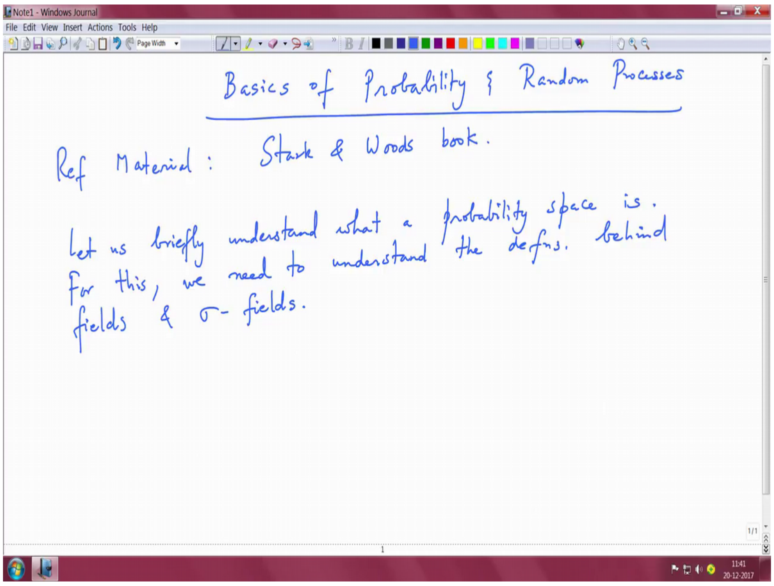
Mathematical Methods and Techniques in Signal Processing – I

Prof. Shayan Srinivasa Garani
Department of Electronic Systems Engineering
Indian Institute of Science, Bangalore

Lecture – 24
Basics of probability and random variables

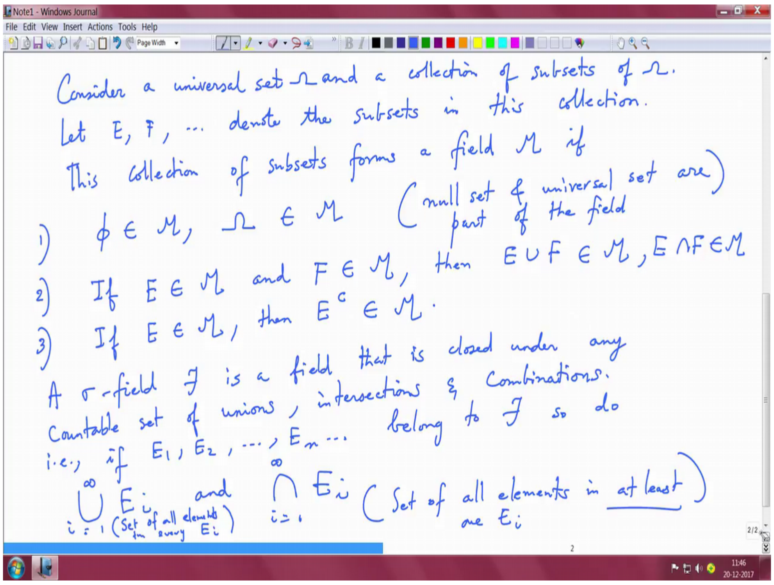
Let us, delve into the basics of probability and random processes as required for this course. So, I am not going to go in to details of probability and random processes because it is a separate course in itself, but I am just going to cover the material just for your understanding, so that you are comfortable to the rest of the course. So, let us first understand what a probability space is and for which we have to formally get into the details.

(Refer Slide Time: 00:49)



So, let us briefly understand what a probability space is. For this purpose, we need to understand the definitions behind fields and sigma fields, this is more in a modern set theoretic setup, one can think about the combinatorial way of looking into probability, but the modern version is a set theoretic approach. So, therefore, I will cover this from the set theoretic approach, just basics. So, that you are comfortable seeing through this material. So, now let us start with our understanding of what fields are, right.

(Refer Slide Time: 20:08)

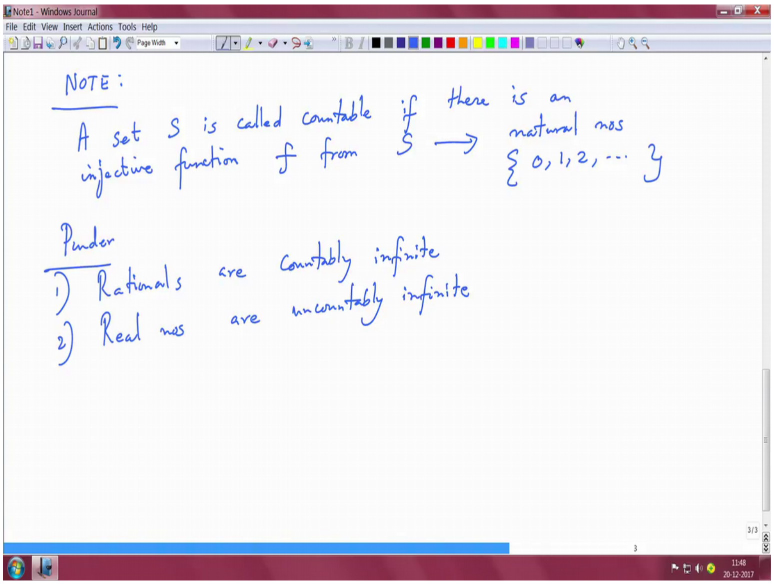


Now, consider a universal set omega and a collection of subsets of omega, you have a universal set and there looking into a collection of subsets of omega. Let E F so on; denote these subsets in this collection. Now, this collection of subsets forms a field script M if the following properties hold. So, what are the properties? So, the null set belongs to the field and the universal set belongs to the field, that is null set and universal set are part of the field.

Second, if the subset E belongs to the field M and subset F belongs to the field M, then the union of E and F belongs to field and so is the intersection of E and F that belongs to the field. The other property is, if E belongs to the field, then E complement belongs to the field. Now, these are the properties, a sigma field F is a field that is closed under any countable set of unions intersections and combinations of these unions and intersections, that is if E 1 E 2 and so on t E n maybe you can extend this, belong to the field I mean I said we can extend this E 1 e 2 so on, E n because you can have infinite subsets.

See, if they belong to the field so do the union of all these sets and the intersections to a set of elements in at least 1 of E i, this would be the intersection and here is a set of all elements in every E i right. So, this implies set of all elements in at least one E i, this is set of all elements in every E i.

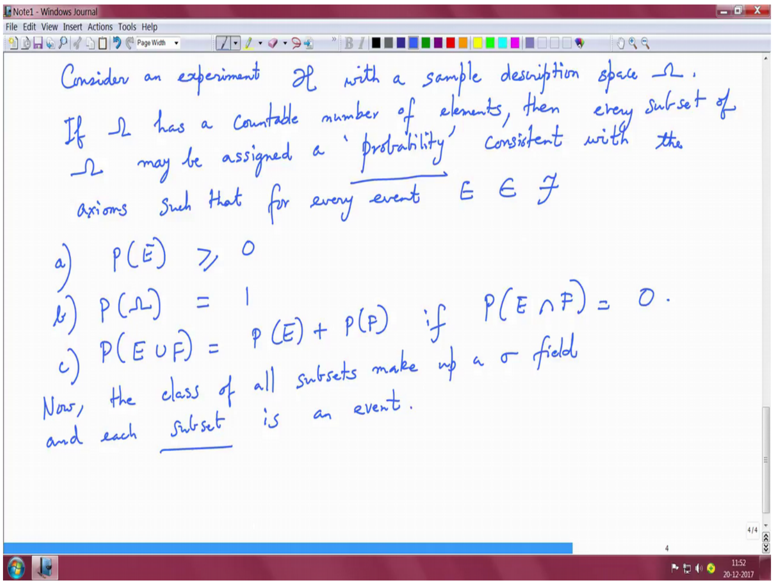
(Refer Slide Time: 07:07)



Now, there is this notion of countability, that I introduced and what we mean as follows, a set S is called countable if there is an injective function f from the set S to the set of natural numbers which is a 0, 1, 2, so on.

So, I think you have to ponder some things here, the set of rational numbers are countably infinite, whereas a set of real numbers are uncountably infinite. So, the details of, the proof of this countably infinite and uncountably infinite you can get into the details by looking into any analysis book and you may get some idea how to go about these.

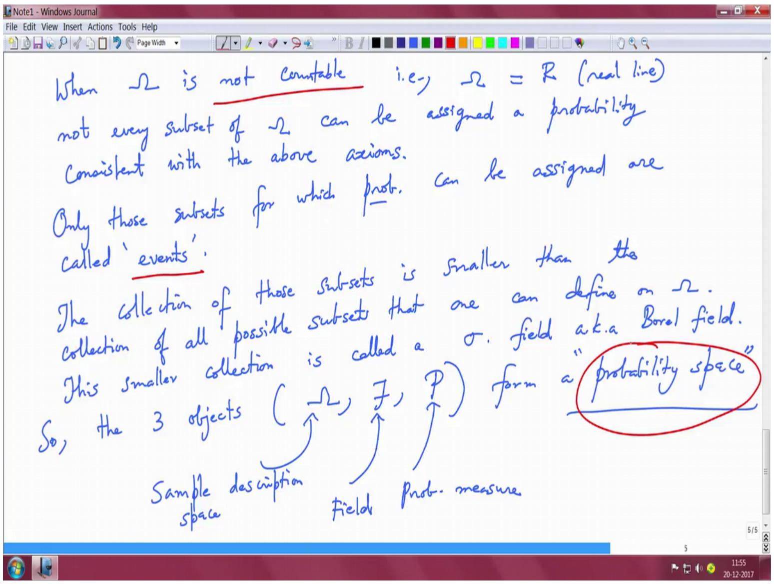
(Refer Slide Time: 08:50)



Now, let us go a little further, consider an experiment given by script H with a sample description space omega, if omega has a countable number of elements then every subset of omega may be assigned a probability, consistent with the axioms such that for every event E belonging to the field F, script F. The following hold 1, the probability of the event should be greater than or equal to 0.

Now, we are assigning some matrix, some measure to this. So, probability of the event has to be greater than or equal to 0, probability of omega, which is your universal set has to be 1 and probability of E union F equals probability of E plus, probability of F, if the probability of E into section F is 0. Now, the class of all subsets make up a sigma field and each subset is an event. So, here we are defining first, what an event is, in terms of subsets and that is an important discussion that we need to make.

(Refer Slide Time: 11:48)

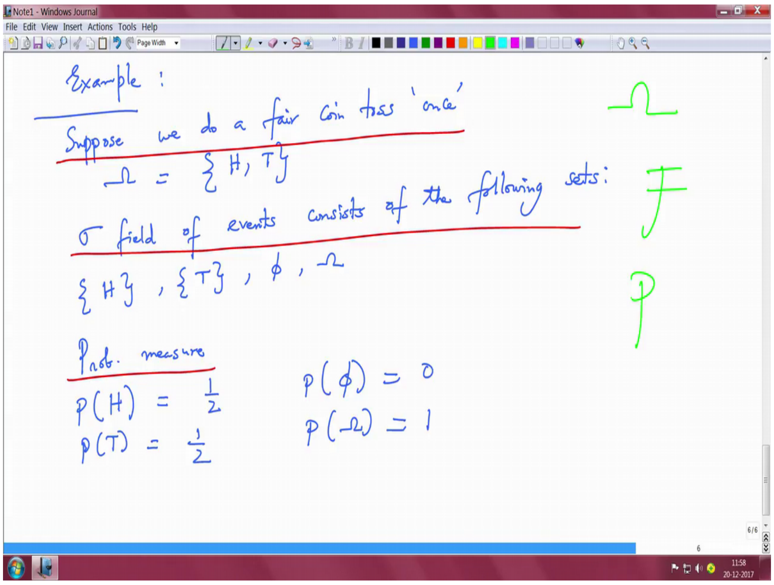


When omega is not countable, is an important point and when can you have something which is not countable, that is omega is possibly the real line not every subset of omega can be assigned a probability consistent with the above axioms, we listed the axioms in the previous slide.

So, only those subsets for which probability can be assigned or called events, now the collection of those subsets is smaller than the collection of all possible subsets, that one can define on omega and this smaller collection is called a sigma field. It is also known as a Borel field. So, the 3 objects omega, the field and the probability measure form a probability space, just to recall omega is the sample description space, F is the field and we have seen what properties of field must satisfy and then, this is the probability measure, consistent to the axioms which we described.

So, when somebody talks about probability space, I think you have to understand that we mean that we are referring to these 3 objects, which are part of this space that is the sample description space, the field and the probability measure. This is an important concept in the context of the modern set theoretic framework for probability and often you will require this notion in your research or in your higher understanding. Let us, just revisit in this context the simple coin toss experiment and see how we can relate these objects.

(Refer Slide Time: 15:46)

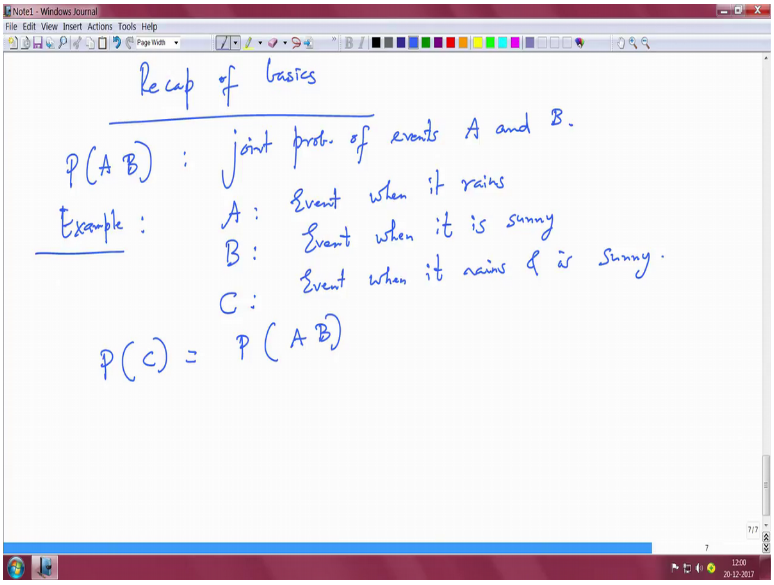


So, suppose we do fair coin toss once, that means, we have a collection of 2 subsets, which are the head and the tail outside of the null set and the universal set. So, universal set is basically comprises of the head and the tail. The sigma field of events consists of the following sets, it could be a head, it could be a tail, it could be a null set, it could be the universal set.

See the distinction between the probability description space and the sigma field and then, we have the probability measure, the probability of measure is basically measure on this sigma field of events. So, the probability of the head since it is a fair coin toss it is 1 half and probability of tail is also 1 half, probability of a null set is a 0, probability of the universal set is 1. So, this gives you a sort of an idea, what we are referring to. So, this is our object omega, this is our object, this is a field, this is the probability measure P and this is the probability space for a fair coin toss, the experiment.

Now, let us recall a few other basics and we will try to recap a few things.

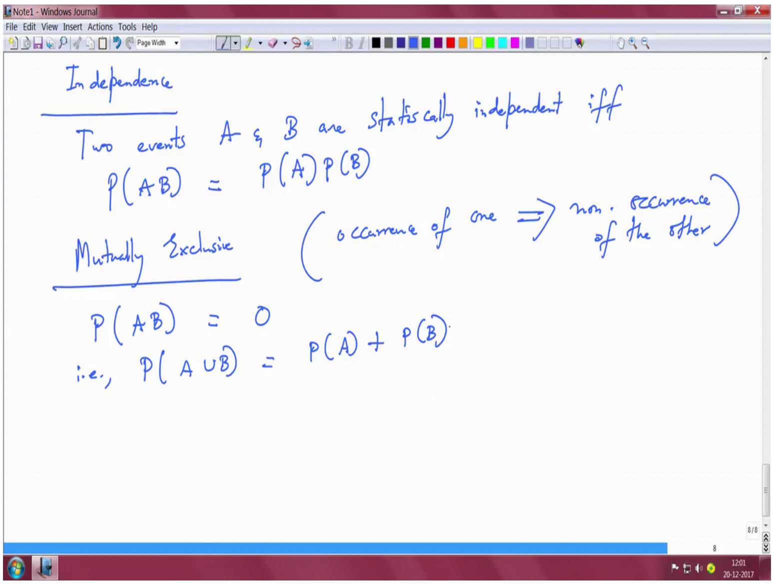
(Refer Slide Time: 18:06)



Probability of the joint occurrence of 2 events A and B is given by probability of A and B how to read this as probability of A and B, this is basically the joint probability of events A and B. So, an example is A is an event, when it rains B is an event, when it is sunny C could be a joint event, when it rains and is sunny probably you might sense a rainbow. So, probability of this event C is the joint of A and B, probability of the occurrence of event A and B and that is what we mean, this is basically an intersection.

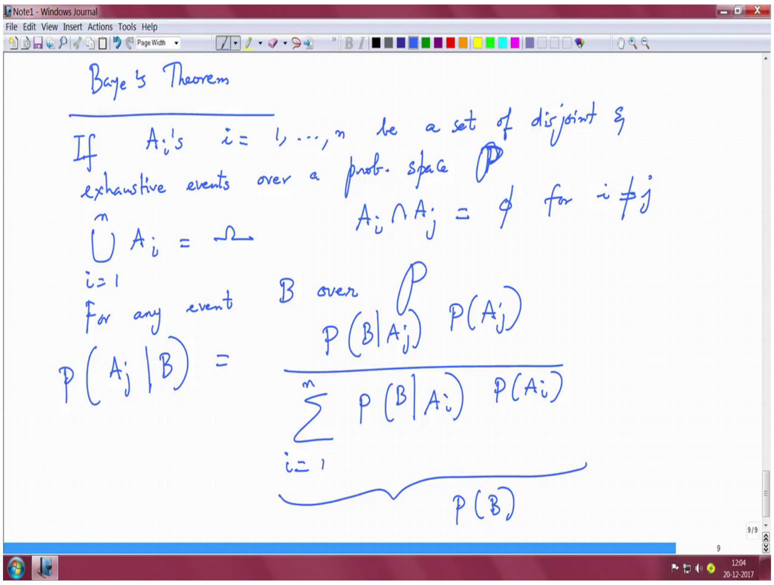
Now, one can think about how to do this from a combinatorial way, that means, you can count the number of events when it rains, count the number of events when it is sunny, count the number of events when it is raining and it is sunny over the number of experiments that you conduct and then you can bring this notion of what the probability measure has to be, so this is some basics. Let us also recall a few definitions.

(Refer Slide Time: 20:04)



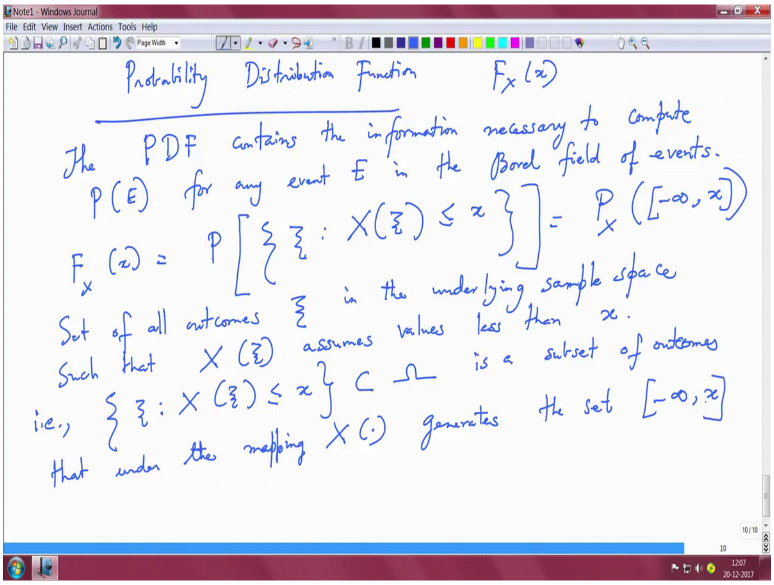
Independence, 2 events A and B are statistically independent. If and only f probability of A and B that has joint occurrence of probabilities, the joint occurrence of events A and B, the probability of the joint occurrence of the events A and B can be decomposed as probability of A times probability of B. There is also a notion of mutually exclusive property. So, what it means is, occurrence of 1 implies non occurrence of the other. So, in this case probability; the joint probability of A and B is 0 and therefore, the probability of A union B is probability of A plus probability of B and then we should also know the Baye's theorem, I think all of you would have seen this from your high school but just to recap.

(Refer Slide Time: 21:45)



If Ai's i equals 1 to n be a subset or to be a set of disjoint and exhaustive events over a probability space P some script P here, then the union over all these Ai's is omega. Let us assume that A i intersection A j is null for i naught equal to j, that is take 2 events and it there they do not jointly occur channel, then for any event B over this probability space P the probability of A j given this event B is probability of B given A j times, probability of A j divided by summation, i going from 1 to n probability of B given A i times probability of A i. So, this is basically probability of B and we have expressed the joint of probability of A j and B in into different ways using the Baye's theorem. The proof is pretty straight forward from the basic definition of conditional probability and I will leave this as sort of an exercise for you to ponder about.

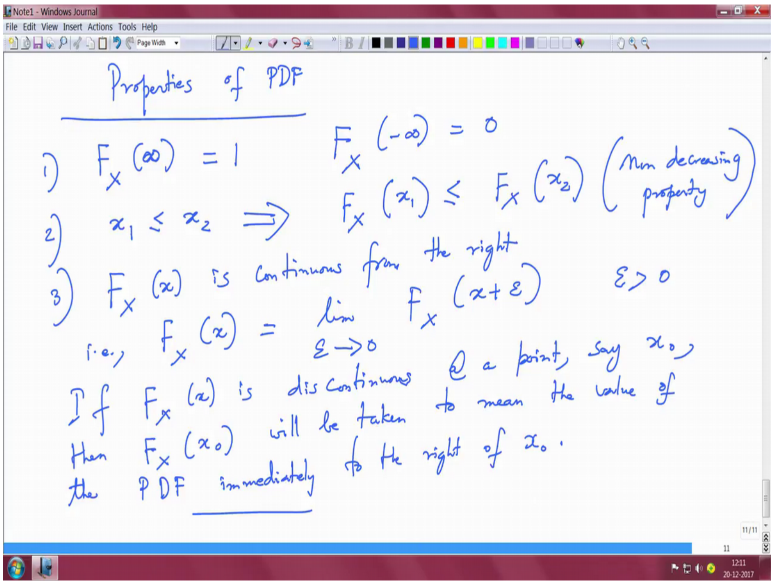
(Refer Slide Time: 24:00)



So let us, now get into the understanding of what probability distribution function is. This is also called the cumulative distribution function. The probability distribution function contains the information necessary to compute the probability of an event E, for any event E in the Borel field of events . So, if you get into the mathematical notation, the probability distribution function is basically the probability over all zeta such that under the mapping x of zeta you are looking at a collection of all zeta such that, this mapping is less than or equal to some chosen x. So, this is given as the probability, the random variable over the event minus infinity to some small x some chosen number.

So, this is basically set of all outcomes zeta in the underlying sample space such that x of zeta assumes values less than some chosen small x. I give some small x and I want to figure out under the mapping x, the set of all outcomes zeta such that x's zeta is less than or equal to that chosen quantity. Now, as you can see the set of all zeta such that x of zeta is less than or equal to x is contained within omega and this is a subset of outcomes, that under the mapping x generates the set minus infinity to x. Now, we have a few properties for this probability distribution functions, I am just going to enumerate them.

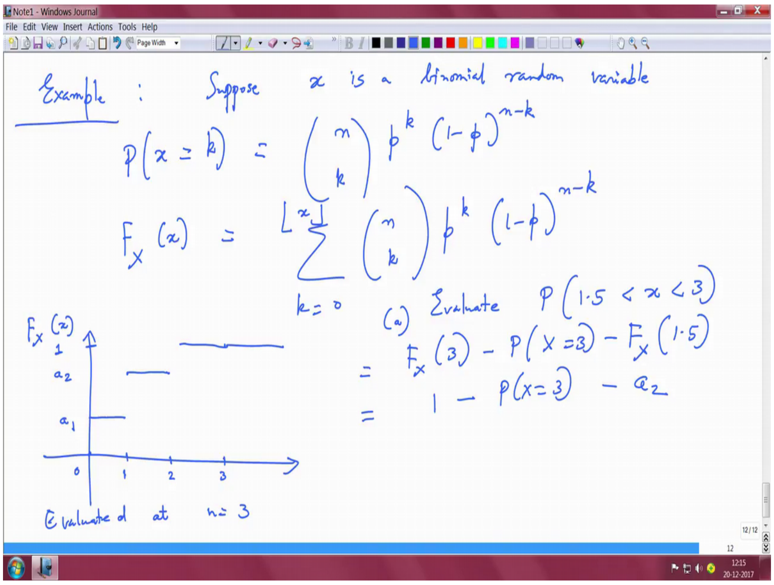
(Refer Slide Time: 27:40)



First, the probability distribution function evaluated at infinity is 1, F infinity of f x of minus infinity is a 0. if I consider 2 quantities small x 1 less than or equal to small x 2, this implies that the probability distribution function evaluated at small x 1 is less than or equal to the probability distribution function evaluated at small x 2.

This is basically the non decreasing property of the cumulative distribution function or the probability distribution function. Then the there is the third property, capital F x of small x is continuous from the right, that is the probability distribution function is right continuous. What it means is, capital F subscript capital x of x is basically limit as epsilon goes to 0 of the distribution function, probability distribution function evaluated at x plus epsilon slightly at epsilon greater than 0. If the distribution function is discontinuous at a point say x naught, then the distribution function at x naught will be taken to mean the value of the probability distribution function immediately to the right of x naught. So, if at that point x naught, if it is discontinuous then you will consider the value of the probability distribution function immediately to the right of the point x naught and that is why we have the right continuous property here.

(Refer Slide Time: 31:01)

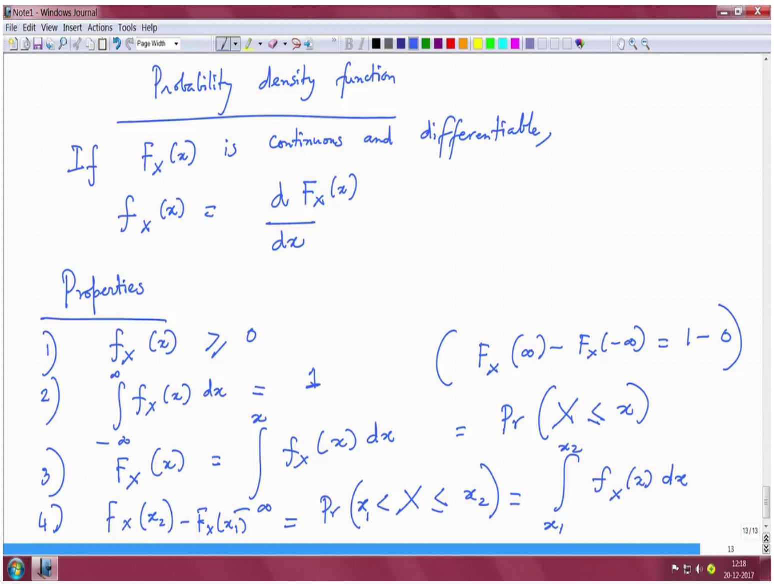


Now, let us consider a simple example, suppose x is a binomial random variable then probability that x equals some k is given by n choose k p power k 1 minus p power n minus k this is the E probability mass function and the probability distribution function is given by this quantity where you basically look at the floor of x, some all the masses from k equals 0 to h o c floor of x. Now, how should this look like if you just evaluate for n equals 3. So, when k equals 0 is 0. Let us, plot this probability distribution function, it has to let us say we have 3 values 0, 1, 2 and 3. At 3 definitely this has to be 1, because you do not have anything more than 3, so therefore, this has to be 1.

Let us assume that we accumulate the mass till 1 and that quantity is a 1 and then there is a discontinuity then we have we accumulate a mass at 2 and that is a 2 and then we have a mass at 1, a mass of 2 and basically after 3 it is basically a constant. So, if you want to evaluate probability that 1.5 is less than x, less than 3. So, x is some number given to you. Then, this is basically the distribution if you want evaluate this probability this is the distribution at 3 you assume slightly though it is less than 3 you assume the accumulation at 3 minus the probability that x equals 3 minus the cumulation or the probability distribution function evaluated at the point 1.5 because you to subtract that mass here.

So, this would be basically 1 minus probability that x equals 3 is what you can compute minus this point F x of 1.5 is this quantity a 2 reading off from this label here. This is how you would actually compute the probability distribution function. So, I gave you the probability mass function here because I mean this is basically the probability that x takes a certain value this may not be so direct and straightforward, when we look at continuous random variables because first we start with the probability distribution function and the distribution function is continuous and differentiable, then we can basically get the probability density function. So, let me define what the probability density function is, if the probability distribution function is continuous and differentiable then the density function is basically a derivative of the probability distribution function and we assume the derivative exist . So, therefore, we can take the derivative.

(Refer Slide Time: 35:22)

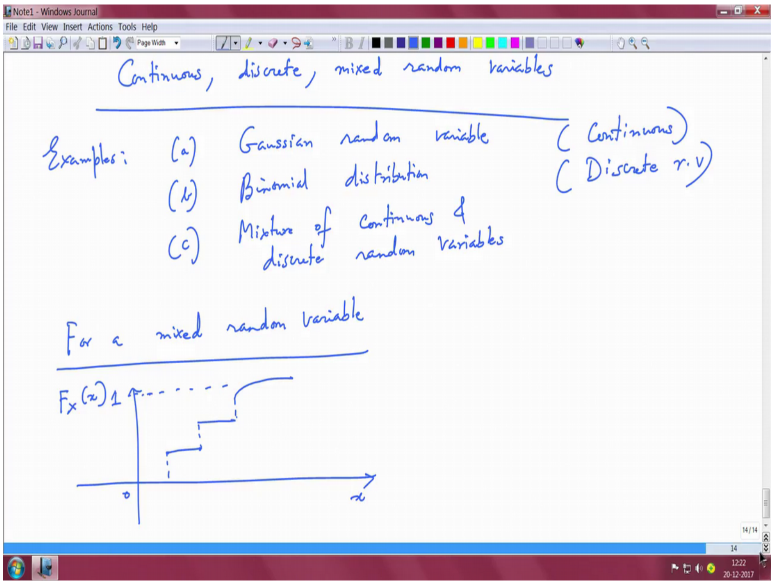


So, the properties for the probability density function are as follows, 1 the density function; probability density function is nonnegative, right it makes no sense to say that the mass probability mass function or density function is negative. If there is no meaning to that negative quantity there because it is a derivative of the distribution function. So, there is no meaning to this mass which is negative, so it is nonnegative. Then it integrates the density should integrate 2 1, why because you evaluate. How do you compute this? This quantity, this is basically from the probability distribution function it is F x of infinity minus F x of minus infinity which is basically 1 minus 0 and it is that is why you get it as 1.

Now, the distribution function is linked to the density function, if you integrate the density function from minus infinity to a point x, some constant x then you get the distribution; probability distribution function evaluated at x. This is giving you the probability that the random variable X capital X is less than or equal to some constant small x. Now, if you want to compute the difference in the probability distribution functions evaluated 2 points x 1 and x 2, which is giving you the probability essentially that random variable X is between x 2 and x 1, that can be obtained by integrating the probability density function between limits x 1 and x 2.

So, these are some basic properties you would have seen all of these, but now in the context of sets etcetera you will be able to better appreciate what is really happening here and if you think about the probability density function, so the probability for continuous random variables, probability at a certain point is essentially 0, I mean that makes sense because if you shrink this integral, if x 1 and x 2 coincide or x 2 is just epsilon from x 1 and if you just collapse that interval basically that gives you 0 for the probability and therefore, the probability for a continuous random variable the probability that the random variable takes a certain value is basically 0 and that should be obvious.

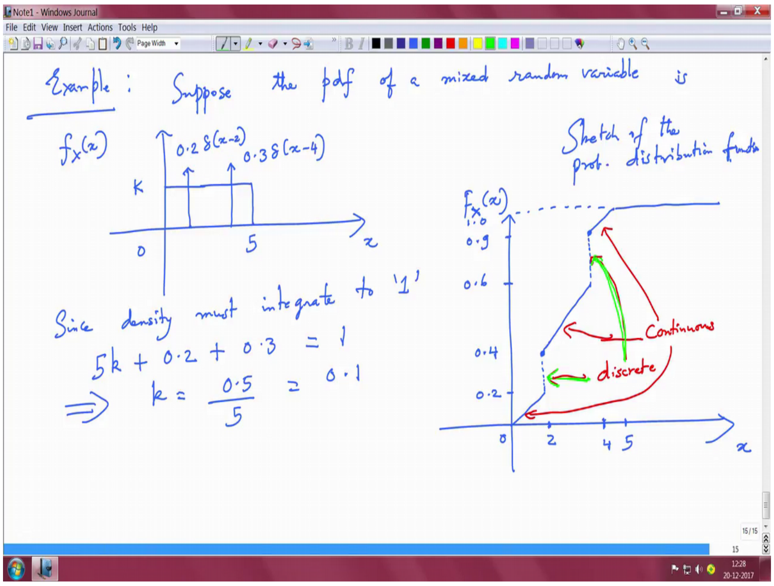
(Refer Slide Time: 39:44)



So, with this I will sort of discuss some subtleties between continuous discrete and makes random variables. So, random variables can be continuous, they could be discrete and they could be mixed. So, examples the Gaussian random variable is 1 example for a continuous random variable, then we have the for example, look at the binomial distribution that is an example of a discrete random variable, sometimes you can have a mixture of continuous and discrete random variables. So, a mixed random variable I can give you an example, so for a mixed random variable if you look at the sketch of the probability distribution function, I can have jumps indicating if is discreet at some points and then I may have something which is continuous and of course, this has to integrate to 1.

So, this is an example for a mixed random variable. So, for a continuous random variable you can just imagine that the pdf is basically continuous, for the discrete case we have seen an example of a staircase function is where things are discrete.

(Refer Slide Time: 42:29)



Now, we will try to work out a simple example considering a mixed random variable, suppose the pdf of a mixed random variable is given as follows. This is uniform from 0 to 5, there is a height some K that I do not know, at 2 I have a mass and impulse, which is height 0.2, this is delta x minus 2 and at 4 I have another impulse, which is 0.3 delta x minus 4. Now, under any circumstance the density must integrate to 1.

So, since density must integrate to 1, so let us look at the integration here. So, this is the area of this rectangle which is 5 k, 5 is the base and k is the height, 5 times k plus you have a mass of 0.2 at 2 (Refer Time: 44:10) mass of a 0.2 and another mass of 0.3 at x equals 4. So, this must sum to 1. So, this implies k is 0.5 upon 5, which is 0.1 and if you sketch the pdf that would be pretty interesting because let us sketch this, sketch of the probability distribution function. So, between 0 and 2, I have a ramp because I am just integrating this rectangle here. So, the I am with a slope of 0.1 I just have a ramp here and I accumulate a mass of point 2 here on the y axis, then at 0.2 I have a jump of point 2. So, I basically accumulate a mass here of 0.2, I think I can indicate this jump perhaps through dotted lines.

So, I accumulate a jump of 0.2 at x equals 2. So, I go to 0.4, now between 2 and 4 I do not have any impulse. So, I basically keep integrating it, till I accumulate a mass of 0.6 and at 4 I have an impulse with a mass of 0.3's accumulate a mass of 0.3 here and I get to this point which is 0.9 and after 4, between 4 and 5 just immediately after 4 I just have a ramp with a slope of 0.1 and then I touch this, this is 1, so this point is 5.

So, this is a sketch of the probability distribution function. So, this gives you an idea, so at these points it is a discrete and at these points it is continuous, this is continuous and these 2 points are basically discrete. So, maybe I can put a green color here, indicating it is discrete and red color indicates it is continuous. So, this is sketch of the probability distribution function for a mixed random variable. So, hopefully with these basics you may be able to understand and appreciate the subtle details, when you actually look into the probability distribution functions and probability density functions, which will be useful for your analysis, further during the course when you have to deal with these probabilities.