Measure Theoretic Probability 1
Professor Suprio Bhar
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Lecture 17
Distribution Function of an RV

Welcome to this lecture. Before we continue the discussion about the topics in this lecture, let us quickly recall what we have done in the previous lecture. So, in the previous lecture I looked at the random variables or random vectors defined on a probability space and then what we considered was the measurable structure of such functions and we combined it with the probability measure on the domain side to construct a probability measure on the range side.

So, for random variables the range will be the measurable space real line together with the Borel σ-field, for random vectors if they are taking values in $\mathbb{R}^d$, then the range side will be $\mathbb{R}^d$ together with the Borel σ-field of $\mathbb{R}^d$.

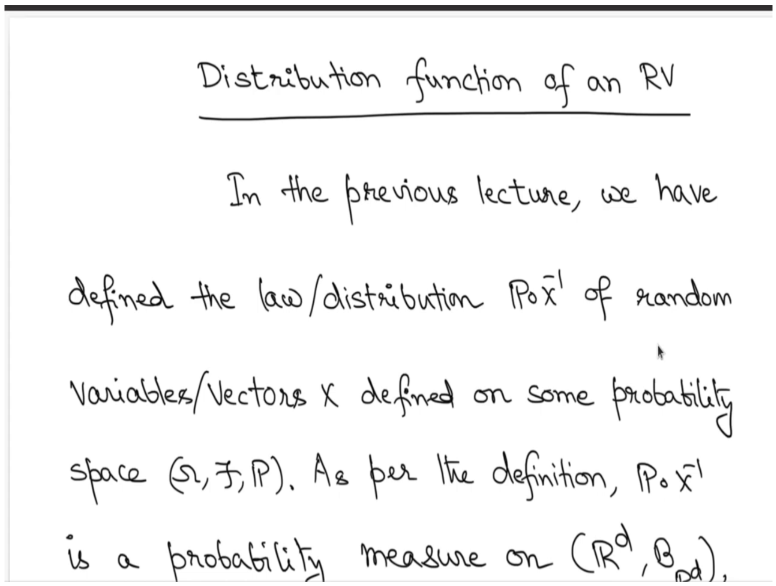
So, on such range spaces, range side measurable spaces we constructed these probability measures that was combined together with the measurable structure of X, the random variables or random vectors together with the probability measure $\mathbb{P}$, and we called it $\mathbb{P} \circ X^{-1}$, the notation was $\mathbb{P} \circ X^{-1}$, and we called it the law or distribution of the random vector or random variable X.

So, we constructed this and we saw one simple example given by the Dirac mass, and then we left as exercise, some of the other related examples. So, please compute the laws and check how they come out as combinations of Dirac masses. So, those exercises will give you examples of convex linear combinations of Dirac masses. So, in this lecture we are now going to continue discussions about the probability measure that we obtained and we are going to construct a very important function out of this.

Again, as seen in the previous lecture many of the probability of the events that we considered in your basic probability courses can now be reinterpreted as probability measure of the corresponding set that is appearing on the Borel σ-field, or the real line or the $\mathbb{R}^d$.

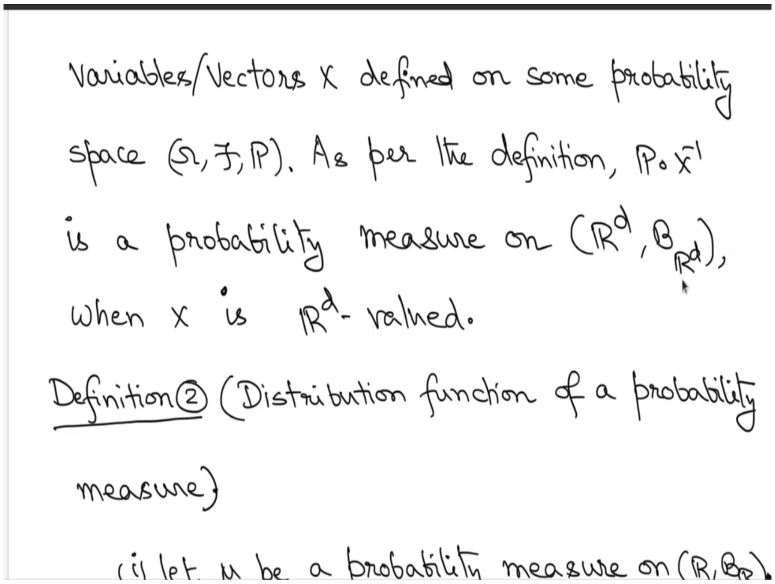
A similar thing is going to happen here we are going to consider a very important function that you have seen in your basic probability course and we are going to obtain it from this specific probability measure the law or distribution of the random variable or random vector X, the probability measure which is $\mathbb{P} \circ X^{-1}$. So, let us continue and move on to the slides.

(Refer Slide Time: 02:43)



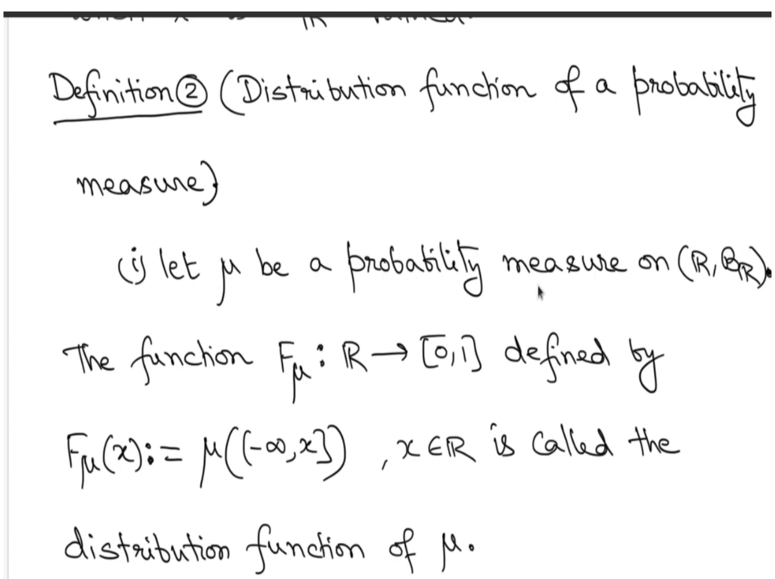
So, in the previous lecture we have defined the law or distribution $\mathbb{P} \circ X^{-1}$ of the random variables or random vectors X.

(Refer Slide Time: 02:53)



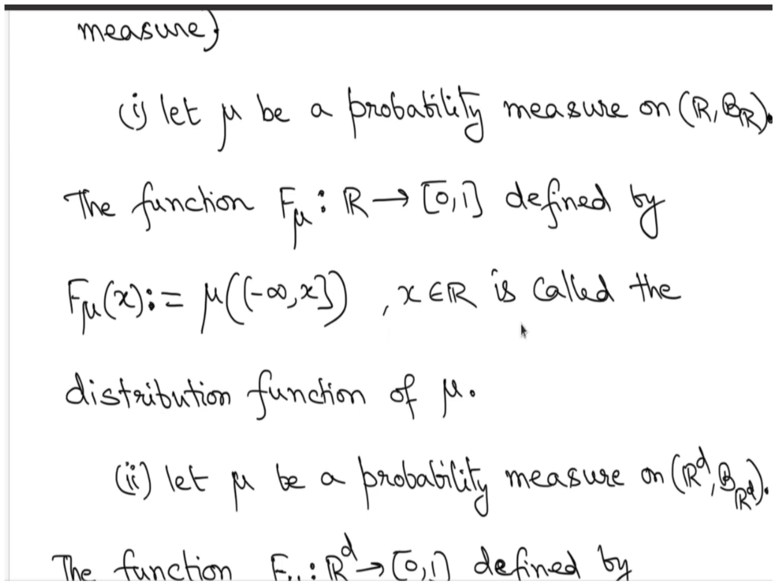
And they are supposed to be defined on some given probability space $(\Omega, \mathcal{F}, \mathbb{P})$. And as for the definition $\mathbb{P} \circ X^{-1}$ became a probability measure on this space. So, if you are considering this $\mathbb{R}^d$ valued random vector then you just get $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ that is measurable space. So, on top of that you get this probability measure.

(Refer Slide Time: 03:15)



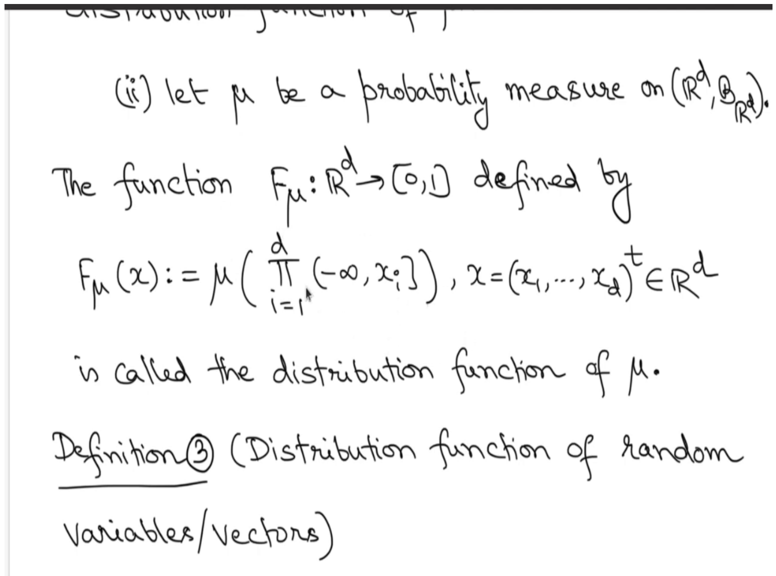
And now we are going to define something called the distribution function of a probability measure. So, this is a general definition of a function corresponding to a probability measure on $\mathbb{R}^d$ or $\mathbb{R}$. So, either of these spaces. So, let us first look dimension one just to understand the concept. So, start with the measurable space Borel σ -field on top of the real line, so that is the measurable space and consider μ to be a probability measure on there.

(Refer Slide Time: 03:42)



So, now consider this function F_μ defined on the real line and taking values between 0 and 1. So, what is this function? This function for any real number x takes the size of the set $(-\infty, x]$. So, that, that interval you look at, $(-\infty, x]$. That interval you look at, look at the size of that under the probability measure mu, and consider that as the value of the function at the point x. And this is called the distribution function of the probability measure mu.

(Refer Slide Time: 04:18)



And for the d-dimensional case you have the exact analog of the same definition. So, now you are defining the function on $\mathbb{R}^d$ and it will again take values between 0 and 1. What you have to do? Look at a point $x \in \mathbb{R}^d$, so it is a vector with components x 1 up to x d, and for such a point x, such a vector in $\mathbb{R}^d$, what you look at is this rectangle.

So, the ith coordinate is simply this interval $(-\infty, x_i]$. So, it is a product of such intervals. So, it is the default product of intervals of this type where the ith component is simply coming from the ith coordinate of point x. So, you get this rectangle and this is a set in $\mathbb{R}^d$, this is a Borel set in $\mathbb{R}^d$ and therefore you can talk about the measure of that under the probability measure mu.

And then you get a value and that value you assign it to this function at the point x and this is called the distribution function of μ in the d-dimensional setup. So, exact analog of whatever you have seen in the dimension one.

(Refer Slide Time: 05:29)

Definition ③ (Distribution function of random variables/vectors)

let X be a random variable/vector on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The function $F_{\mathbb{P}\circ X^{-1}}$ of the law $\mathbb{P}\circ X^{-1}$ is called the distribution function of X.

But then, with that as the motivation you can now define distribution function of random variables or random vectors. So, what do you do? Take a random variable or random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Correspondingly you get this law which is $\mathbb{P} \circ X^{-1}$, and that is defined on $\mathcal{B}_{\mathbb{R}}$ for the case of random variables for $\mathbb{R}^d$ valued random vectors you get it on $\mathbb{R}^d$ with the Borel σ-field of $\mathbb{R}^d$.

Now, corresponding to that probability measure you get a distribution function as defined above. Consider that. So, that is nothing but, as per our notation $F_{\mathbb{P}\circ X^{-1}}$. This function is called the distribution function of X.

So again, for a random variable or a random vector you get the law which is the probability measure on the range side and on the range side you have defined, for any probability measure you have defined this distribution function, take that. So, call that as the distribution function of X.

(Refer Slide Time: 06:26)

distribution function of X.

<u>Note ⑤</u>: To simplify the notation, we write F_X instead of $F_{\mathbb{P}\circ X^{-1}}$.

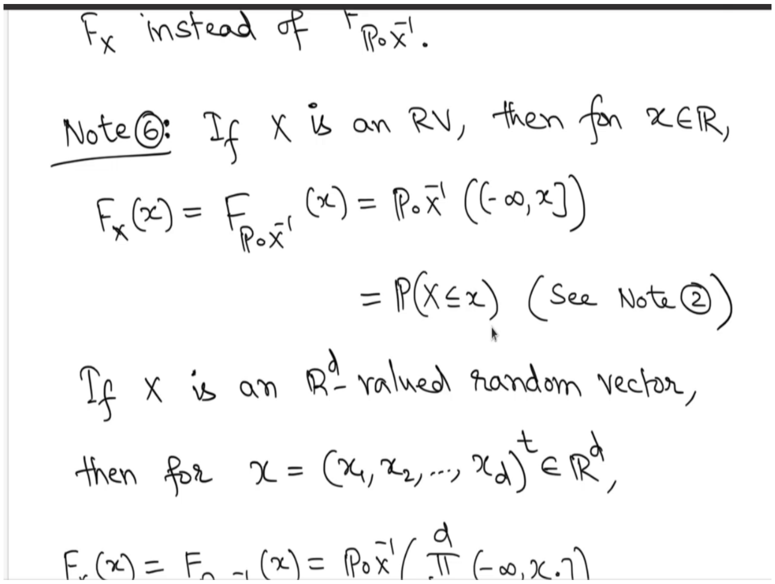
<u>Note ⑥</u>: If X is an RV, then for $x \in \mathbb{R}$,

$$F_X(x) = F_{\mathbb{P}\circ X^{-1}}(x) = \mathbb{P}\circ X^{-1}((-\infty, x])$$

$$= \mathbb{P}(X \leq x) \quad (\text{See Note ②})$$

And now, to simplify the notation what we are going to do, is to forget about this probability measure now. So, it will be given to us but we will simplify the notation and instead of writing $F_{\mathbb{P}\circ X^{-1}}$ we will simply write F_X. So, this is a distribution function of X.

(Refer Slide Time: 06:49)

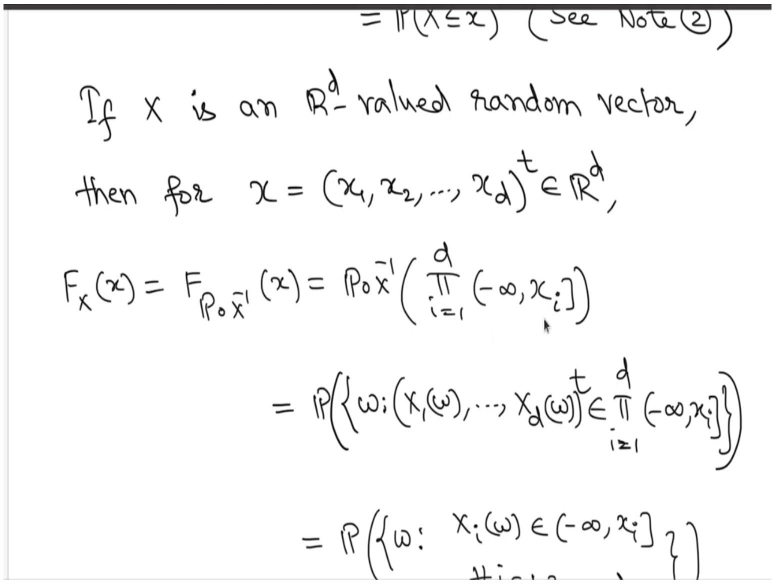


But then follow the notations mentioned in the previous lecture and let us rewrite this, the distribution function of X, let us try to connect it with probability of events. What is this? So, if X is a random variable, so you are working in dimension 1, you get back some nice familiar expressions.

So, for any point x in the real line, so as defined above $F_X(x)$ is nothing but the $F_{\mathbb{P}\circ X^{-1}}(x)$. This is just notation.

But then as for the definition this is $\mathbb{P} \circ X^{-1}(-\infty, x]$. So, you are looking at the size of the set, size of the interval $(-\infty, x]$ under the probability measure $\mathbb{P} \circ X^{-1}$. And as for the notation introduced in the previous lecture, this is nothing but the probability of the event $X \leq x$. And therefore, you get back your familiar distribution function corresponding to a random variable. So, this simply turns out from the definition of the law.

(Refer Slide Time: 07:49)



And similar expressions can be derived for the random vector case. So, again if you are working in d-dimensions, take a point x, which is made up of these components $(x_1, \ldots, x_d)$, this is a vector in $\mathbb{R}^d$. So, try to compute the distribution function at the point x.

So, that is nothing but, as per the definition $\mathbb{P} \circ X^{-1}$ of this default product of the intervals. So, you are looking at this default product of the intervals and looking at the probability measure of that. So, under the measure $\mathbb{P} \circ X^{-1}$.

(Refer Slide Time: 08:20)

$$F_X(x) = F_{\mathbb{P}\circ X^{-1}}(x) = \mathbb{P}\circ X^{-1}\left(\prod_{i=1}^{d}(-\infty, x_i]\right)$$

$$= \mathbb{P}\left(\left\{\omega : (X_1(\omega), \ldots, X_d(\omega))^t \in \prod_{i=1}^{d} (-\infty, x_i]\right\}\right)$$

$$= \mathbb{P}\left(\left\{\omega : \begin{array}{l} X_i(\omega)\in(-\infty,x_i] \\ \forall i = 1,2,\ldots,d\end{array}\right\}\right)$$

$$= \mathbb{P}\left(\{\omega : X_i(\omega)\le x_i, \forall i=1,\ldots d\}\right)$$

$$= \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d).$$

But then, what you are saying is that you are looking at all points in the domain, ω such that the random vector evaluated at the point ω, that is nothing but $\{X_1(\omega), X_2(\omega), \ldots, X_d(\omega)\}$. So, that is the d-components. So, that thing should land up in this default product.

(Refer Slide Time: 08:42)

$$= \mathbb{P}\left(\left\{\omega : (X_1(\omega), \ldots, X_d(\omega))^t \in \prod_{i=1}^{d} (-\infty, x_i]\right\}\right)$$

$$= \mathbb{P}\left(\left\{\omega : \begin{array}{l} X_i(\omega)\in(-\infty,x_i] \\ \forall i = 1,2,\ldots,d\end{array}\right\}\right)$$

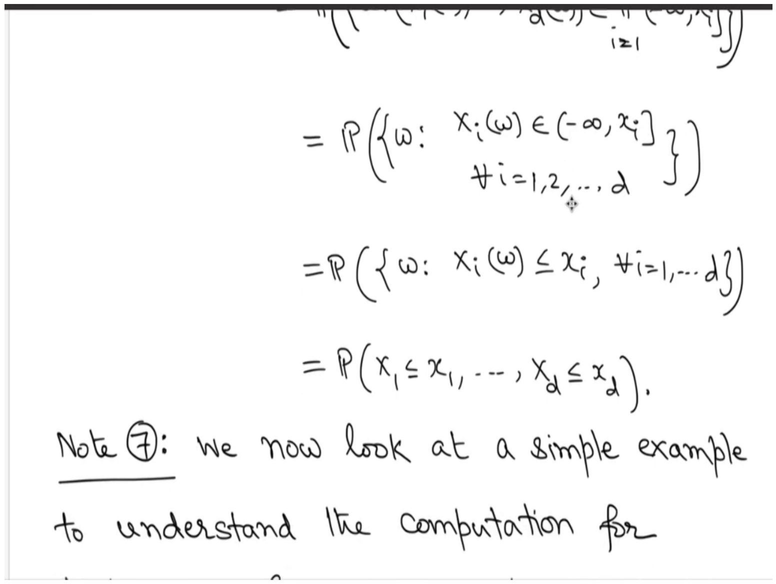
$$= \mathbb{P}\left(\{\omega : X_i(\omega)\le x_i, \forall i=1,\ldots d\}\right)$$

$$= \mathbb{P}(X_1 \le x_1, \ldots, X_d \le x_d).$$

Note 7: We now look at a simple example

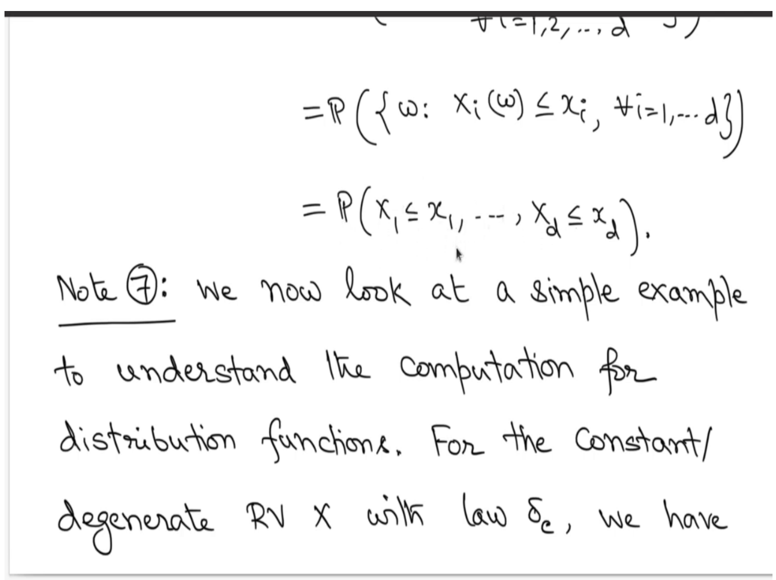
And therefore, what you are saying is that you are looking at all points on the domain side such that individual components $X_i(\omega)$ exactly falls in these intervals $(-\infty, x_i]$.

(Refer Slide Time: 08:55)



And that can now be rewritten in terms of $X_i(\omega) \leq x_i$. This is simple inequalities, you can write down. But this should happen for all the coordinates $i = 1, \ldots, d$.

(Refer Slide Time: 09:10)



But then, you can now rewrite using your familiar notation suppress the ω and simply rewrite it in terms of $\mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d)$, and that is nothing but the distribution function of the random vector x. Again, you are getting back the familiar expression.

(Refer Slide Time: 09:30)

$$= \mathbb{P}(X_1 \le x_1, \dots, X_d \le x_d).$$

Note ⑦: we now look at a simple example to understand the computation for distribution functions. For the constant/degenerate RV X with law δ_c, we have

for $x \in \mathbb{R}$,

$$F_X(x) = \mathbb{P} \circ X^{-1}((-\infty, x]) = \delta_c((-\infty, x])$$

Now, we go ahead and look at certain simple examples to understand these computations of distribution functions. So, in the previous lecture we computed the law for this constant or degenerate random variable, which was taking value some constant given constant value, and the law was exactly the Dirac measure situated or supported at that point, at that constant value. So now we want to compute the distribution function corresponding to this random variable.

(Refer Slide Time: 09:49)

for $x \in \mathbb{R}$,

$$F_X(x) = \mathbb{P} \circ X^{-1}((-\infty, x]) = \delta_c((-\infty, x])$$

$$= \begin{cases} 0, & \text{if } c \notin (-\infty, x] \\ 1, & \text{if } c \in (-\infty, x] \end{cases} = \begin{cases} 0, & \text{if } x < c \\ 1, & \text{if } x \ge c. \end{cases}$$

we now study the properties of distribution functions of probability

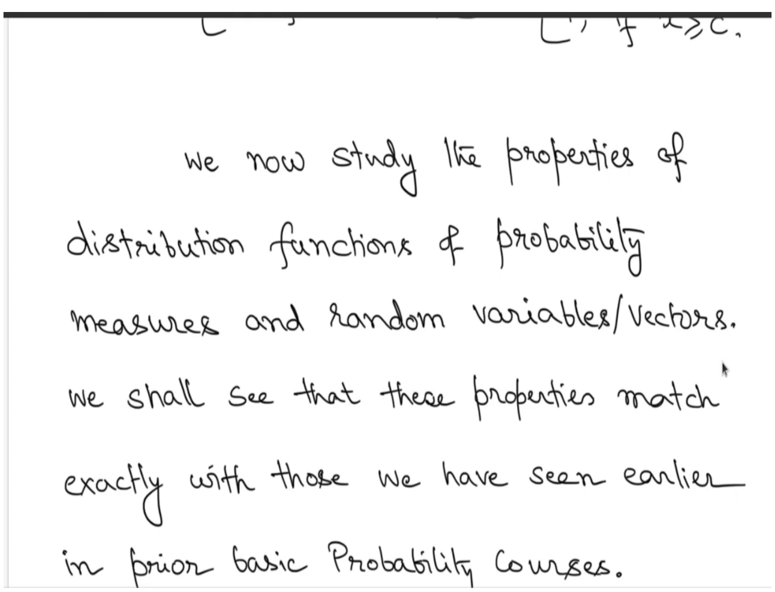
So, what do you do? You follow the definition. So, the distribution function of X evaluated at any x in the real line, so that is nothing but $\mathbb{P} \circ X^{-1}$ of the interval. And therefore, here in this case, $\mathbb{P} \circ X^{-1}$ is the Dirac mass, Dirac measure supported at the point c.

And what does Dirac do? It checks whether the point c is in this interval or not. Now that means that if the point c is not in the interval $(-\infty, x]$, if it is not there you will assign the value 0. If the point is there in the interval, you will assign the value 1. So, this is simply following the definition of the Dirac mass.

But then you can rewrite this condition, c belonging to these sets or not in terms of certain simple inequalities. What are these inequalities? $c \notin (-\infty, x]$ if the point $x < c$. And it will be exactly equal to 1, that is, the point c is exactly in the set if the x is exceeding the value c, or equal to c.

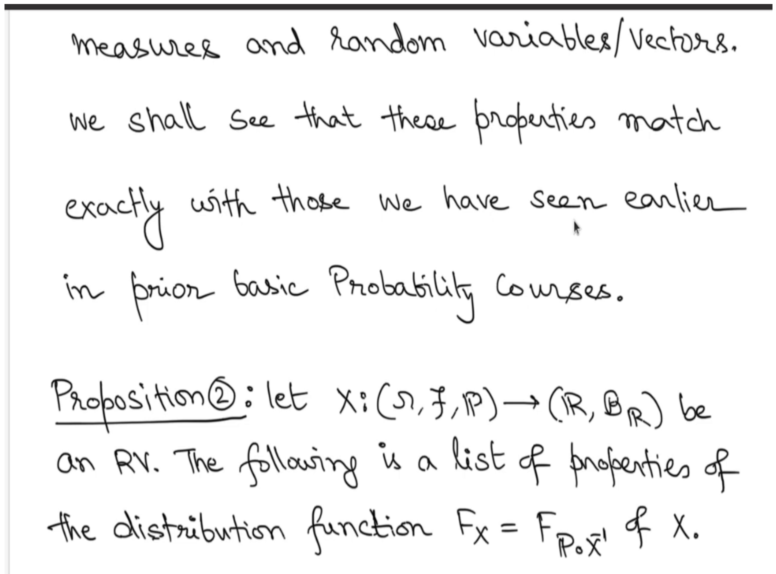
And that is exactly the distribution function of the degenerate random variable taking the constant value c. So, you get back your familiar values for the distribution functions. But this is through the law now. All these computations are going via the law.

(Refer Slide Time: 11:26)



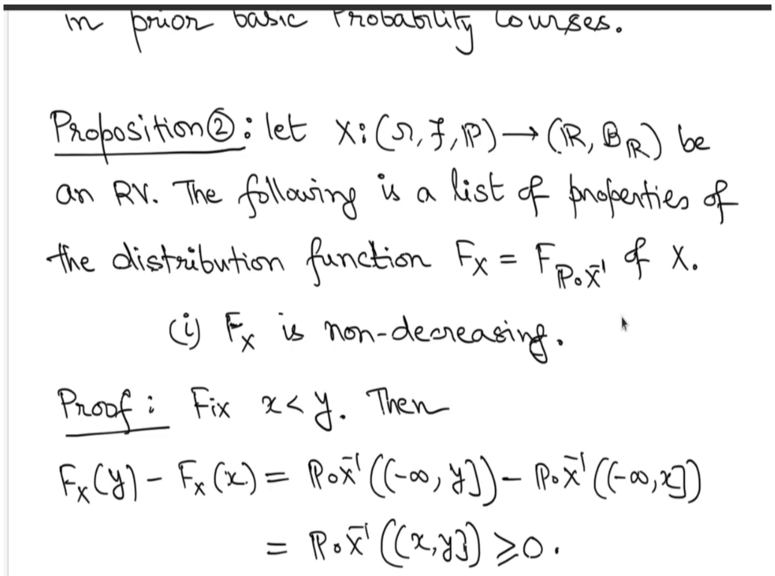
And as we shall see many of the properties of the random variables or random vectors are getting captured by this law. So, we are now going to study the properties of distribution functions corresponding to probability measures or corresponding to random variables or vectors.

(Refer Slide Time: 11:45)



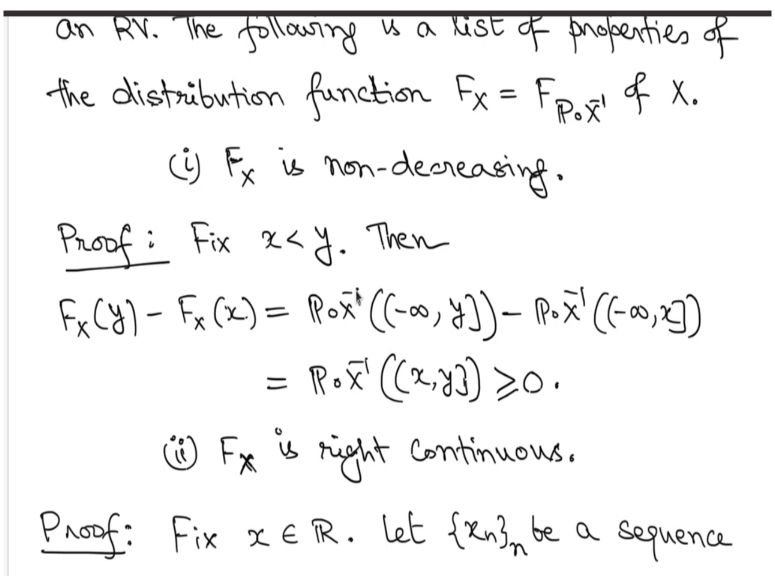
And we are going to see that all these properties for this random variables or vectors that you have already seen in your basic probability course are reappearing here and that is all, I mean, you are just matching up with whatever you have seen earlier. But important restriction, so far we have not talked about discrete or continuous random variables. We are doing everything in general. So, that is a very big advantage of using this measure theoretical approach.

(Refer Slide Time: 12:14)



So again, so let us start with these notations. So, in proposition 2, we are looking at a specific random variable, and we are listing certain properties of the distribution function, the well-known properties.

(Refer Slide Time: 12:25)



So, what is this? The first property says that the distribution function is non-decreasing. So, how do you go about proving this? So, what you have to do? You have to evaluate it at two points, x and y. So, choose $x < y$, and then you evaluate the value then try to check whether the function values you can compare.

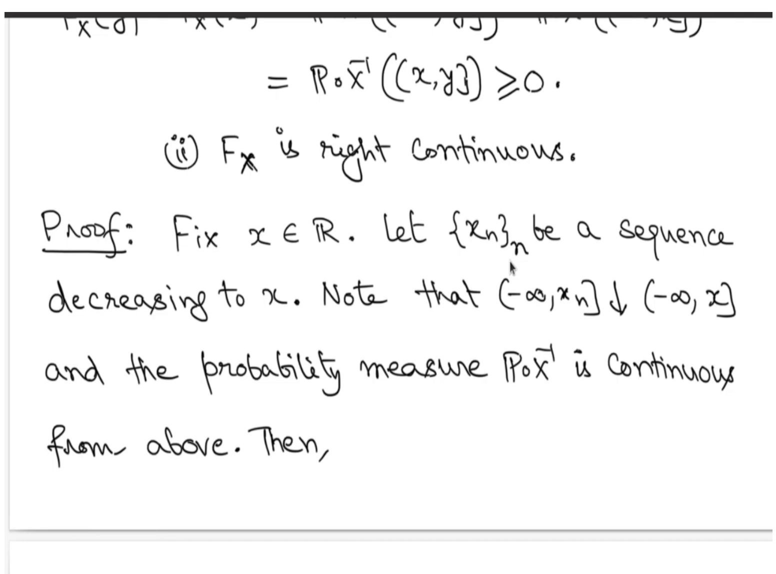
So, look at this difference of these two quantities. So, you evaluated at y and subtract out the, the value at x. But then, put in the definition it is $\mathbb{P} \circ X^{-1}$ of this interval minus $\mathbb{P} \circ X^{-1}$ of this interval.

Now, use the finite additivity of the probability measure $\mathbb{P} \circ X^{-1}$ and write the interval $(-\infty, y]$ as $(-\infty, x] \cup (x, y]$, and one of the terms will cancel off with this and that will just leave you $\mathbb{P} \circ X^{-1}(x, y]$.

Now, this interval whatever it is, it is a set, it is a Borel set on R and its size under the probability measure $\mathbb{P} \circ X^{-1}$ cannot be negative, so always the size is non negative. This is simply following the basic definition that are measure associates non negative values.

So therefore, you get that $F_X(y) \geq F_X(x)$.And this property follows quite simply from the basic definition that $\mathbb{P} \circ X^{-1}$ is a measure, probability measure and in particular it is a measure. So, therefore the size of these sets will be non-negative.

(Refer Slide Time: 14:12)

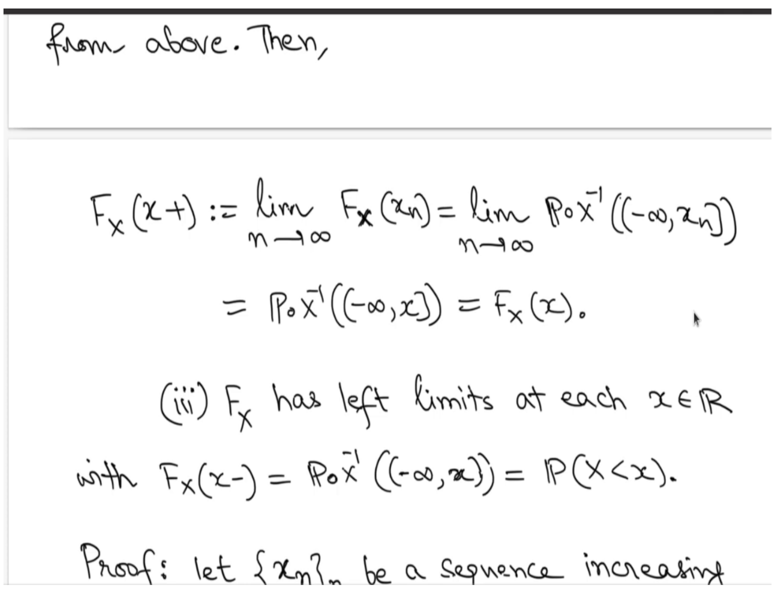


So, let us try the next property, that F_X is actually right continuous. So, how do you go over proving this? So, what you have to do? You have to verify the right continuity at all the points on the real line. So, fix any arbitrary point x. You want to verify that it is right continuous.

Now, the way to check right continuity will be to approximate the point x from the right side. So, choose such a sequence after coming to the point x from the right. So, it is basically, you can choose it as a decreasing sequence of real numbers, decreasing to the point x.

But then, if you look at such intervals $(-\infty, x_n]$, they decrease and decrease to the interval $(-\infty, x]$. This is a simple observation. And then you use the fact that the probability measure $\mathbb{P} \circ X^{-1}$ is continuous from above. So, you can go through the probability measures of these approximating intervals and after you take the limit you will get back the probability measure of this resultant interval. So, that is continuity from above

(Refer Slide Time: 15:16)

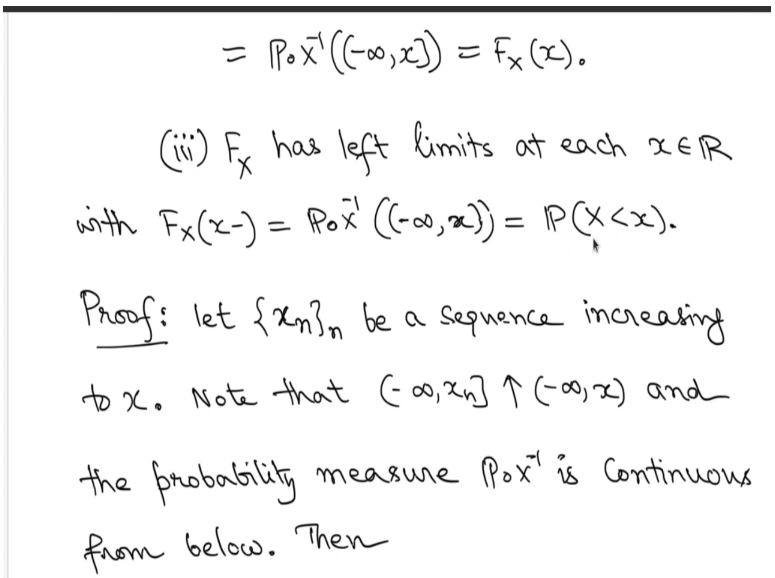


And therefore, what you end up having is the right limit at the point x. So, let us compute that. So, the right limit at the x is nothing but the limit of these points x_n. So, again, all you are going to show is that no matter what sequence $\{x_n\}$ you choose from the right approximating the point x, you get the same value. So, it does not matter. So, here this limit is now independent of the choice of the x ns

So, let us do the computations and we will immediately see that it is a really independent of the choice of x_n's. So, now put in the definition, it is $\mathbb{P} \circ X^{-1}$ of such sets as we mentioned, but then use the continuity from above and that is nothing but $\mathbb{P} \circ X^{-1}$ of this. So, it is really independent of the choice of the sequence $\{x_n\}$.

And now this quantity, whatever it is, as per definition, it is nothing but the distribution function evaluated at the point x. So, therefore the right limit is matching with the function value and therefore the function is right continuous at this arbitrary point x. So, since x is arbitrary, the point x is arbitrary, you get the right continuity at all the points.

(Refer Slide Time: 16:22)



But now, let us look at the left limit versions of this. And it will turn out that there are possibly certain differences. It need not be left continuous. So, what happens here? So, you again look at the distribution function of X, and you try to evaluate these left limits at each point on the real line.

You will get the left limit but the left limit is exactly the length of such an interval, $(-\infty, x)$ and which can be written as the $\mathbb{P}(X < x)$.

(Refer Slide Time: 17:07)

with $F_X(x-) = \mathbb{P} \circ X^{-1}((-\infty, x)) = \mathbb{P}(X < x)$.

Proof: let $\{x_n\}_n$ be a sequence increasing to x. Note that $(-\infty, x_n] \uparrow (-\infty, x)$ and the probability measure $\mathbb{P} \circ X^{-1}$ is continuous from below. Then

$$F_X(x-) = \lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} \mathbb{P} \circ X^{-1}((-\infty, x_n])$$

$$= \mathbb{P} \circ X^{-1}((-\infty, x)) = \mathbb{P}(X < x).$$

So, how do you show this? So, to show this we follow the similar procedure to the argument in right continuity. So, what we did there was approximating it from the right. here we will be approximating it from the left.

So, now $\{x_n\}$ will be a sequence which will be coming from the left hand side of x increasing to x. So, look at this. So, now these intervals, $(-\infty, x_n]$, they will increase and increase to the interval $(-\infty, x)$. So, this is a very important distinction from the case of right continuity.

So, here you do not get the boundary point x. So, these intervals, $(-\infty, x_n]$, the union of those they will actually turn out to be $(-\infty, x)$. Great.

(Refer Slide Time: 17:50)

to x. Note that $(-\infty, x_n] \uparrow (-\infty, x)$ and the probability measure $\mathbb{P} \circ X^{-1}$ is continuous from below. Then

$$F_X(x-) = \lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} \mathbb{P} \circ X^{-1}((-\infty, x_n])$$

$$= \mathbb{P} \circ X^{-1}((-\infty, x)) = \mathbb{P}(X < x).$$

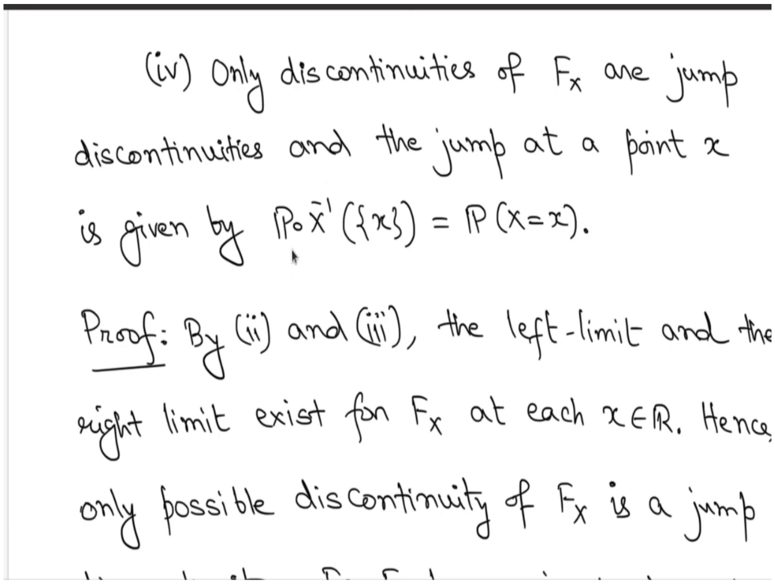
(iv) Only discontinuities of F_X are jump discontinuities and the jump at a point x

And then what happens? You look at the corresponding sizes and here you apply continuity from below because the sets are now increasing. So, then the left limit evaluated at the point x is again the limit taken over such sequences and as you will see it will again be independent of the choice of the sequence.

So, you put in the definition. So, it is nothing but the size of the set $(-\infty, x_n]$ evaluated in terms of the probability measure. So, it is the size of that. Great. But you use the continuity from below, it will exactly be the, the size of the limit set. So, it is $(-\infty, x)$. And you rewrite it, it is exactly $\mathbb{P}(X < x)$.

So, the left limit is $\mathbb{P}(X < x)$. And this is again expected from your usual understanding, or usual results that you have seen in your basic probability theory.

(Refer Slide Time: 18:46)



So now, we computed the right limit, we computed the left limit. You can now talk about the discontinuities that can happen for such functions. So, you have a right continuous non decreasing function, you have shown that it is right continuous, it has left limits but there might be some issues with the discontinuity at these points.

So, the claim is this that the only discontinuities of F_X, the distribution function of X are jump discontinuities. Again, you get back the familiar properties of the distribution function. And now we can also say something about the jump size.

Again, you get back the familiar property that the jump size at any point x is exactly the probability of the event $X = x$, which is nothing but now the size of the singleton set $\{x\}$ in terms of the probability measure $\mathbb{P} \circ X^{-1}$. So, how do you prove this?

(Refer Slide Time: 19:38)

is given by $\mathbb{P}\circ X^{-1}(\{x\}) = \mathbb{P}(X=x)$.

<u>Proof:</u> By (ii) and (iii), the left-limit and the right limit exist for F_X at each $x \in \mathbb{R}$. Hence only possible discontinuity of F_X is a jump discontinuity. If F_X has a jump at x, then the size of the jump is

$$F_X(x+) - F_X(x-) = \mathbb{P}\circ X^{-1}((-\infty,x]) - \mathbb{P}\circ X^{-1}((-\infty,x))$$

So, to talk about the discontinuities of this non decreasing right continuous function, you just look at the left limits and the right limits. Since the left limits and the right limits exist for any arbitrary point, you do not get any other type of discontinuities. Only possible discontinuity is a jump discontinuity.

So, where the jump discontinuity simply means that the left limit and the right limit do not agree. So, the jump discontinuity can occur only when the left limit and the right limit at a point do not agree. So, let us try to see that.

(Refer Slide Time: 20:10)

only possible discontinuity of F_X is a jump discontinuity. If F_X has a jump at x, then the size of the jump is

$$F_X(x+) - F_X(x-) = \mathbb{P}\circ X^{-1}((-\infty,x]) - \mathbb{P}\circ X^{-1}((-\infty,x))$$
$$= \mathbb{P}\circ X^{-1}(\{x\}) = \mathbb{P}(X=x).$$

(v) $F_X(\infty) := \lim F_X(x) = 1$

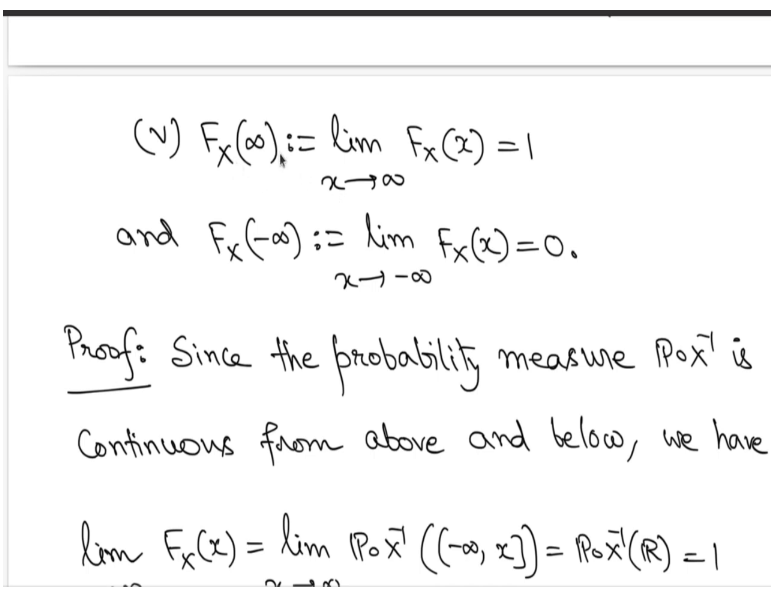
So, if F_X has a jump that means that you get a difference of values here. So, the left limit and the right limit might not match. So, look at this difference now. So, as for the calculations above it is nothing but the size of this interval $(-\infty, x]$, and then you are subtracting out the size of the interval $(-\infty, x)$. So, this is again all the expressions that we have derived earlier.

But then you again use the finite additivity of the measure $\mathbb{P} \circ X^{-1}$ and write the interval $(-\infty, x]$ as the disjoint union of the open interval $(-\infty, x)$ and the singleton set $\{x\}$.

So therefore, you will exactly end up with the size of the singleton set $\{x\}$ under the probability measure $\mathbb{P} \circ X^{-1}$ and that is nothing but $\mathbb{P}(X = x)$.

So, whenever there is a mass situated at the value x, whenever X takes the value x with some appropriate probability, positive probability then you get a jump. So, that is exactly whatever you have seen in your basic probability theory.

(Refer Slide Time: 21:17)



But then, you can also consider limits at ∞ and $-\infty$. Again, this will follow from the basic properties of the probability measure $\mathbb{P} \circ X^{-1}$ and this will match with the properties expected from your basic probability theory. So, what happens here? So, look at the limit at ∞.

(Refer Slide Time: 21:37)

$$F_X(\infty) := \lim_{x \to \infty} F_X(x) = 1$$

and $F_X(-\infty) := \lim_{x \to -\infty} F_X(x) = 0.$

<u>Proof:</u> Since the probability measure $\mathbb{P} \circ X^{-1}$ is Continuous from above and below, we have

$$\lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} \mathbb{P} \circ X^{-1}\left((-\infty, x]\right) = \mathbb{P} \circ X^{-1}(\mathbb{R}) = 1$$

and

So, what is this? So, you are looking at the limit of the distribution function values as point x goes to ∞. So, put in the definition. So, it is the size of the interval $(-\infty, x]$ closed under the probability measure $\mathbb{P} \circ X^{-1}$ and you are letting x go to ∞.

But these intervals, $(-\infty, x]$ increase to the whole real line. So, remember these sets are contained in the real line and if you look at the union of these, if you, if you go over sequences you will exactly get back the whole real line. And you now use continuity from below here.

So, when you are taking limit at ∞, you use continuity from below. And hence what you end up with is the probability of the whole real line under the probability measure $\mathbb{P} \circ X^{-1}$ and that is exactly equal to 1. So, you get back the limit at ∞ to be 1.

(Refer Slide Time: 22:29)

and

$$\lim_{x \to -\infty} F_X(x) = \lim_{x \to -\infty} \mathbb{P}\circ X^{-1}\left((-\infty, x]\right) = \mathbb{P}\circ X^{-1}(\phi) = 0.$$

(vi) For any $-\infty \le a < b \le \infty$,

$$F_X(b) - F_X(a) = \mathbb{P}\circ X^{-1}\left((a,b]\right) = \mathbb{P}\left(a < X \le b\right).$$

Proof: If $-\infty < a < b < \infty$, then the result is a part of statement (i). When $a = -\infty$ or $b = \infty$, the proof is left as an exercise.

And now, what is the limit at $-\infty$? You have to look at the limit of the sizes of these sets, $(-\infty, x]$. But these sets are now decreasing. If you choose a sequence of $\{x_n\}$ going to $-\infty$, you will immediately be able to show that these sets decrease and decrease to the complete intersection which is the empty set, and therefore the probability of that is exactly 0. So, $\mathbb{P} \circ X^{-1}$, the size of that, size of the set, empty set under the $\mathbb{P} \circ X^{-1}$ is 0. So, therefore the limit at $-\infty$ is 0.

(Refer Slide Time: 23:02)

(v) $F_X(\infty) := \lim_{x \to \infty} F_X(x) = 1$

and $F_X(-\infty) := \lim_{x \to -\infty} F_X(x) = 0.$

Proof: Since the probability measure $\mathbb{P}\circ X^{-1}$ is continuous from above and below, we have

$$\lim F_X(x) = \lim \mathbb{P}\circ X^{-1}\left((-\infty, x]\right) = \mathbb{P}\circ X^{-1}(\mathbb{R}) = 1$$

So, we are going to call this limit at ∞, as $F_X(\infty)$ and limit at $-\infty$ as $F_X(-\infty)$. So, remember the original distribution function for this random variable X was defined on the real line, but we have just extended the values of the function to the points ∞ and $-\infty$, and therefore you can think of the distribution function F_X as a function on the extended real line. So, this we can use later on. So, this factor we will be using later on.

(Refer Slide Time: 23:32)

$$\lim_{x\to-\infty} F_X(x) = \lim_{x\to-\infty} \mathbb{P}\circ X^{-1}((-\infty,x]) = \mathbb{P}\circ X^{-1}(\phi) = 0.$$

(vi) For any $-\infty \le a < b \le \infty$,

$$F_X(b) - F_X(a) = \mathbb{P}\circ X^{-1}((a,b]) = \mathbb{P}(a < X \le b).$$

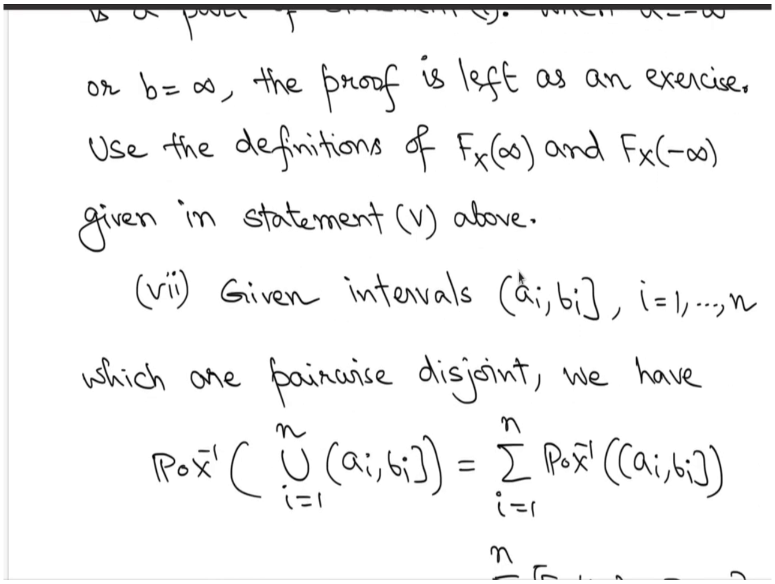
Proof: If $-\infty < a < b < \infty$, then the result is a part of statement (i). When $a=-\infty$ or $b=\infty$, the proof is left as an exercise. Use the definitions of $F_X(\infty)$ and $F_X(-\infty)$

But now, you can derive other familiar properties, like the difference of the function values for any two points, $a < b$ is exactly this quantity. Again, you follow the definitions that for a, b when they are finite, so when they are real numbers, you can put in the calculations once more, put in the definition $\mathbb{P} \circ X^{-1}(a, b]$.

So, this is simply calculating by the finite additivity of the probability measure $\mathbb{P} \circ X^{-1}$. So, you first put in the definition of F_X in terms of $\mathbb{P} \circ X^{-1}$, apply the finite additivity. So, this was discussed in statement Roman 1. So, you get back exactly this type of probability of events.

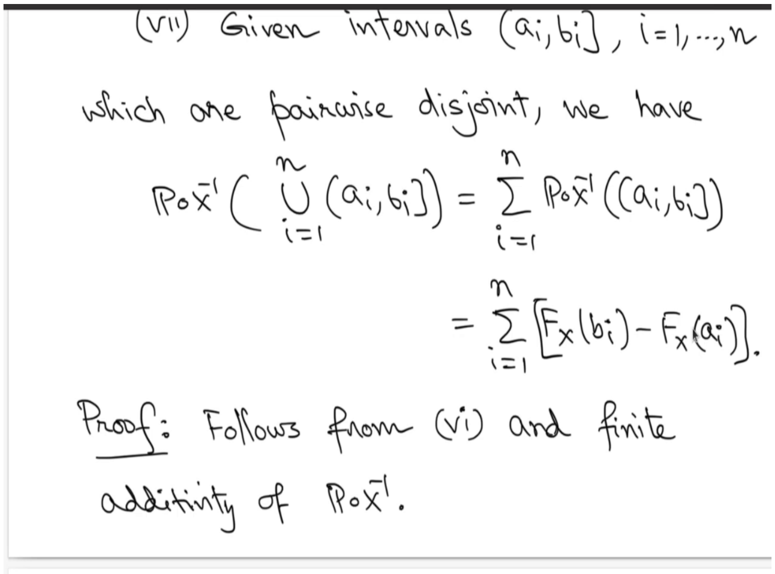
But then let us see if we can interpret these results when $a = -\infty$. Now that we have talked about the function values ∞ and at $-\infty$, we would like to make sense of this equality, if it holds in the cases when $a = -\infty$ or $b = \infty$. So now the idea is this, the left hand side makes sense because you have defined it and all you have to do is to take appropriate limits on the right hand side.

(Refer Slide Time: 24:47)



So, this part of the proof is left as an exercise. Please use the definitions of the function values at ∞ and $-\infty$ as discussed above. So, please try to see this, that these appropriate versions of this equality holds when $a = -\infty$ and $b = \infty$. Of course, you have to use the, your familiar interpretations of such intervals whenever you are dealing with the cases $a = -\infty$, $b = \infty$. So please use the appropriate interpretations as given in the Week 1 discussions.

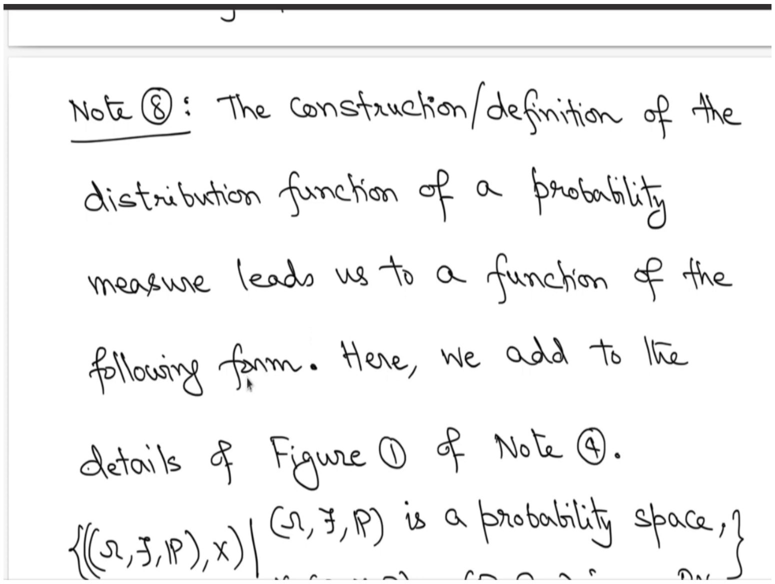
(Refer Slide Time: 25:24)



But then you can extend this idea that you have obtained for one single interval open a to closed b, two finitely many such intervals. So, if you choose such pairwise disjoint intervals you can look at the size of the union. So, this is a disjoint union, finite disjoint union of such sets and look at the size of this under the probability measure $\mathbb{P} \circ X^{-1}$.

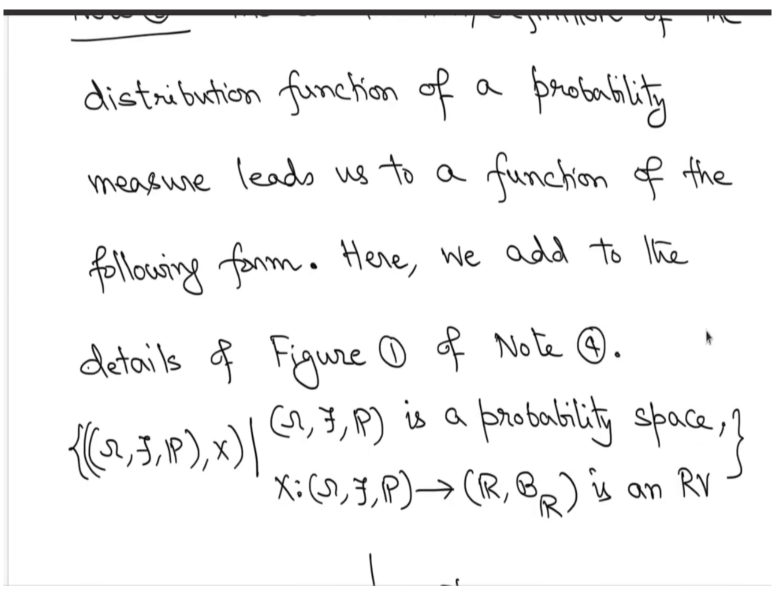
So, it is nothing but the $\mathbb{P} \circ X^{-1}$ of the individual sizes. So, it is a finite summation for using finite additivity, and now individually, these sets have the size which is given by the difference of the function values. So, for all these types of finite disjoint union of left open, right closed intervals, you can get back the probability measure by the difference of the function values. So, this is again simple application of the finite additivity.

(Refer Slide Time: 26:15)



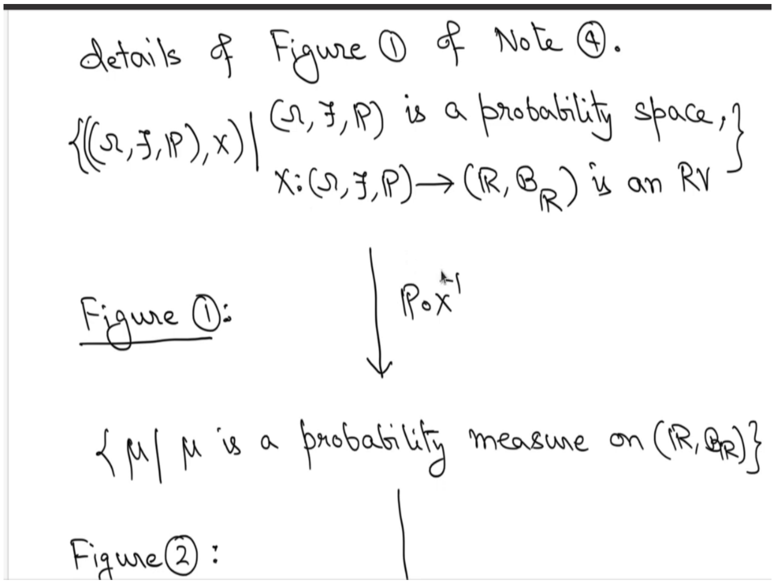
But now, let us observe this, that the construction or the definition of the distribution function as defined from the probability measure associated to the random variable or the random vector. This is leading us to a function of this following form.

(Refer Slide Time: 26:32)



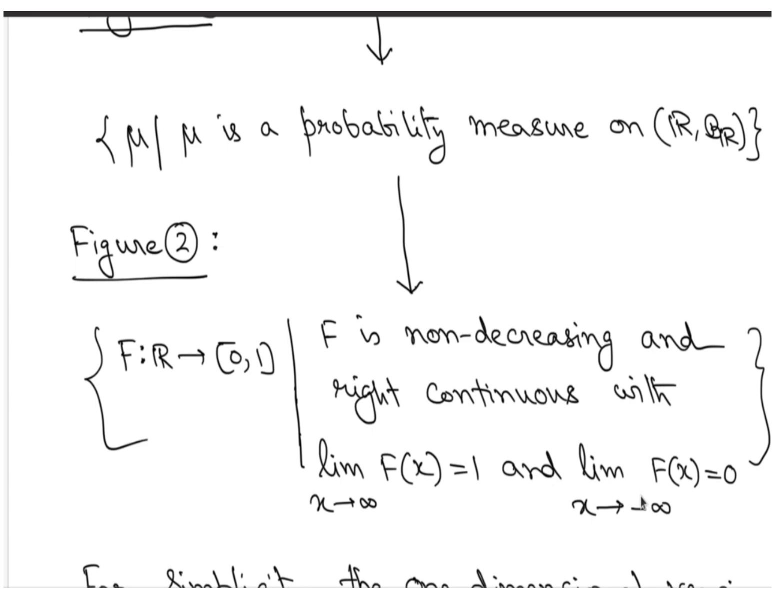
So, here we are just continuing on to add the details as discussed in Note 4 earlier.

(Refer Slide Time: 26:37)



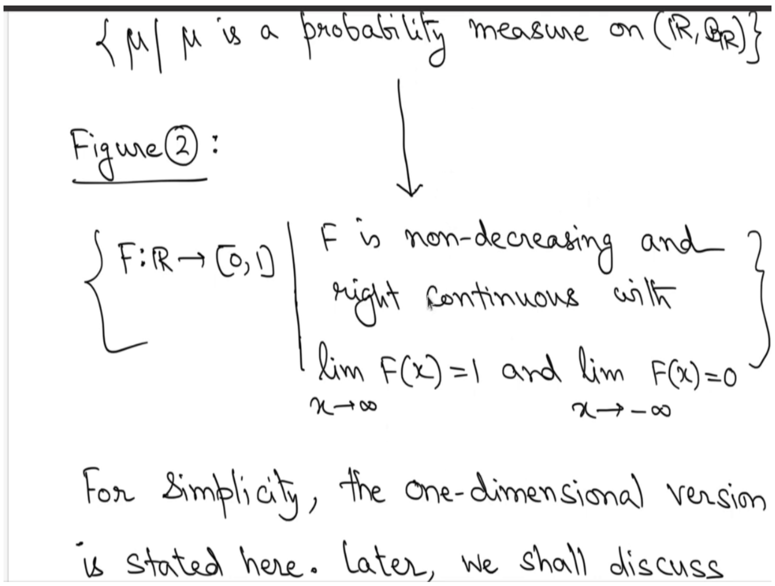
So, I will just recall what we did in Note 4. So, in Note 4 we said that given this collection of possible probability spaces and random variables on top of it, we have this collection, the first collection. So, the second collection is the collection of all probability measures and we may make this association, we created this function that sends elements from the first collection to the elements in the second collection. So, you get probability measures but, just by looking at the laws of these random variables *X* on the domain side.

(Refer Slide Time: 27:11)



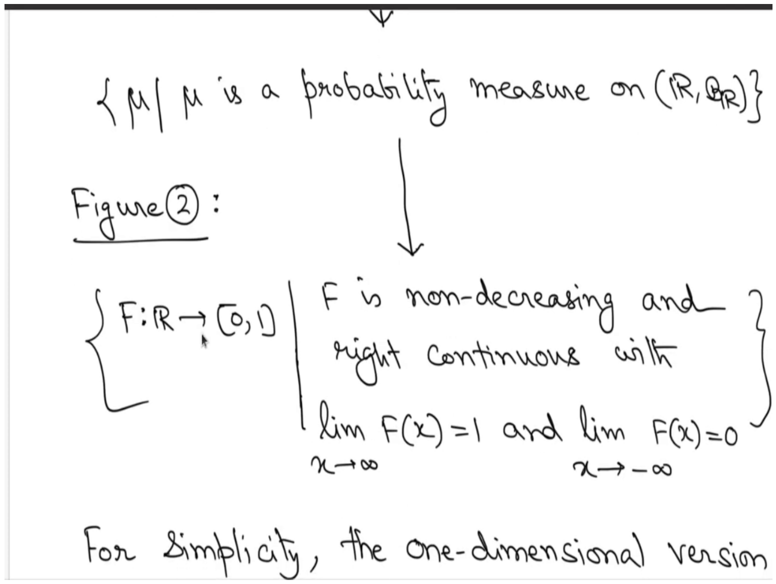
But then, for every probability measure now you have now sent it to a non decreasing right continuous function with the properties that the limit at ∞ is 1 and limit at $-\infty$ is 0. So, this is just extending the Figure 1 that was mentioned in Note 4. We have just added Figure 2, we have associated a function appropriate, function with nice, some nice properties with every probability measure on the real line.

(Refer Slide Time: 27:27)



So here, we have just stated the one-dimensional version of this. So, we have taken things for the random variables and therefore you get laws in terms of probability measures on the real line and then corresponding to that you get distribution functions with these properties, again, defined on the real line.

(Refer Slide Time: 27:54)

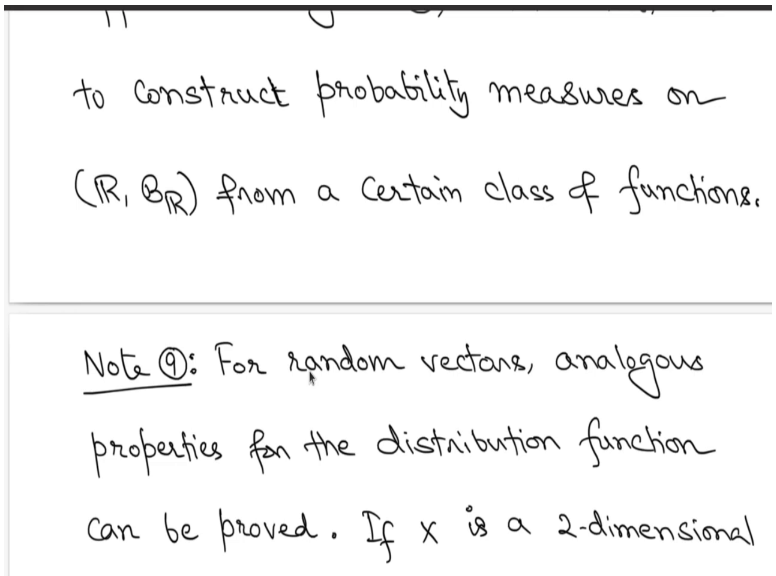


But later, we will talk about higher dimensional versions. But now importantly as mentioned in Note 4 earlier, we mentioned that we are going to construct a function going in the opposite direction of Figure 1. We are also going to construct a function which goes in the opposite direction of Figure 2.

So, what we are going to do is to given functions of this type, non decreasing right continuous with limits at $+\infty, -\infty$ given as 1 and 0, given such functions we are going to construct probability measures on the real line.

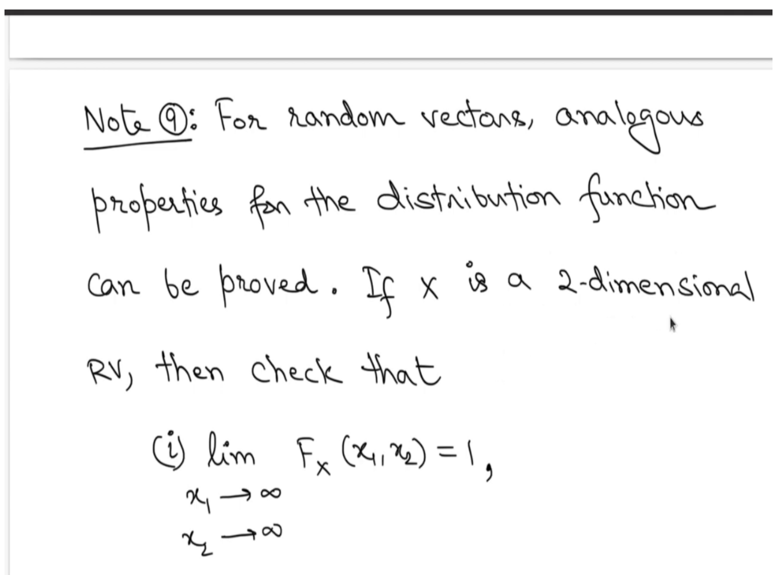
So that will basically make certain identification between collections of random variables, collections of probability measures and collections of distribution functions. So, we will make all these connections and that will be via certain functions which go in the opposite directions of Figure 1 and Figure 2. So, we will see that.

(Refer Slide Time: 28:45)



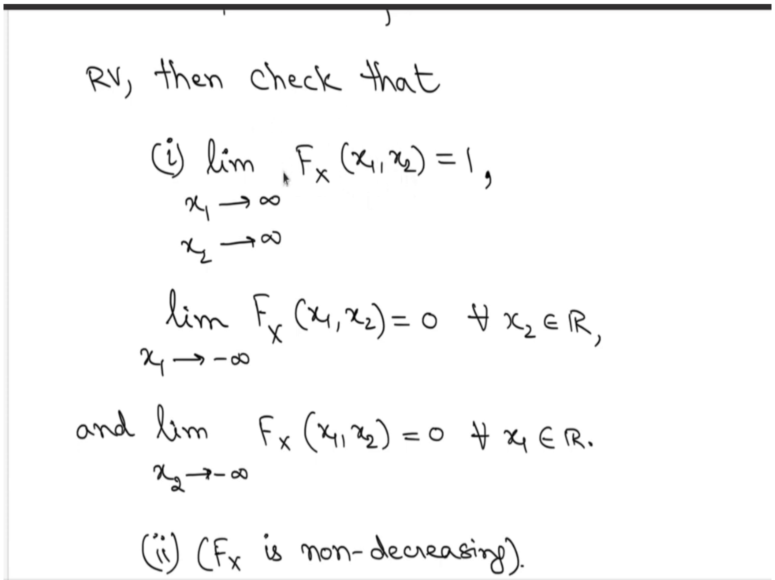
And this will actually allow us to construct probability measures on the real line from certain class of functions. So, we will see that later.

(Refer Slide Time: 28:55)



Now, we have discussed the properties of the distribution function for random variables. But then analogous properties can be proved for random vectors. So, all you have to do is to follow the definition and get back your familiar properties. Great. Just for simplicity let us look at the two-dimensional case.

(Refer Slide Time: 29:16)



So, this is a two-dimensional random variable or random vector. So, what you can check now, this I have left as exercises, what you can check now is that if you let $x_1 \to \infty$ and $x_2 \to \infty$ simultaneously, so the limit for the distribution function $\lim_{x_1, x_2 \to \infty} F_X(x_1, x_2)$ defined on $\mathbb{R}^2$, so I am saying that this will be taking the value 1 at the limit.

So, you can define the function value, extend the function value at (∞, ∞). So, please try to check this. Follow the definition, through the definition given by $\mathbb{P} \circ X^{-1}$. Here, X is a two-dimensional random vector.

(Refer Slide Time: 29:58)

$$\text{(i)} \lim_{\substack{x_1 \to \infty \\ x_2 \to \infty}} F_X(x_1, x_2) = 1,$$

$$\lim_{x_1 \to -\infty} F_X(x_1, x_2) = 0 \quad \forall x_2 \in \mathbb{R},$$

$$\text{and} \lim_{x_2 \to -\infty} F_X(x_1, x_2) = 0 \quad \forall x_1 \in \mathbb{R}.$$

(ii) (F_X is non-decreasing).

For $x_1 < x_2$, $y_1 < y_2$, observe that

$$\mathbb{P} \circ X^{-1}((x_1, x_2] \times (y_1, y_2]) \geq 0$$

Similarly, you can talk about limits by fixing one of the coordinates and letting the other one go down till $-\infty$. So, if you do that, again, by the similar calculations as done for the random variables case, the one-dimensional case, you can again show that both these limits will be 0.

So, if you fix one coordinate, let the other go down to $-\infty$ you will get the value 0. So, again these properties simply follow from the, continuity from below and above of the probability measure $\mathbb{P} \circ X^{-1}$. So please check this and try to write down these proofs.

(Refer Slide Time: 30:34)

(ii) (F_X is non-decreasing).

For $x_1 < x_2$, $y_1 < y_2$, observe that

$$\mathbb{P} \circ X^{-1}\left((x_1, x_2] \times (y_1, y_2]\right) \geq 0.$$

But, $\mathbb{P} \circ X^{-1}\left((x_1, x_2] \times (y_1, y_2]\right)$

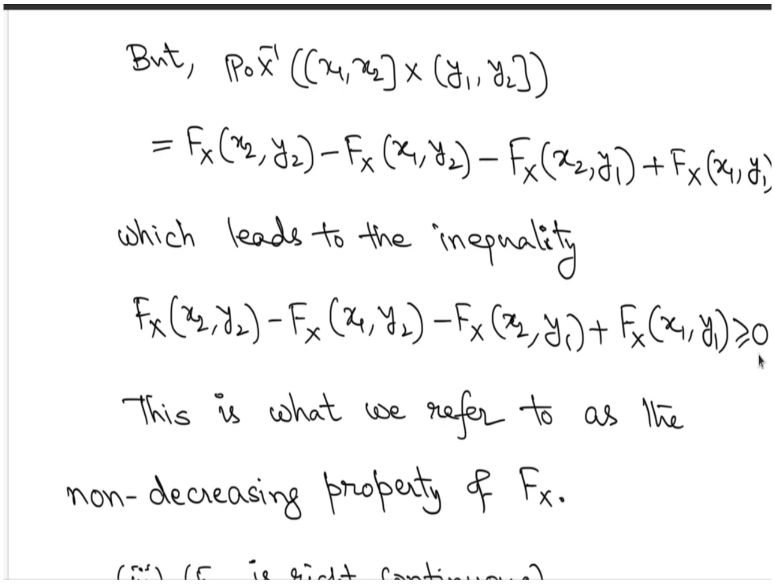
$$= F_X(x_2, y_2) - F_X(x_1, y_2) - F_X(x_2, y_1) + F_X(x_1, y_1)$$

which leads to the inequality

But then for the two-dimensional case, what is the version of the non decreasing property? So here what you can do is that you start with two values, two real numbers x_1 and x_2 and again choose another two real numbers y_1 and y_2.

So, you are going to choose these points x_1, x_2 in the first coordinate and y_1, y_2 in the second coordinate. So, look at this kind of a product of intervals. So, this is a two-dimensional set. So, the first coordinate lands up in $(x_1, x_2]$, second coordinate lands up in $(y_1, y_2]$. So, now this is a nice set in the Borel σ- field on $\mathbb{R}^2$, and if you look at the size of that, that will be non, non negative that is simply by the definition of a probability measure.

(Refer Slide Time: 31:23)



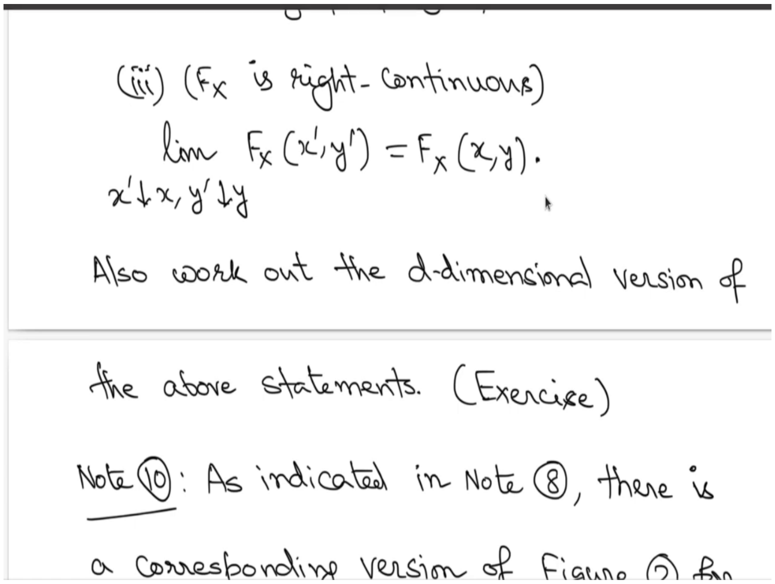
But then what you can observe now is that $\mathbb{P} \circ X^{-1}$ of this interval can be written in terms of the certain relations involving the distribution function. So, what you do is that you rewrite the, you compute the function values, the distribution values at these points. So, start with x_2, y_2, subtract out these two quantities and add up this quantity.

And what will happen is that since the probability of the product of intervals is non-negative, you get this inequality that this combination of the function values must also be non-negative. This is what we refer to as the non decreasing property of the distribution function F_X.

So, this is in the two-dimensional case, so you get back the this, non decreasing property following the properties of the corresponding probability measure, the measure associated to these sets, the product of intervals like this that rectangle sets is non-negative.

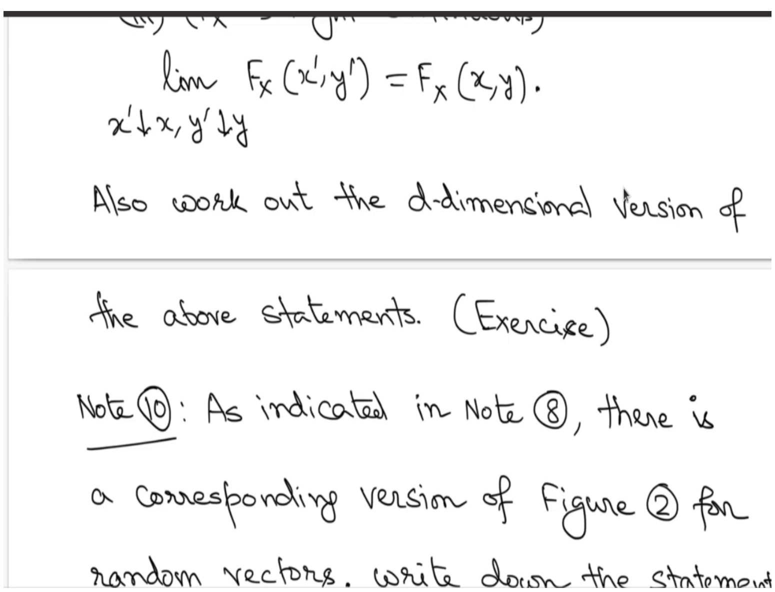
And question is, how do you prove such a relation? How do you go from probability of product of intervals to this thing? The hint is that you use the inclusion exclusion principle. So, please try to work this out.

(Refer Slide Time: 32:32)



Now, again you have this familiar property of right continuity, so all you have to do you have to let the limit for each of the coordinates, if you have, take the limits, each coordinate simultaneously you get back the actual function value. So again, this will be using the property of the corresponding probability measure $\mathbb{P} \circ X^{-1}$. All you have to use is the property continuity from above.

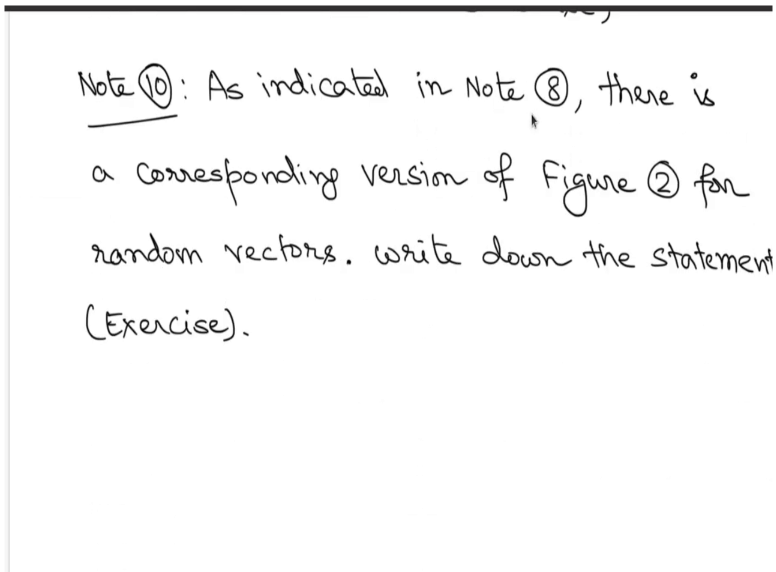
(Refer Slide Time: 32:56)

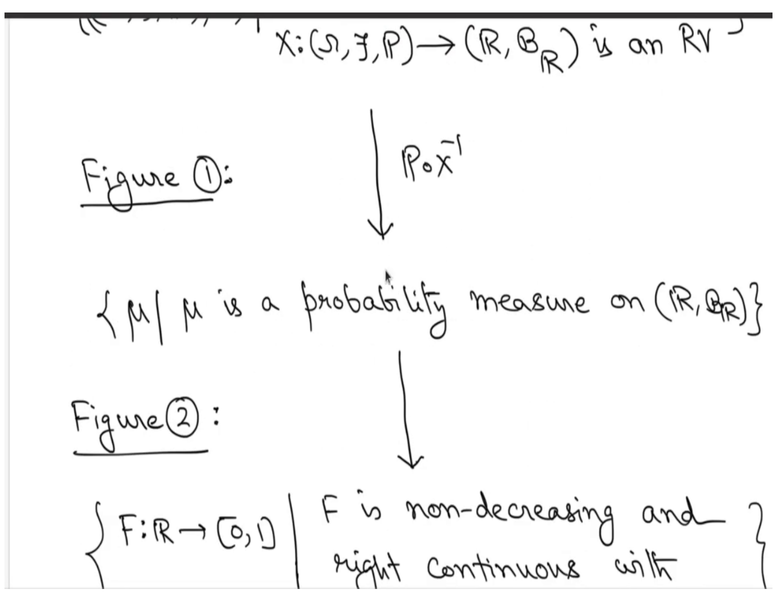


But then, you can also work out the d-dimensional versions of this. So, we have started with the one dimension, just for simplicity we have just given a brief outline or brief ideas about the two dimensional case, but now we are saying that the analogous versions hold for the d-dimensional case also.

So, please try to work out the non decreasing property, the right continuity, and other properties. So, please write it down. So, this is for the d-dimensional case, and that is left as exercises.

(Refer Slide Time: 33:28)





Now, a final comment is this, that as indicated in Note 8, so what was Note 8? So let us go back to that. So, the Note 8 was this connection between random variables to probability measures and to probability measures to the corresponding distribution function. So, this was stated for the dimension 1 case. So, we are now going to state certain things about the d-dimensional case.

So, there is a corresponding version for the d-dimensional case, for the random vectors case. So, you actually have to start with random vectors on some probability space, go to the corresponding law which is a probability measure on $\mathbb{R}^d$ together with the Borel σ-field on $\mathbb{R}^d$, and then corresponding to that probability measure you get a distribution function. You get this connection starting with random vectors, go to the probability measure which is its law, and then go to the corresponding distribution function.

So, this connection, please try to write it down for the d-dimensional case. So, in Note 8 it is written for the one-dimensional case please write down the d-dimensional version. We are going to continue the discussions about properties of the distribution function in the next lecture, specifically we are going to talk about the jumps of distribution functions. So, we stop here.