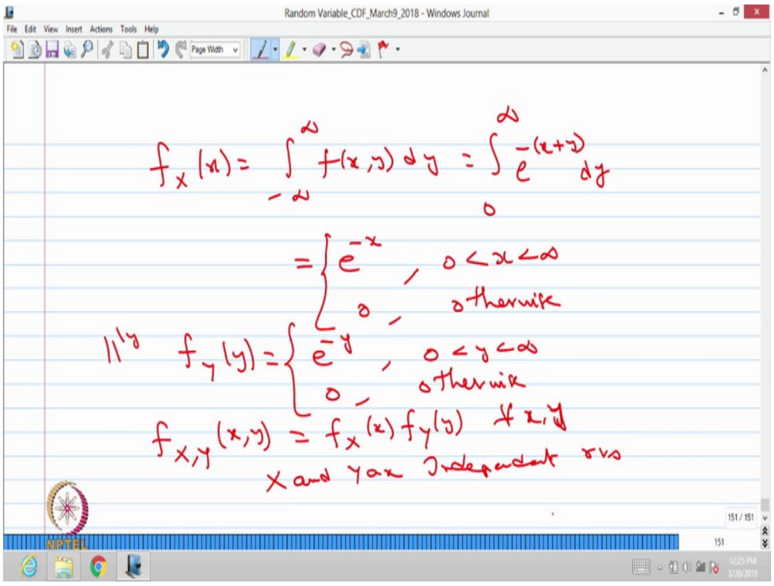
(Refer Slide Time: 09:44)



We can go for creating a simple example of iid random variables. So, this is example number 5, let X and Y be continuous type random variables with joint probability density function is given by $e^{-(x+y)}$.

When x is lies between 0 to infinity and y is also lies between 0 to infinity; 0 otherwise, we started with the two dimensional continuous type random variable. Random variables with the joint probability density function, that is $e^{-(x+y)}$, when x lies between 0 to infinity y between 0 to infinity, you can verify whether this is going to be the joint probability density function by integrating joint probability density function with respect to x and y that is same as $\int_0^{\infty}\int_0^{\infty} e^{-(x+y)} dxdy = 1$.

(Refer Slide Time: 11:34)

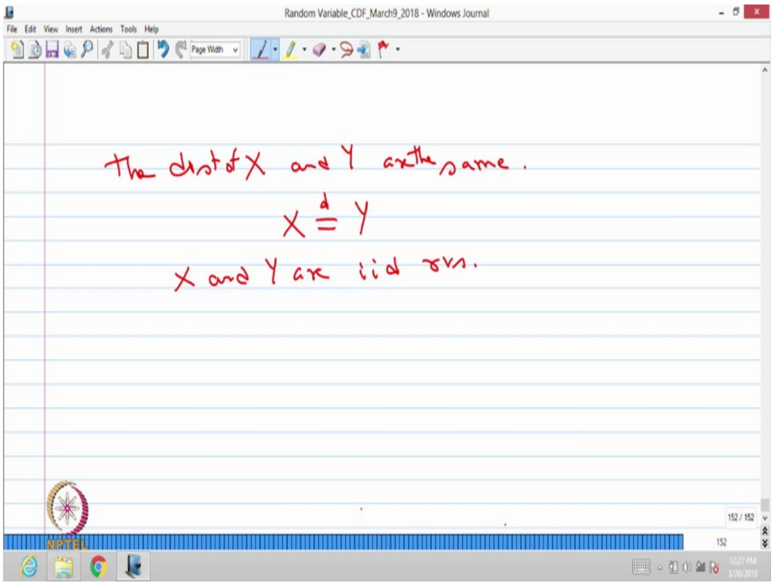


Therefore this is joint probability density function, let us go for finding out the marginal distribution by integrating the joint probability density function with respect to y, that is $\int_0^{\infty} e^{-(x+y)} d\,y$, if you do this integration you will get the answer that is e^{-x}; when x is lies between 0 to infinity. So, this is the probability density function of X. Similarly, if you do the exercise of finding the probability density function of Y, you will get e^{-y}, when y lies between 0 to infinity, otherwise it is 0.

If you cross check the joint probability density function, whether that is same as product of probability density function of X and Y or all x and y, the joint probability density function is $e^{-(x+y)}$, and the probability density function of X is e^{-x}probability density function of Y is e^{-y} so, this condition is true for all x, y . Therefore, we can conclude X and Y are independent random variables. Not only that probability density function of X is e^{-x}, probability density function of Y is e^{-y}.

So, if you find the distribution; that means, the CDF, or the probability density function for a continuous type, or probability mass function for discrete type, the distributions are same the distributions of X and Y are same, also they are independent random variable.

(Refer Slide Time: 13:53)



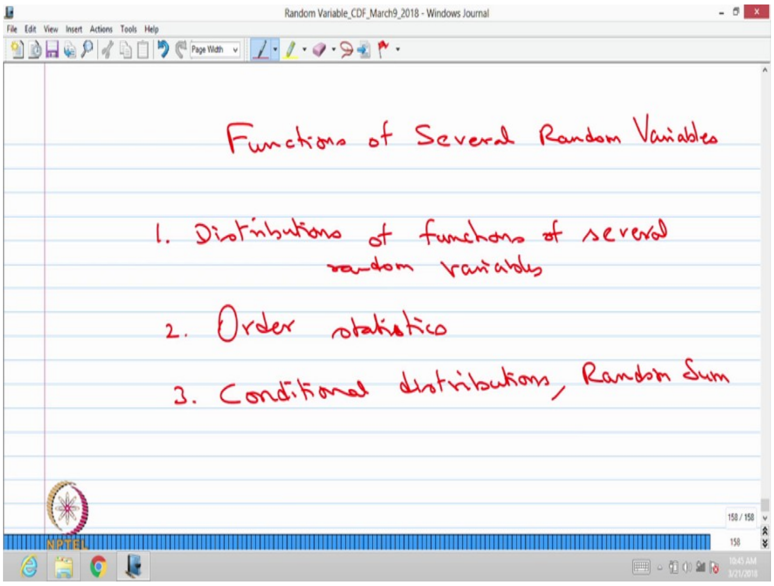
Therefore, we can conclude the distribution of X and Y are same, are the same we can conclude in notation, we can use X is identical to Y in distribution, they are identical distribution the distribution of X and the distribution of Y are same we can use the word d above the equal symbol; that means, both are having the identical distributions. And also they are independent therefore; we can conclude X and Y are iid random variables. So, whenever we write iid random variables that means, those random variables are mutually independent, also they are having the same distributions, then we call it as iid random variables.

Introduction to Probability Theory and Stochastic Processes

Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Module - 06
Functions of Several Random Variables
Lecture – 29

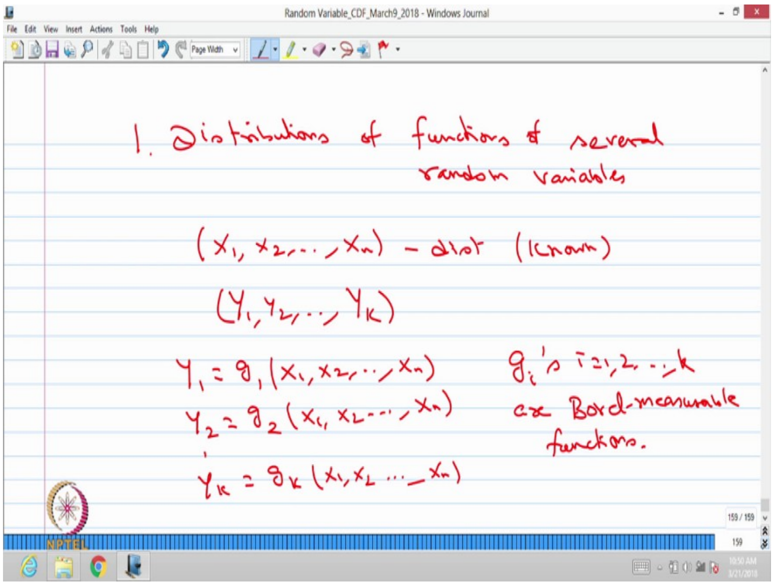
(Refer Slide Time: 00:24)



In this week we are going to discuss the model on functions of several random variables. In the last model we discussed more than one random variable together that form a random vector, and we discuss the joint distribution of n dimensional random variables. Now, in this module we are going to discuss functions of random variables. In this topic we are going to discuss three aspects; one is the distributions of a functions of several random variables. Then the second lecture we are going to discuss order statistics, this is also one special type of functions of several random variables. Third, we are going to discuss conditional distribution.

The way we discussed conditional probability of event, we are going to discuss conditional distributions, when we discuss several random variables, and followed we are going to discuss random sum. So, these are all the three topics which we are going to cover it in this model, in functions of several random variables.

(Refer Slide Time: 02:35)



The first lecture is on distributions of functions of several random variables. In any random experiment we create n random variables, but later when we start solving the problem. So, we may feel we need some more random variables which is other then what we have started in the beginning. We can always create a new random variable from the scratch; by the definition of random variable then we can get the n dimensional random variable.

But it is easy to create new set of random variable with the help of the earlier set of random variable which we have already created; that means, we can create some sort of composite function of random variables, such a way that we will land up the new set of random variables.

In that case our interest is to find out what is the distribution of the new set of random variables, when we know the distribution of the earlier set of random variables; that means, through the distribution of earlier set of random variables we are going to find the distribution of new set of random variable. When I use the word earlier set of random variables and new set of random variable, I am using the notation called the earlier set of random variables is $X_1, X_2, \ldots, X_n$; that means, each one is a random variable we have n random variables are together jointly, therefore, this is n dimension random variable.

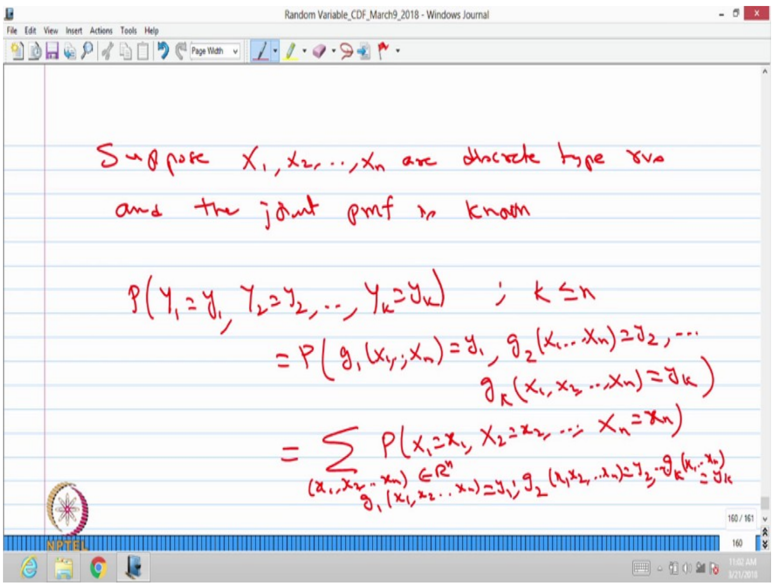
The new set of random variable, I say $Y_1, Y_2, \ldots, Y_k$; that means, the new set of random variable need not be the same size of the earlier set of random variable. So, we know the distribution of this random variable, that is known. We are going to find out what is the distribution of $Y_1, Y_2, \ldots, Y_k$ when we know the distribution of $X_1, X_2, \ldots, X_n$. For that what we are going to do? We are going to create the relation of Y_i's in terms of X_i's; that means, Y_i you can think of it is a composite function with the n random variables.

Similarly, you can think of Y_2 as another random variable, which is a function of $X_1, X_2, \ldots, X_n$ random variable, like that we have k random variables; that means, obviously, X i's are the random variable, Y_i is, i is equal to 1 to k are going to be random variable, provided the function g_i's as to be Boral measurable function.

So, we make the assumptions g_i's where i is equal to 1 to k, are Boral measurable functions, the same thing what we have done in it in the 1 dimensional random variable. When X is the random variable g is a Boral measurable function, then Y = g(X). That is also going to be a random variable the same concept we are using for the several random variables. Therefore, when X_i's are known with the distribution and g_i's are Boral measurable functions then Y_i's are going to be the random variables, therefore, $Y_1, Y_2, \ldots, Y_k$ are the k dimensional random variables.

The question is how to find the distribution of Y_i's, that I am going to give it as the result.

(Refer Slide Time: 07:16)



Suppose X_i's $X_1, X_2, \ldots, X_n$, are discrete type random variables, and the joint probability mass function is known, I am discussing the concept of distributions of several random variable for discrete type random variable first, then later I am going to discuss further continuous type random variable.

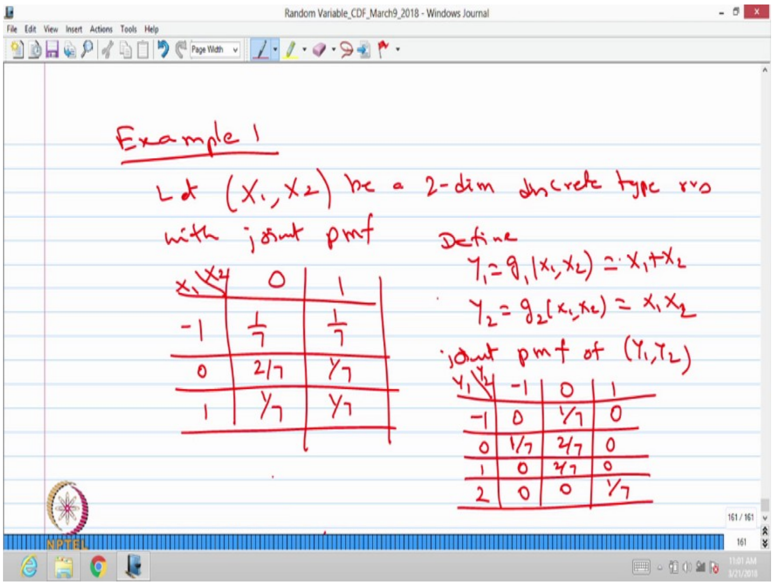
Since we know the joint probability mass function of X_i's you can directly write down the probability mass function of Y_1, Y_2, and Y_k's jointly, that is same as probability of; you can replace $Y_1 = g_1(X_1, \ldots, X_n)$, that takes a value y_1. Y_2 takes a value y_2; that means, $g_2(x_1, \ldots, x_n)$, that takes a value y_2 and so on.

And for the $g_k(X_1, \ldots, X_n)$, that takes a value y_k, that is same as $\sum_{(x_1, \ldots, x_n) \in R^n} P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$, not only that $g_1(x_1, \ldots, x_n)$, that is same as y_1.

Similarly, $g_2(x_1, \ldots, x_n)$, that is same as y_2 and so on $g_k(x_1, \ldots, x_n)$, that is same as y_k. If you make a summation over this conditions; that means, all the x_i's belonging to R^n, and $g_1(x_1, \ldots, x_n)$, that is same as y_1, $g_2(x_1, \ldots, x_n)$, that is same as y_2; like that $g_k(x_1, \ldots, x_n)$, that is same as y_k, finding of the joint probability $P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$ that is going to be the $P\{Y_1 = y_1, Y_2 = y_2, \ldots, Y_k = y_k\}$. So, this is a Probability mass function for k dimensional random variable. So, this is the joint probability mass function of k dimensional random variables $Y_1, Y_2, \ldots, Y_k$.

So, this is the way one can find the distribution of a functions of several random variables when the random variables are of the discrete type. And g_i's are Boral measurable functions so that Y is also going to be discrete random variable one can get the joint distribution in this way. Let us go for one simple example.

(Refer Slide Time: 11:58)



How it works the example 1? Let (X_1, X_2) be two dimensional discrete type random variables with joint probability mass function which is given by X_1, X_2, the possible values of X_1 is -1, 0 and 1.

And the possible values of X_2 that is 0 and 1, the joint probability mass values for possible values of $X_1 \wedge X_2$ this is 1/7, and this is also 1/7, this is 2/7, and this is 1/7, 1/7 again 1/7. If you sum all the values it is going to be 1 2 3 4 5, 5 plus 2 is 7 summation is one, and if you make a row sum and column sum, you will get the marginal distribution of $X_1 \wedge X_2$.

Now, we are going to define the new random variables, that is Y_1 as a function $g_1(X_1, X_2)$, that is $X_1 + X_2$, that is Y_1, we are defining a second random variable as a $g_2(X_1, X_2)$, that is $X_1 X_2$; $X_1, X_2$are of the discrete type random variable and $Y_1 = X_1 + X_2$; $Y_2 = X_1 X_2$. Therefore, you will get $Y_1 \wedge Y_2$ are discrete type random variables.

Now, our interest is to find out the joint probability mass function of Y_1, Y_2. Since it is a discrete type you can make a table so Y_1, Y_2. Since X_1 takes a value -1, 0, 1; X_2 takes a value 0 and 1. Y_1 takes a value -1, 0, 1 and 2. So, minus 1, 0, 1 and 2. Similarly one can find the possible values of Y_2, that is -1, 0 and 1. Now you can start filling up, suppose Y_1 takes a value minus 1, Y_2 takes a value -1, what are all the possibilities in the X_1, X_2? So, that the Y_1 is going to be -1, Y_2 is going to be -1.

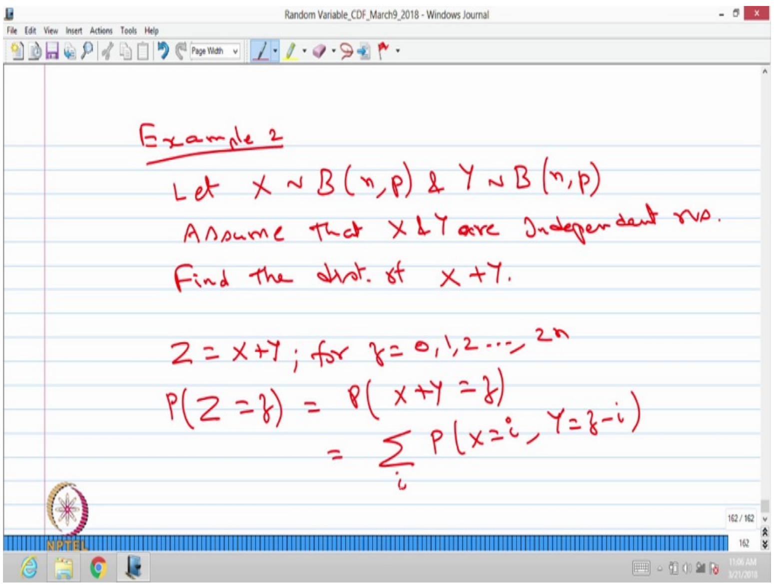
There is no way you will get this possibility, therefore, the probability of empty set is 0, whereas when Y_1 = -1, Y_2 = 0, that is possible with the probability 1/7. Similarly, $Y_1 = -¿$ 1, $Y_2 = ¿1$, that is not possible therefore, the probability is 0.

Similarly, you can fill up other elements 1/7, this is 2/7 you can verify. 0, this is 0, 2/7 again there is no possibility therefore, the probability is 0. There is no possibility this is 1/7 again you can cross check whether this is going to be the whole summation is double summation over $Y_1, Y_2 h$as to be 1. 1/7 +1/7 +2/7 +2/7 +1/7, therefore, the addition is one.

Here also you can find the marginal distribution of $Y_1 \wedge Y_2$from the joint distribution. So, since this is the discrete type with the two dimensional random variable; and again we make a Y_1, Y_2as two dimensional random variable, we are getting the joint probability mass function of Y_1, Y_2. Suppose if it is n dimensional and Y_i's are k dimensional, still you can able to find as long as k ≤n, as long as k ≤ n, you can get the joint probability mass function.

So, that can be written in the previous slide; k ≤ n you can get the joint probability mass function of Y_i's where i is running from 1 to k.

(Refer Slide Time: 17:58)



This is a very easiest example. Now we will go for the little different example 2, let X be a random variable which is binomial distributed with a parameters n, p. And another random variable Y that is also binomial distributed with parameter n, p, same p. And I make the assumption, assume that the random variables X and Y are independent.

I assume that X and Y are independent random variables. The question is, find the distribution of X + Y; find the distribution of X + Y. It is easy to do because X is a discrete type random variable, Y is a discrete type random variable, and we have not supplied the joint probability mass function of X and Y. Whereas, we make the assumption both are independent random variable.

So, you can use the independent concept if two random are independent then the joint probability mass function is same as product of probability mass function of X and Y. So, you can use that concept. Therefore, I can directly compute the distribution of X+ Y. I will take Z = X+ Y. Since X is a binomial distribution, the possible values are 0, 1, 2 and n, and Y is also binomial with the possible values are 0, 1, 2 and so on till n, therefore, the possible values of Z is going to be 0, 1 and so on till 2 n.

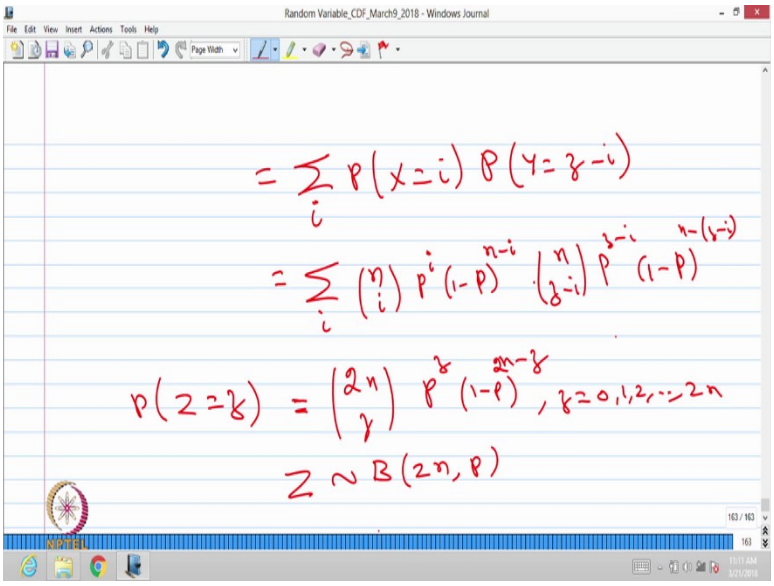
So, one can find the probability mass function of the random variable Z for z takes a value 0, 1, 2 and so on till 2 n. So, this is the probability mass function which we are going to find, which is going to be positive and all others points it is going to be it is 0. So, we will be read about a when z takes a value 0, 1, 2 so on till 2 n. So, this is same as, I can replace Z = X+ Y takes a value small z. That is same as, if I introduce one dummy index i, that is P{X = i, Y = z-i }, and all possible values of i, I can get the P{X + Y = z}. I am just replacing P{ X +Y = z}, that is same as $\sum_i P\{X=i, Y=z-i\}$, that is going to be same as P{ X +Y = z}.

I am not going to write what are all the possible values of i very clearly, because for some i the Y may not be within the range of 0 to n or X cannot be in the range of 0 to n, therefore, whenever the X takes a values 0 to n as well as Y takes a value 0 to n, then only you will have some sort of joint probability mass function, otherwise those probabilities are going to be 0, therefore, I am not going to write what is the possibility values of i. I leave it as it is.

Now I am going to use the independent concept; that means, the $P\{X=i, Y=z-i\}$, that is same as since these two random variables are independent. I can make it $P\{X=i\} P\{Y=z-i\}$, that is what I made the assumption these two random variables are independent otherwise I can proceed further.

So, either to solve this problem, I should have given both the random variables joint probability mass function for X, Y or I would have made the assumption X and Y are independent, then only I can able to find the distribution of X + Y.

(Refer Slide Time: 23:24)



So, this is same as $\sum_i P\{X=i\}P\{Y=z-i\}$.

Now, you can use binomial distribution probability mass function for X as well as Y. That is what we have not discussed many problems in the earlier modules, when we started discussing the standard distributions, because the whole course we will be using those distribution again and again. So now, those who remember probability mass function of binomial distribution, you can directly write that is ${}^{n}C_i p^i (1-p)^{n-i}$. This is a probability mass function for a binomial distribution with the parameters n, p.

Similarly, you can write probability mass function of Y, that is ${}^{n}C_{z-i} p^{z-i}(1-p)^{n-z+i}$. You can go for simplification; this is p^i; this is p^{z-i}you can take the p^z outside. Similarly, $(1-p)^{n-z+i}$ and this is $(1-p)^{n-i}$. So, you can do some simplification, and after simplification you can get a the result that is ${}^{2n}C_z p^z (1-p)^{2n-z}$, the way I said $p^i p^{z-i}$that is killed you will get p^z.

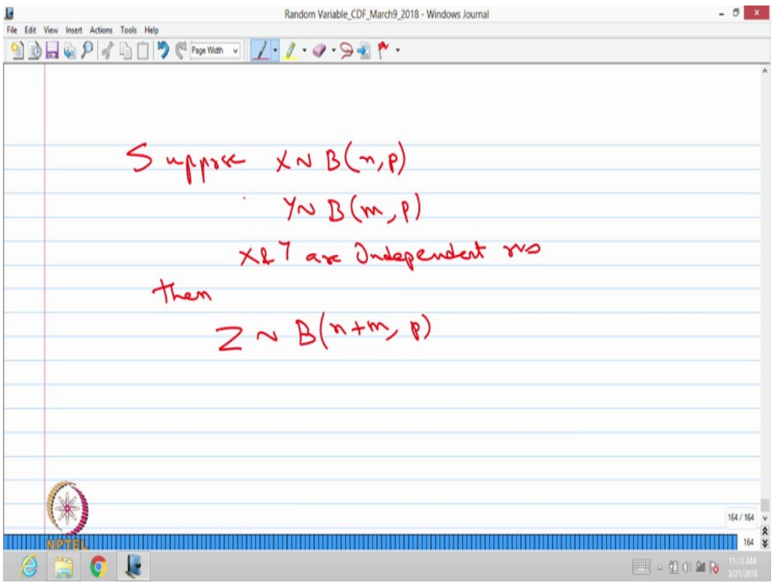
Similarly, $(1-p)^{n-z+i}$, if you simplify you will get $(1-p)^{2n-z}$. The only thing is the summation over i, ${}^{n}C_{i}$with the multiplication ${}^{n}C_{z-i}$, that is same as ${}^{2n}C_{z}$, that is the result which we are using $\sum_{i} {}^{n}C_{i}{}^{n}C_{z-i} = {}^{2n}C_{z}$.

So, this is the probability mass function of the random variable Z. And possible values of z is 0, 1, 2 and so on, till 2 n by seeing the probability mass function, which is greater than 0 in this value, 0 otherwise. You can conclude the random variable Z, which is also binomial distributor with the parameters, you have to map the probability mass function of this with the probability mass function of binomial distribution, then you will conclude that is parameters 2 n, p.

So, this is the use of a named or standard or common distribution, you do not know the distribution of Z you are finding the distribution of Z. After you get the probability mass function, this is same as probability mass function of a binomial distribution. Therefore, we conclude Z is also binomial distributor with a parameters 2 n, p. Now, one can discuss some vertex scenario.

Suppose X takes the binomial distribution with the parameter n, p, whereas, Y takes binomial distribution with some other parameter instead of n, suppose you treat some other positive integer m, but the next parameter is same p. let me write.

(Refer Slide Time: 27:57)



Suppose X is binomial distribution with n, p, and Y is binomial distribution with some other parameter m, p, and again I making the assumption both the random variables are independent. Then also you will get binomial distribution with a parameters n+m, p; that means, if you have a two independent binomial distribution with a different n and m. Whereas, the probability of success in each Bernoulli trial p is same for both the random variables, then the sum is also going to be a binomial distribution with the parameters sum of those first parameter, the second parameter is p.

That means, this can be extended for any n random variables, suppose each random variable is binomial distributed with some number some positive integer with p. And all are going to be a different number for first parameter, then the sum is going to be again binomial distribution, with sum of first parameters with second parameters p.

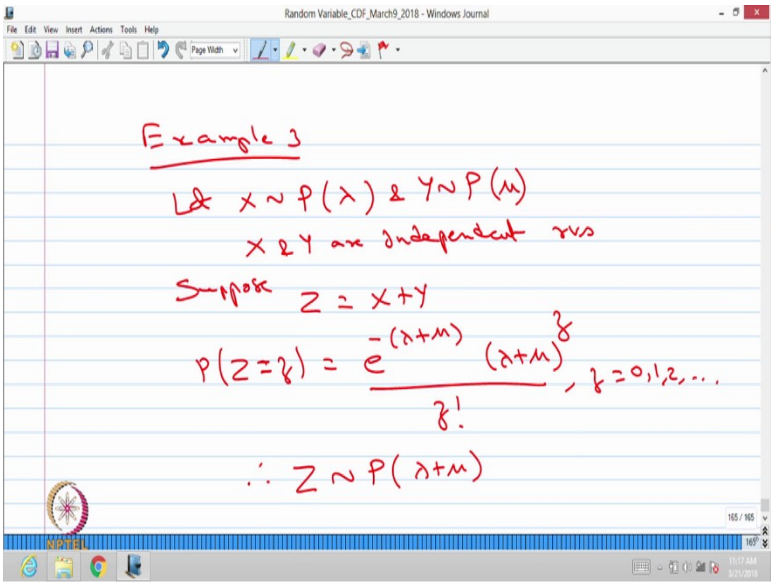
So, this can be generalized into any n mutually independent random variables. Since we have two random variables, we are using the word independent random variable. Once you have more than two random variables, for any n mutually independent random variables; each one is binomial distributed with the same probability of success p, with the different parameters in the first one then the sum is also going to be binomial distribution.

Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 30

I will move into one more example of a discrete type.

(Refer Slide Time: 00:03)



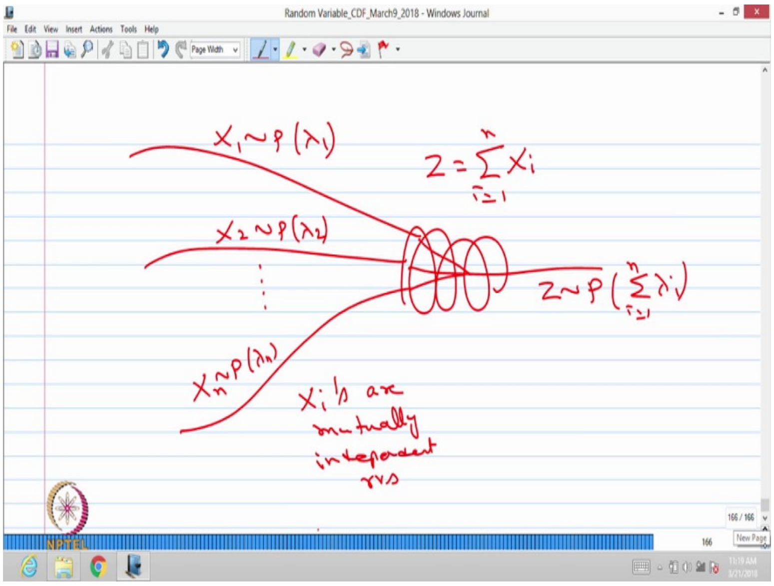
This is also going to be a very important result, that is, let X is Poisson distribution with a parameter λ and Y is again Poisson distribution with a parameter μ. And I make the assumption X and Y are independent random variables; independent random variables.

Suppose I create a random variable Z = X + Y. Similar derivation what we have done it for the binomial distribution. The similar derivation you can do and you can conclude $P\{Z = z\} = \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^z}{z!}$ where z can takes a value 0, 1, 2 and so on.

I am not giving the derivation we can do the similar derivation of the previous example you can get the probability mass function of Z is going to be this form, other than this z values it is going to be 0. Now we can map this with is there any standard distributions or common distribution matches we can find out.

So, this is going to be same as the probability mass function of Poisson distribution with a parameter $\lambda + \mu$. Therefore, one can conclude Z is also Poisson distribution with a parameter $\lambda + \mu$. That means, if you have two independent random variables both are Poisson distributed with some parameters then the sum is also going to be a Poisson distribution with a parameter is sum of their parameters.

(Refer Slide Time: 02:34)

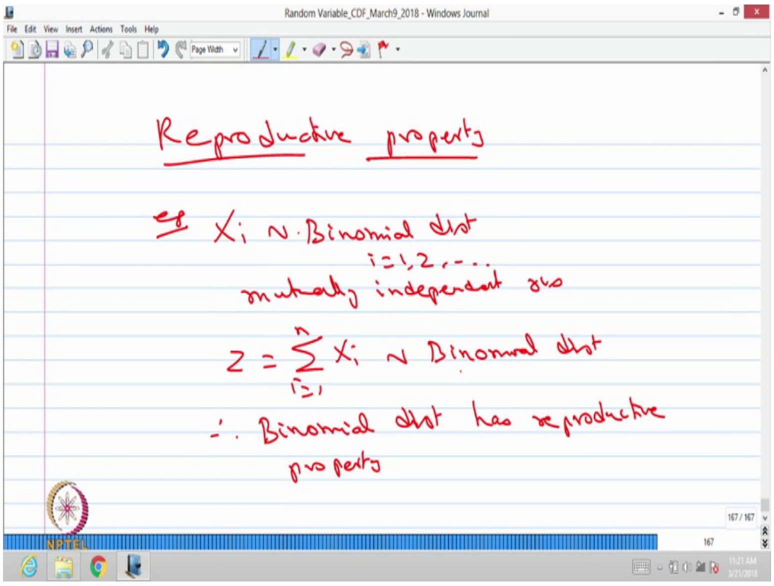


The same concept can be extended for any n random variables; that means, if n random variables are mutual independent, that means there is a one random variable that is the X_1 that is Poisson distributor with a parameter λ_1. There is another random variable X_2 that is also Poisson distributed with a parameter λ_2. Like that I have a n-th random variable that is also Poisson distributed with a parameter λ_n.

If I make a random variable which is nothing, but $\sum_i X_i$; that means, I will land up only one with the only one random variable by summing all the random variables that is Z, and this is going to be a Poisson distribution with sum of their parameters.

As long as all the X_i'sare mutually independent random variables; as long as all the random variables are mutually independent. Then the summation is going to be again Poisson distribution with a parameter is $\sum_i \lambda_i$. From these we are going to give one important property, that is called Reproductive property.

(Refer Slide Time: 04:03)



What the reproductive property says that if you have sequence of random variables and if you make a sum of those few of the variables out of it and all are having some distribution and after making the summation you are getting the same distribution of same as X_i's, or the original sequence of random variable then we conclude this random variable has; this particular random variables has a reproductive property.

That means for example, each X_i's are a binomial distributed and I have many random variables. All are mutually independent. I make the assumption all the random variables are mutually independent. Then if I make a random variable as the sum of few random variables out of this collection if that is also follows a binomial distribution. So, we can conclude binomial distribution has reproductive property.

Similarly, one can say the Poisson distribution also has a reproductive property whereas the Bernoulli distribution does not have a reproductive property. Because if you have a Bernoulli distributed random variable all are mutually independent, if you make n such random variable as a summation then that is going to be a binomial distribution, no more Bernoulli distribution. Therefore, Bernoulli distribution does not have a reproductive property.

Similarly, one can go for some common continuous type random variables. If you have a normal distributions all are mutually independent, if you make a summation then that is also going to satisfies; the reproductive property. That means, summation is also going to be a