Measure Theoretic Probability 1
Professor Suprio Bhar
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Lecture 16
Law or Distribution of an RV

Welcome to this lecture. This is the first lecture of Week 4. We are now at a very important juncture of this course. So, let us first take stock of the situation, whatever we have learned in the first three weeks. And this is important because we are going to use all of that knowledge in Week 4, we are going to connect all of that in Week 4. So, let us try to recall one by one.

In Week 1, we studied σ-fields on non intercepts. Together, with these σ-fields, the non intercepts, this pair gives you measurable spaces. Now, with these ideas, what we did in Week 2, was to put a measure on top of measurable spaces. So, from Week 1 to Week 2, we went from measurable spaces to measure spaces. And special examples of measure spaces were probability spaces.

So that was a direct connection from Week 1 to Week 2. But in Week 3, again, we went back to Week 1. We started looking at measurable structures, measurable spaces and we studied measurable functions which can be defined on top of these measurable spaces.

So a priori, when you are talking about measurable functions, there was no connection about the measures. So, if you have a measurable space, you can talk about measurable functions on top of that.

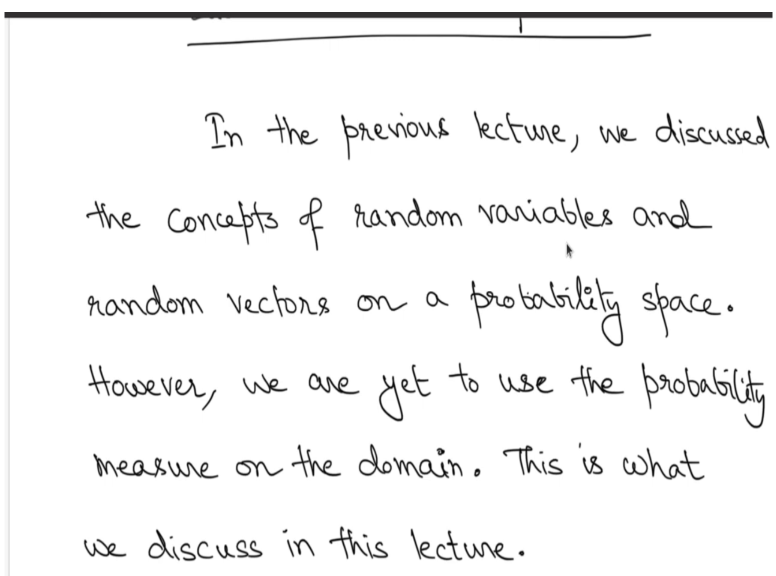
So, from Week 1, we directly went to Week 3. So, there was no direct connection between Week 2 and Week 3. But now, at the end of Week 3, we defined random variables. And you suddenly found that we are talking about Borel measurable functions on top of probability spaces. So, there was an additional probability measure given to you on the domain side.

So now, we mentioned that this has to be clarified, what is the connection of this probability measure, with respect to the measurable structure. So, this is exactly connecting to the tie up between Week 2, which involves measures and Week 3, which involves measurable functions.

So again, from Week 1 to Week 2, we went from measurable spaces to measure spaces. From Week 1 to Week 3, we went from measurable spaces to measurable functions. Now, we are going to tie up Week 2 and Week 3 by connecting measurable structures, measures spaces and measurable functions. So, recall, from the last lecture of the previous week, that we have defined random variables and random vectors as Borel measurable functions on top of probability spaces.

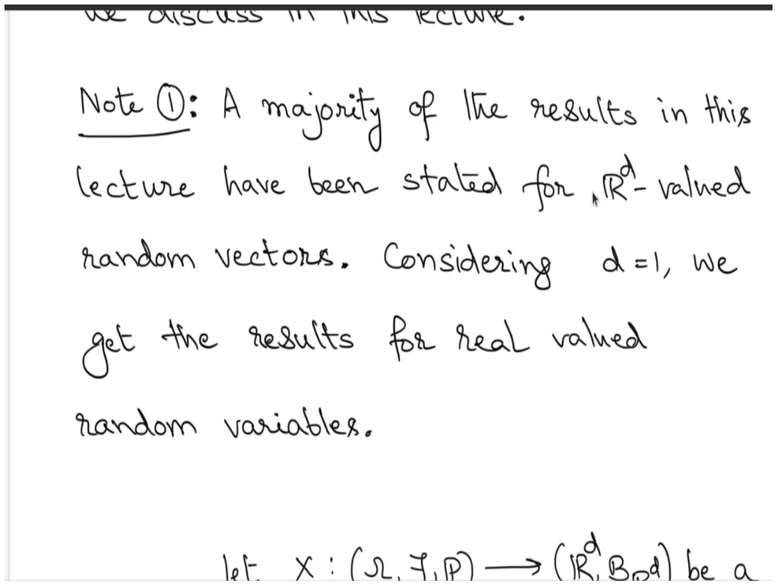
So what you are going to do is to connect these measurable structures, measurable spaces, measurable functions, together with these measures, which is available on the domain side. So, that is what we are going to talk about in this lecture. And we are going to see that this gives you a very, very important construction of a measure on the range side. So, let us go ahead and start looking at the lecture notes.

(Refer Slide Time: 03:41)



So again, just to quickly recall, in the previous lecture, we discussed the concepts of random variables and random vectors on a probability space. But again the main point is that we are yet to use the probability measure give to us. So, the probability measure that was given to us on the domain side, this we have not used. So, let us see how we are going to use this.

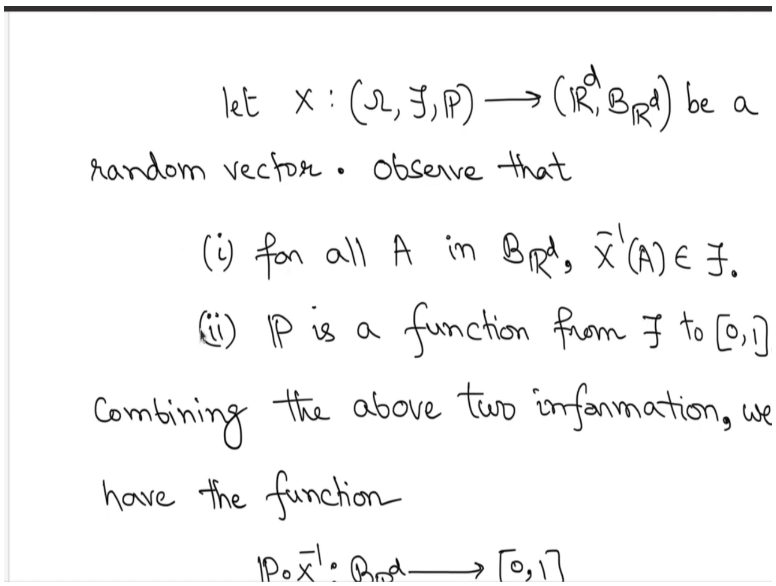
(Refer Slide Time: 04:06)



So first, one quick point is that, all of the results that we have discussed in this lecture notes are stated for general d-dimensional random vectors. But once you specify dimension $d = 1$, you can always restrict your attention to random variables. So, it is not a big issue.

When you are talking about $\mathbb{R}^d$ valued random vectors, you simply rewrite the results for $d = 1$, you will get the results for random variables. So, it is not a big issue. We are just stating things in general framework for $\mathbb{R}^d$ valued random vectors.

(Refer Slide Time: 04:44)



So let us start. The main idea is this. So, start with a random vector, taking values in $\mathbb{R}^d$. So, what is this? So, you have a probability space, $(\Omega, \mathcal{F}, \mathbb{P})$ and the range side, you have the $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. So, with respect to these σ-fields on the domain side and the range side, the random variable X is measurable. But you are just given this additional probability measure.

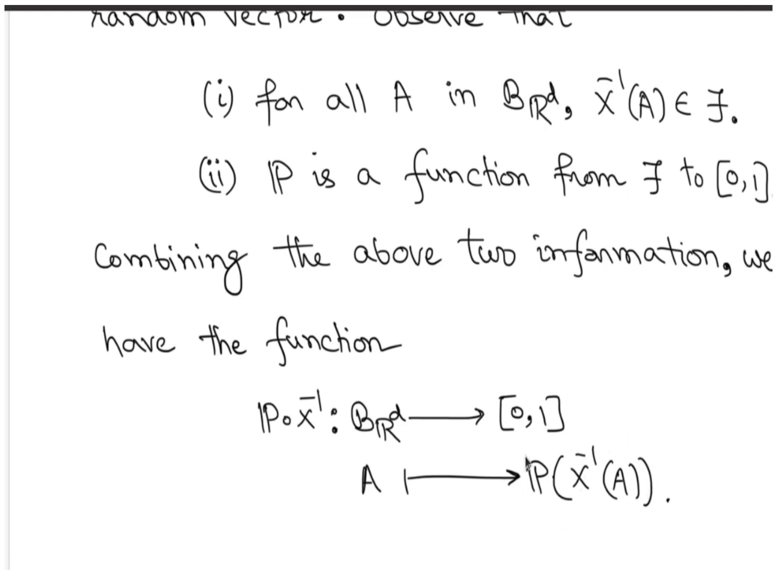
Now, let us put together all the information that we have. So, from the measurable structure of X, that function, you know that for all the Borel sets, that is on the range side, for all the Borel sets A, the pre image lies in the domain side σ-field.

So, this is simply pulling back or looking at pre images of sets, Borel sets on the range side. If you pull them back, they are on the domain side σ-field $\mathcal{F}$. So, that is something following from the measurable structure of the measurable function X. So, that is the first piece of information.

So, the second piece of information that you are going to use is the fact that the probability measure given to you is a function from the σ-field $\mathcal{F}$ and it associates values between 0s and 1s. So, for any arbitrary set in your σ-field on the domain side, the, you are going to associate values between 0 and 1. So, that is what the probability measure does, with appropriate additional properties like countable additivity and so on.

But at the end of the day, for any arbitrary set on the domain side, as long as that set is in your σ-field, $\mathcal{F}$, you can talk about the probability of that set. And that is what we are going to use now.

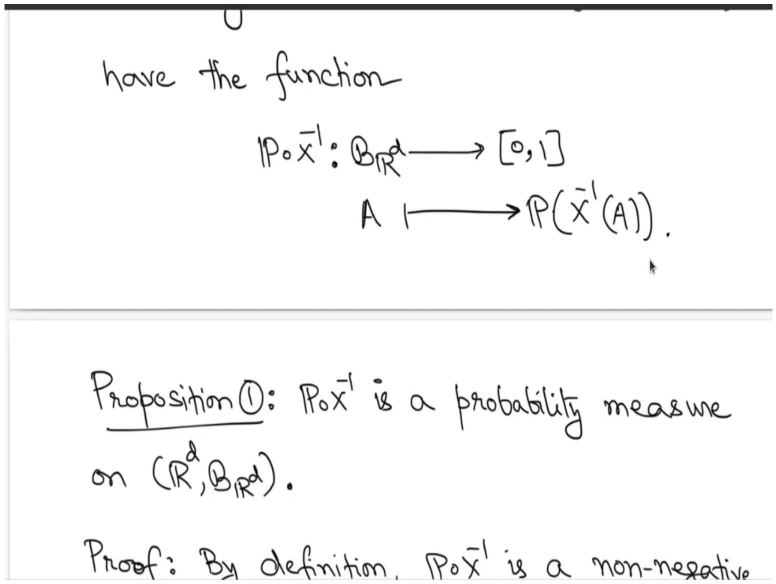
(Refer Slide Time: 06:26)



So, put these two pieces of information together, and look at this composition operation in some sense. What we are looking at is this $\mathbb{P} \circ X^{-1}$ operation. What does it do? It takes a Borel set on the range side, pulls it back, to the domain side.

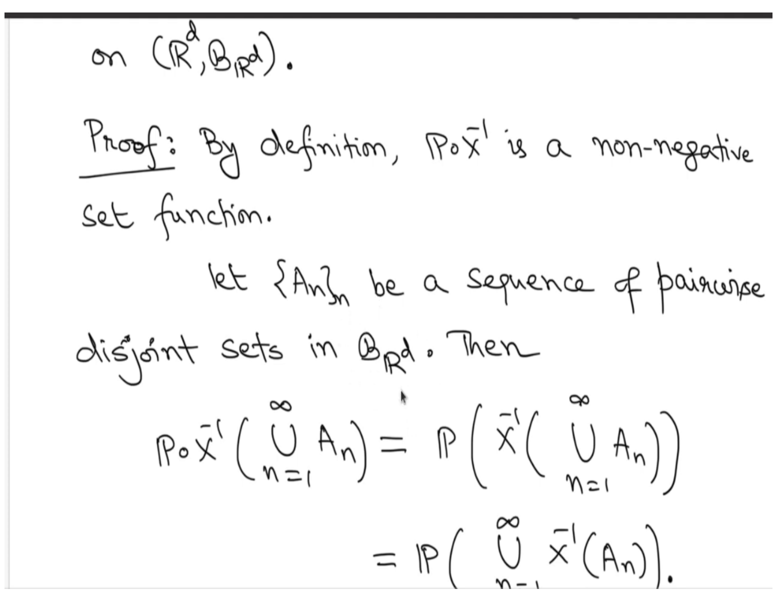
So, you get $X^{-1}(A)$, which is a set on the domain side σ-field, $\mathcal{F}$. You can talk about the probability measure of that. So, that is the value. That is the value between 0 and 1. So, this function is well defined. So, there is no issue in defining this function.

(Refer Slide Time: 06:53)



But we are going to say that this function is very, very important and this gives you, in fact, a probability measure on the range side. So, here again, for any Borel set A on the range side, you are associating a value which is given by probability of $X^{-1}(A)$. So, you are looking at probability of the pre image of the set A under X. And I claim that this is a probability measure that is on the range side.

(Refer Slide Time: 07:26)



So how do you go about proving this? So, you have already mentioned that $\mathbb{P} \circ X^{-1}$, that is well defined, so that is not a problem and it also takes non negative values. In fact, we have mentioned specifically that it takes values between 0 and 1. So, it is non negative.

So, you have a non negative set function defined on $\mathbb{R}^d$ with $\mathcal{B}_{\mathbb{R}^d}$. So, that measurable space on top of that, you have this important set function $\mathbb{P} \circ X^{-1}$.

But then, you want to claim that this is a probability measure. So, there are two steps in proving this thing. The first step, you are supposed to show that $\mathbb{P} \circ X^{-1}$ is a measure. And to do that, you have to first verify countable additivity.

So, the second step will be looking at the $\mathbb{P} \circ X^{-1}$ of the whole set and that should turn out to be 1. So, let us first verify the countable additivity. So, to do that, what do you do? You take a sequence of pairwise disjoint sets in $\mathcal{B}_{\mathbb{R}^d}$. So, you are going to verify the countable additivity here.

(Refer Slide Time: 08:22)

Let $\{A_n\}_n$ be a sequence of pairwise disjoint sets in $\mathcal{B}_{\mathbb{R}^d}$. Then

$$\mathbb{P} \circ X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right)$$

$$= \mathbb{P}\left(\bigcup_{n=1}^{\infty} X^{-1}(A_n)\right).$$

Since A_n's are pairwise disjoint, so are the sets $X^{-1}(A_n)$'s. Hence, continuing from above, So now, look at the countable union $\bigcup_{n=1}^{\infty} A_n$. So, now, look at $\mathbb{P} \circ X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)$. Now, putting the definition. So, for any set, you define it this way. So, it is $\mathbb{P} \circ X^{-1}$. So, it is defined this way that you are looking at $\mathbb{P}\left(X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right)$.

So, this is for any arbitrary set. So, therefore, it is also true for the countable union here. But then, here you are going to use this very, very important property about the pre images, that the union simply comes out. So, this was mentioned in our exercise.

So, it is simply the union of the individual pre images. So, that is a very important information that $X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)$ is nothing but the $\left(\bigcup_{n=1}^{\infty} X^{-1}(A_n)\right)$. So, that is what you write inside.

But then, you also make this important observation. As long as these A_n's are pairwise disjoint, the pre images are also pair wise disjoint. So, this is a very, very important observation. That you are starting off with pairwise disjoint sets A_n, and you are looking at this countable union, but you have now looked at this $\mathbb{P} \circ X^{-1}$ of this countable union and you have written it in terms of the probability $\mathbb{P}$ of this union of the sets which are now pairwise disjoint. So, the pre images are not pairwise disjoint. What you are going to use now is the countable additivity of the probability measure $\mathbb{P}$. So, that is given to you.

(Refer Slide Time: 09:46)

$$\mathbb{P}\circ X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right)\right)$$

$$= \mathbb{P}\left(\bigcup_{n=1}^{\infty} X^{-1}(A_n)\right).$$

Since A_n's are pairwise disjoint, so are the sets $X^{-1}(A_n)$'s. Hence, continuing from above,

$$\mathbb{P}\circ X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(X^{-1}(A_n)\right)$$

$$= \sum_{n=1}^{\infty} \mathbb{P}\circ X^{-1}(A_n).$$

So, you are going to use the countable additivity of the probability measure $\mathbb{P}$, to claim that $\mathbb{P} \circ X^{-1}$ of the countable union is nothing but this $\sum_{n=1}^{\infty} \mathbb{P}\left(X^{-1}(A_n)\right)$.

So, it is simply using the fact that the pre images are pairwise disjoint. And you are using the fact that $\mathbb{P}$ is a probability measure. In particular, it is countable additive. So, that is what you do.

(Refer Slide Time: 10:10)

Since A_n's are pairwise disjoint, so are the sets $X^{-1}(A_n)$'s. Hence, continuing from above,

$$\mathbb{P}\circ X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(X^{-1}(A_n)\right)$$

$$= \sum_{n=1}^{\infty} \mathbb{P}\circ X^{-1}(A_n).$$

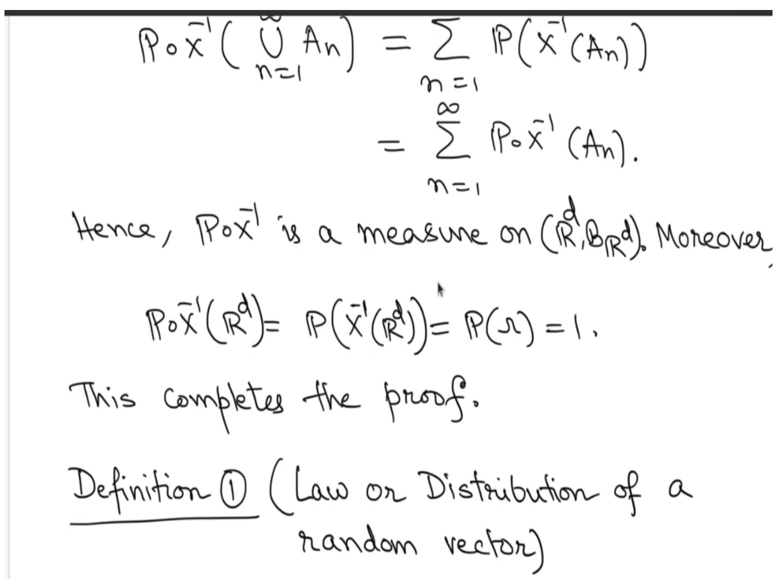
Hence, $\mathbb{P}\circ X^{-1}$ is a measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. Moreover,

$$\mathbb{P}\circ X^{-1}(\mathbb{R}^d) = \mathbb{P}\left(X^{-1}(\mathbb{R}^d)\right) = \mathbb{P}(\Omega) = 1.$$

But then you just re-write using the notation that you have introduced. So, it is nothing by $\mathbb{P} \circ X^{-1}(A_n)$. That is just by the definition. Now, compare the left hand side and the right hand side. It immediately tells you that $\mathbb{P} \circ X^{-1}$ is a set function, non negative set function defined on $\mathcal{B}_{\mathbb{R}^d}$, and it has this countable additivity. So, this you have proved for arbitrary pairwise disjoint sets A_n.

So, you have taken this sequence of pairwise disjoint sets A_n, and you have proved the countable additivity. So, therefore, $\mathbb{P} \circ X^{-1}$ turned out to be a measure on $\mathcal{B}_{\mathbb{R}^d}$.

(Refer Slide Time: 10:49)

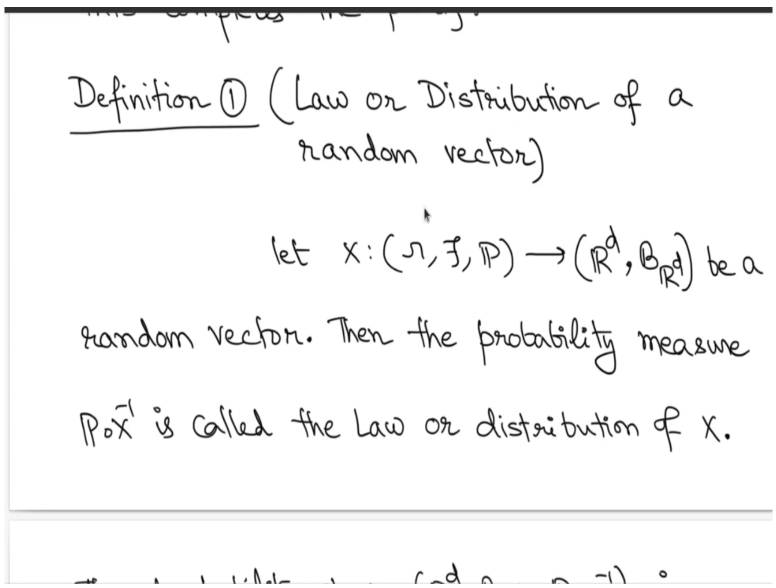


But then you want to additionally claim that this is a probability measure. To do that, you are looking at the probability of the whole set. So, $\mathbb{P} \circ X^{-1}$ of the whole set. What is the whole set? It is $\mathbb{R}^d$. Put in the definition again.

So, it is probability of the pre image of $\mathbb{P}(X^{-1}(\mathbb{R}^d))$, but then X is a mapping from Ω to $\mathbb{R}^d$. So, therefore, if you are looking at the pre image of $\mathbb{R}^d$, you are going to get the whole set anyway. So, this is just using a property of the function. X is a function, you look at the pre image of the whole range, you are just going to get the whole domain. So, these are $\mathbb{P}(\Omega)$.

But since $\mathbb{P}$ is a probability measure on this measurable space $(\Omega, \mathcal{F})$, so therefore, you get, probability of Ω is nothing but 1. And hence, $\mathbb{P} \circ X^{-1}$ must be a probability measure on the range side. So, using the measurable structure of X and the probability measure on the domain side, you have constructed now probability measure on the range side. So, this is a very, very important construction.

(Refer Slide Time: 11:48)



And now, this is what, we are putting into definition. So, we are going to define this thing. Law or distribution of a random vector. So, this is a very important step. So, take this random vector as mentioned above and look at the probability measure that we just discussed.

(Refer Slide Time: 12:03)

random vector)

let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ be a random vector. Then the probability measure $\mathbb{P} \circ X^{-1}$ is called the Law or distribution of X.

The probability space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, \mathbb{P} \circ X^{-1})$ is referred to as the induced probability space.

So, this is $\mathbb{P} \circ X^{-1}$. This is the probability measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. So, therefore, you get a probability space here. $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ together with this probability measure, you get a probability space.

What you are going to look at is this probability measure, and you are going to call it as the law of X, law of the random vector X or the distribution of the random vector X. So, be very certain about this terminology. I am going to call it the law of the random vector X or the distribution of the random vector X. This is nothing but these probability measures, that is getting defined on the range side.

(Refer Slide Time: 12:43)

The probability space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}, \mathbb{P}\circ X^{-1})$ is referred to as the induced probability space.

<u>Note ②</u>: As defined above, for any $A \in \mathcal{B}_{\mathbb{R}^d}$,

$$\mathbb{P}\circ X^{-1}(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

To simplify the notation, we suppress "ω"

So therefore, you get this probability space and what you get is this induced probability space. So, why do I call it a induced probability space? It is simply pushing forward the probability measure that was on the domain side and you are pushing it to the range side. You are pushing it forward in the forward direction. So, it is a induced probability measure. This triple that you end up with, we are going to call it as the induced probability space.

(Refer Slide Time: 13:11)

referred to as the induced probability space.

<u>Note ②</u>: As defined above, for any $A \in \mathcal{B}_{\mathbb{R}^d}$,

$$\mathbb{P}\circ X^{-1}(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

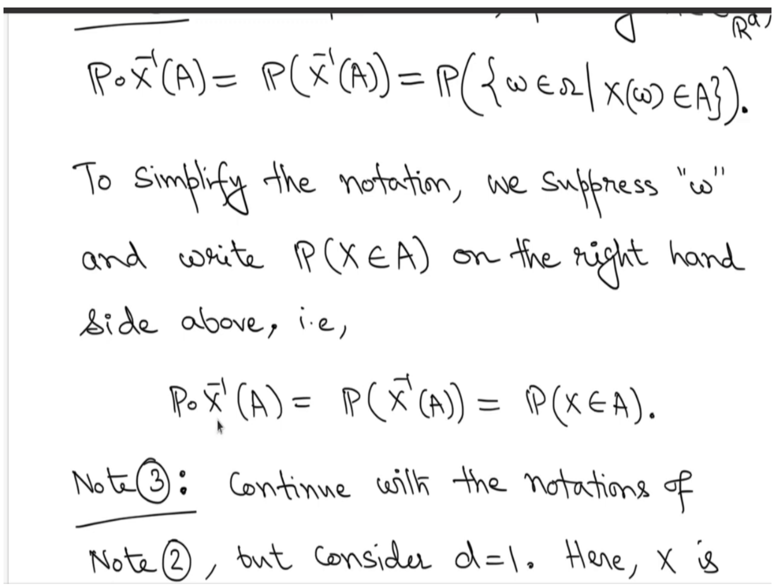
To simplify the notation, we suppress "ω"
and write $\mathbb{P}(X \in A)$ on the right hand side above, i.e,

$$\mathbb{P}\circ X^{-1}(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A).$$

But now, recall these notations that we have talked about in the previous lecture. So, look at $\mathbb{P} \circ X^{-1}(A)$. So, as per the definition, it is nothing but the probability of $X^{-1}(A)$. So, that is what exactly what it is.

But then you put in the definition of the pre image so that is nothing but this thing. So, you are looking at all points in the domain side such that $X(\omega) \in A$. But then, we introduce this concept, or we introduce this notation of suppressing the ω's.

(Refer Slide Time: 13:42)

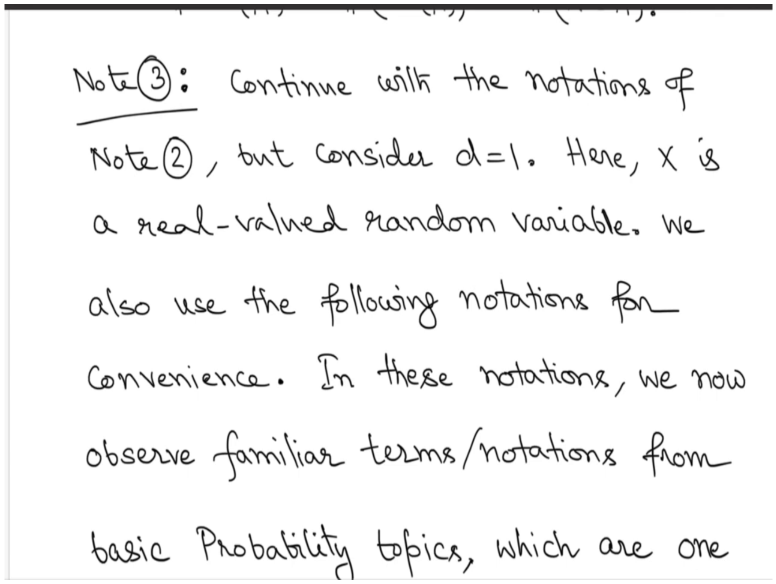


And therefore, what we chose to write was just writing it as $X \in A$. This notation. So, that is exactly what happens now. If you choose to suppress the ω's you simply get back the probability of this event.

This is now a familiar notation to you. This familiar notation, that is already available in your basic probability courses. So, therefore, we have now connected this familiar notation with, coming from the actual law or distribution of the random vector X.

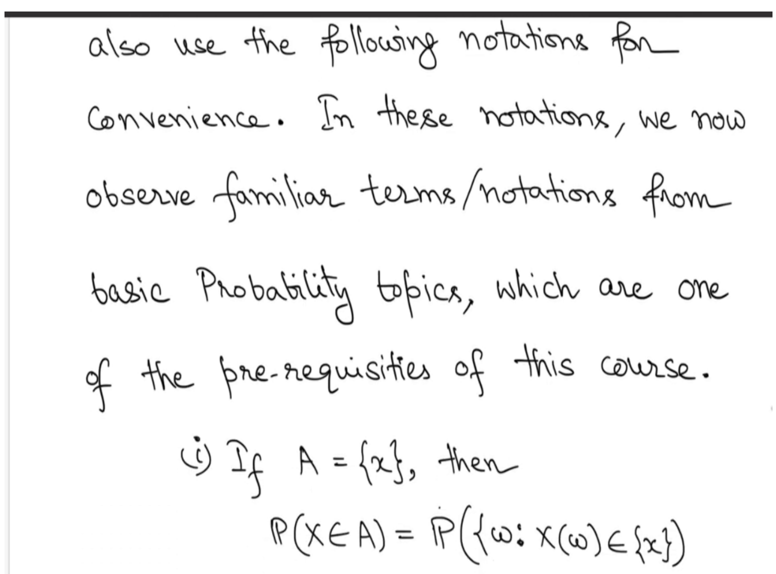
So now, this, all these notations mean the same thing, so $\mathbb{P} \circ X^{-1}(A)$ is nothing but the probability of $X^{-1}(A)$ and that is nothing but probability of $X \in A$, that (())(14:25). So, therefore, all these notations mean the same thing.

(Refer Slide Time: 14:31)



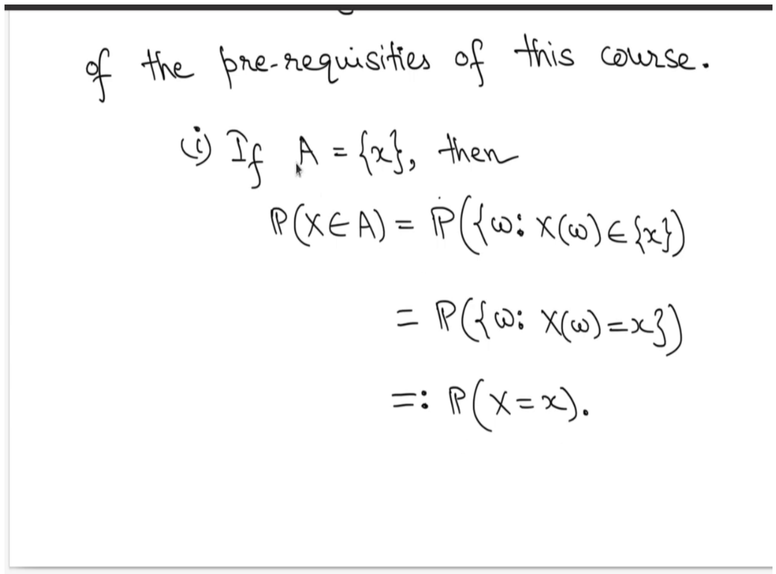
So, now just for simplicity, we continue with the notations but let us consider the case $d = 1$, one dimension case. So, you are talking about a real valued random variable now. So, so far we have, what in general random vectors, $\mathbb{R}^d$ case, but let us come back to dimension 1.

(Refer Slide Time: 14:48)



We can now use certain, following notations for convenience. And these familiar terms will now appear, start appearing. Now, we are going to see familiar terms appearing.

(Refer Slide Time: 14:59)



What are these? So, suppose you take A to be a singleton set $\{x\}$. So, it is on the range side. You are looking at $\mathbb{P}(X \in A)$. And that is, as per definition that we have just talked about is nothing but all points in the domain, all ω such that $X(\omega) \in \{x\}$.

But that nothing but, you can rewrite it using this notation that it is all points ω such that $X(\omega) = x$. Now, if you choose to remove all these ω's, you exactly get back $\mathbb{P}(X) = x$. So, again, this is a familiar notation to you. So, now, all we are saying is that this appears whenever you are looking at pre image of $\{x\}$. So, that is exactly what it is.

(Refer Slide Time: 15:49)

(ii) If $A = (-\infty, x]$, then

$$
\begin{aligned}
\mathbb{P}(X \in A) &= \mathbb{P}(\{\omega : X(\omega) \in (-\infty, x]\}) \\
&= \mathbb{P}(\{\omega : X(\omega) \leq x\}) \\
&=: \mathbb{P}(X \leq x).
\end{aligned}
$$

For $A = (-\infty, x), (x, \infty), [x, \infty)$

Similar notations shall be used.

(iii) If $A = (a, b]$, then

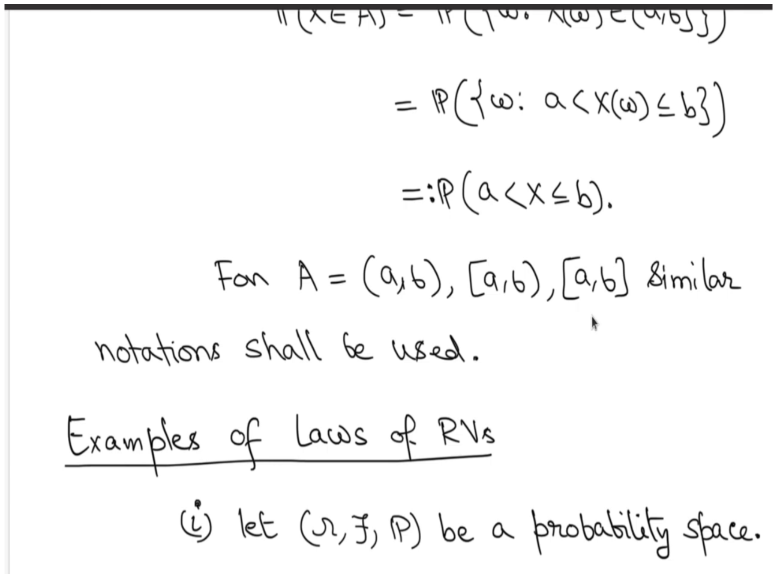
Let us look at other sets of this type. So, let us look at a interval $(-\ \infty,\ x]$. So, $(-\ \infty,\ x]$. So, you are looking at all these kind of specific types Borel sets.

What is this here? Now you are looking at probability of this event, put in the definition, it is nothing but all points in the domain such that $X(\omega)$ falls in this interval. But then rewrite it using inequalities, so you just say all points in the domain such that $X(\omega)\ \leq\ x$.

So since your values are in this interval, then all values must be $\leq x$. And then you choose to suppress the ω, you get back your familiar notation, $\mathbb{P}(X\ \leq\ x)$.

And in fact, you can choose the do the same thing for other types of intervals and you will, you can write down these notations using familiar notations involving these kind of inequalities. These events, using these inequalities, for these type of intervals. $(-\ \infty,\ x)$, $(x,\ \infty)$ and $[x,\ \infty)$. Just write it down, you will get back your familiar expressions.

(Refer Slide Time: 16:53)



But then, what happens if you are looking at this kind of an interval when these things are bounded. We are now choosing to look at (a, b), $[a, b)$ and $[a, b]$. What happens here? Put in the definition once more. So, you are saying that you are looking at all points in the domain, ω, such that $X(\omega)$ lies in this interval.

Choose to rewrite it using the inequalities and remove ω. You get back your familiar notation in terms of the familiar events. So, it is nothing but the probability of X falling between $(a, b]$. And you can choose to write down similar notations for other types of intervals like (a, b), $[a, b)$ and $[a, b]$. So, similar notations, you can find out.

All these things are just connecting to the familiar things that you have already seen in your basic probability courses. With all these notations at hand, what we are now going to do is to look at examples of laws of random vectors or random variables.

(Refer slide Time: 17:51)

(i) let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. and Fix $C \in \mathbb{R}$. Now, the function $X : \Omega \to \mathbb{R}$ defined by $X(\omega) = C, \forall \omega$, is an RV. Here, we have,

$$X^{-1}(A) = \begin{cases} \phi, & \text{if } C \notin A \\ \Omega, & \text{if } C \in A \end{cases} \quad \forall A \in \mathcal{B}_{\mathbb{R}}.$$

Then $\mathbb{P}(X^{-1}(A)) = \{ \ 0, \ \text{if } C \notin A$

So, you have defined it but we would like to see what kind of probability measure appear on the range side. When you push forward the probability measure $\mathbb{P}$, which is appearing on the domain side, you push it by the random variable X, what type of measures appear on the range side.

So we are not choosing to look at this constant function. So, remember, on any measurable space, you have this constant function and that is a nice measurable function. So, therefore, this is a example of a random variable. So, this is a constant or degenerate random variable, which just takes this single value, c.

(Refer Slide Time 18:33)

we have,

$$X^{-1}(A) = \begin{cases} \phi, & \text{if } c \notin A \\ \Omega, & \text{if } c \in A \end{cases} \quad \forall A \in \mathcal{B}_{\mathbb{R}}.$$

Then $$\mathbb{P}(X^{-1}(A)) = \begin{cases} 0, & \text{if } c \notin A \\ 1, & \text{if } c \in A \end{cases} \quad \forall A \in \mathcal{B}_{\mathbb{R}}$$

$$= \delta_c(A), \quad \forall A \in \mathcal{B}_{\mathbb{R}}.$$

Hence $\mathbb{P} \circ X^{-1} = \delta_c$.

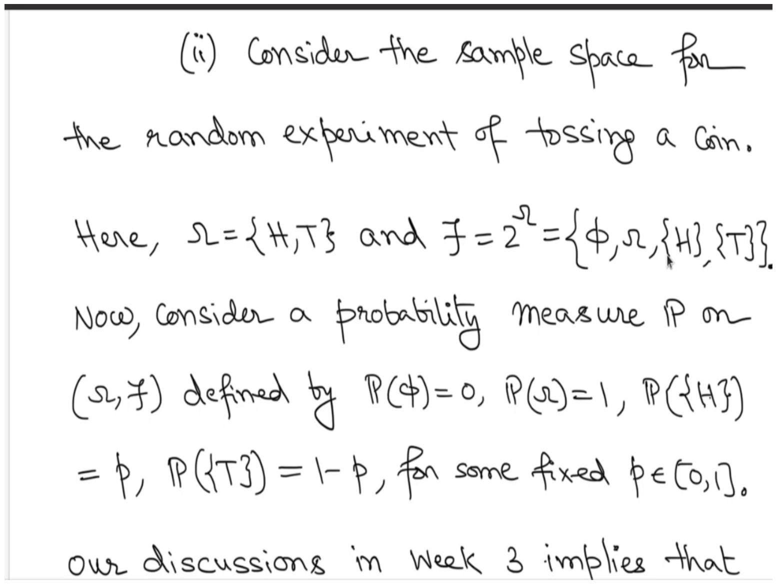
So here, we recall, that the pre-images are nothing but empty set or the whole set depending on the choices that the value c is in the set or not. So, if c is there in the set, you get back the whole set, Ω. If the c is not there in the set A, you get back the pre image as empty set.

But then, look at the probability of that and you know exactly what these values are. So, $\mathbb{P} \circ X^{-1}(A)$ is nothing but 0s or 1s depending on the situation that $c \notin A$ and $c \in A$. But that is nothing but, you recall, that this is nothing but the Dirac measure, direct supported at the point c, acting on A.

You are looking at the size of the set A according to the Dirac measure supported at the point c. What you have found out is that $\mathbb{P} \circ X^{-1}$ is nothing but the Dirac measure.

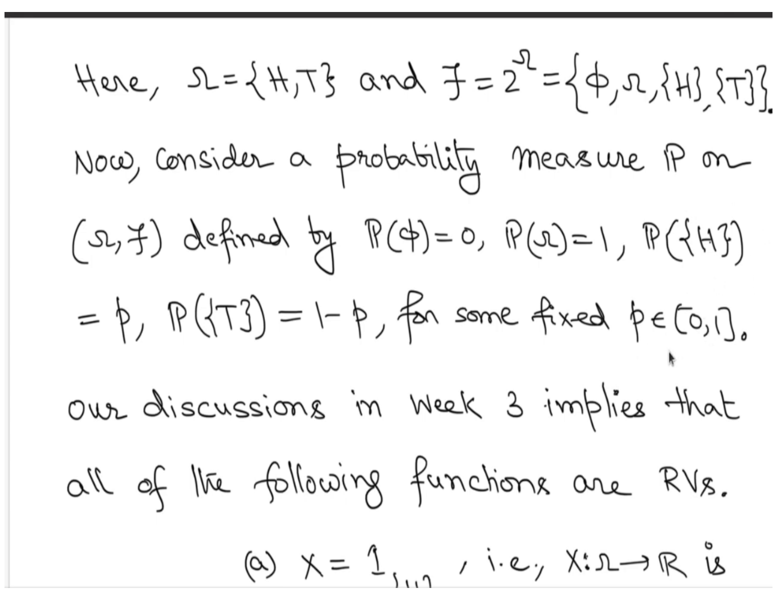
So, for the constant or degenerate random variable, you have explicitly computed all possible pre images and looked at the probability of those, all these events and what you have actually figured out is that $\mathbb{P} \circ X^{-1}$ is nothing but the Dirac measure supported at c. This is the Dirac measure that is appearing on the range side.

(Refer Slide Time: 19:47)



Now, let us look at something more constructive examples. So, take a look at the random experiment of tossing a coin. So, you are going to get heads or tails as outcomes. So, that is your sample space omega. But then, let us choose to consider this kind of a probability measure on the domain side. So, you look at this power set, of course on the domain side. So, this is the finite set, and this is a natural choice. So, the power set appears with this set, so, empty set, the whole set, and the two singleton sets containing heads and the tails.

(Refer Slide Time: 20:23)



So, this is a power set. So, consider this measure on this domain side. So, I am going to associate these values. So, I say that I associate value 0. So, for a probability measure, empty set of course has value 0. Then probability of the whole set, of course, I will associate the value 1. This is a probability measure.

But then you have to specify the values for the singleton sets, heads and the tails. But if you specify the value p, for the heads, then using this property that you can add the probability of the compliment together with the probability of the original set and you will get back 1.

So, you use this property and you will immediately claim that the probability of the singleton set T must be $1 - p$ as long as c is the value of probability of the singleton set heads. So, as soon as you specify this c, all these values are fixed for you. And I claim that this is a probability measure. Stick this.

(Refer slide Time: 21:16)

all of the following functions are RVs.

(a) $X = 1_{\{H\}}$, i.e., $X:\Omega\to\mathbb{R}$ is given by $X(H)=1$, $X(T)=0$.

(b) $X = 1_{\{T\}}$, i.e, $X:\Omega\to\mathbb{R}$ is given by $X(H)=0$, $X(T)=1$.

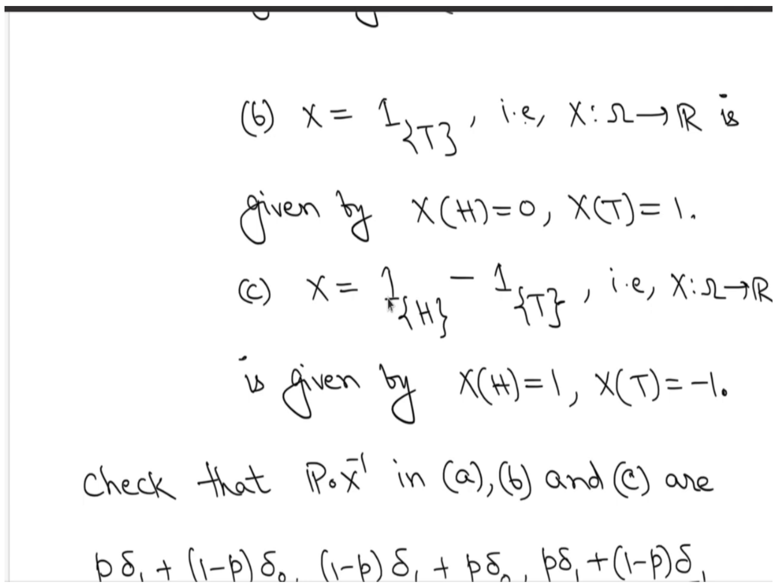
(c) $X = 1_{\{H\}} - 1_{\{T\}}$, i.e, $X:\Omega\to\mathbb{R}$

But now, what you are going to do, is to look at this kind of functions on the domain side. Look at this indicator functions. So, these singletons, as long as these singletons are already appearing on the domain side σ-field, they will give you that these indicators are also measurable functions.

As long as these sets are appearing on the domain side, the corresponding indicators will become measurable. This is something we have discussed in Week 3, as examples of measurable functions.

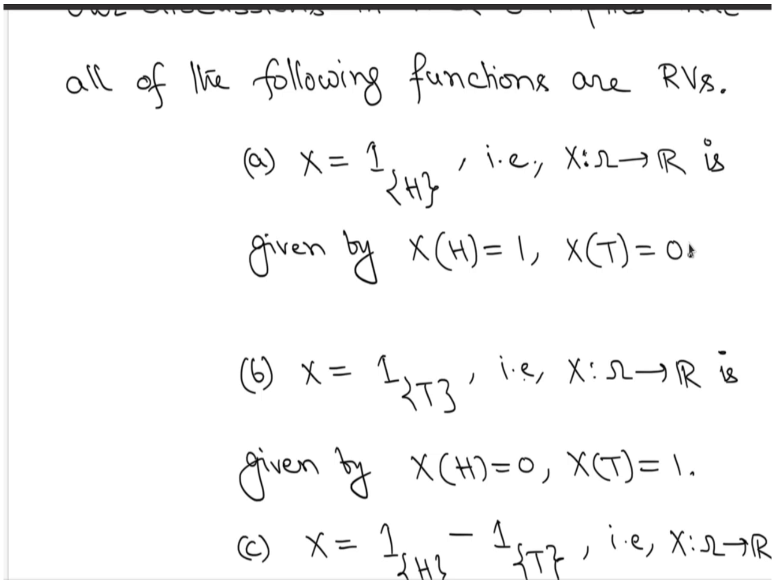
But then, there is a probability measure on the domain side and therefore all these are examples of random variables. So, you are looking at $1_{\{H\}}$, or $1_{\{T\}}$.

(Refer Slide Time: 21:57)



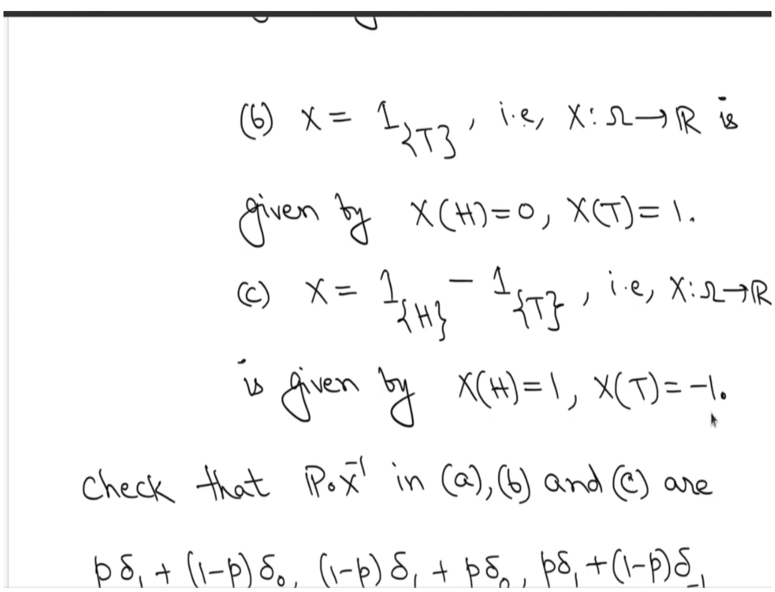
Or, you can choose to look at some combinations of those. As long as you have two measurable functions, their subtraction is also a measurable function. You are using all these nice properties.

(Refer Slide Time: 22:06)



But now, what happens here? You can now look at the functions, $1_{\{H\}}$, for example. Look at this function. So, how does these functions evaluate? On the point H on the domain side, you get the value 1, otherwise you get the value 0. There are the only two points. You get these values.

(Refer Slide Time: 22:26)



Similarly for $1_{\{T\}}$, you can also figure out the appropriate values for the function. Similarly, using the, the subtraction of these two functions, you can write down the actual values of the function X on the points in the domain.

So here, for the head point, you get the value 1, for the tail point, you get the value minus 1. Here exactly, you are associating $\pm\ 1$ value to the heads and the tails. So, this is what associating the numbers, numerical outcomes to random experiments, where there are non numerical outcomes. This is the perfect example of that.

(Refer Slide Time: 23:03)

(a) $X = 1_{\{H\}}$, i.e., $X : \Omega \to \mathbb{R}$ is given by $X(H) = 1$, $X(T) = 0$.

(b) $X = 1_{\{T\}}$, i.e., $X : \Omega \to \mathbb{R}$ is given by $X(H) = 0$, $X(T) = 1$.

(c) $X = 1_{\{H\}} - 1_{\{T\}}$, i.e., $X : \Omega \to \mathbb{R}$ is given by $X(H) = 1$, $X(T) = -1$.

So, you have these explicit measurable functions, X here, a, b, c. In three examples, you have this.

(Refer Slide Time: 23:08)

(c) $X = 1_{\{H\}} - 1_{\{T\}}$, i.e, $X : \Omega \to \mathbb{R}$

is given by $X(H) = 1$, $X(T) = -1$.

Check that $\mathbb{P} \circ X^{-1}$ in (a), (b) and (c) are

$p\delta_1 + (1-p)\delta_0$, $(1-p)\delta_1 + p\delta_0$, $p\delta_1 + (1-p)\delta_{-1}$

respectively (Exercise).

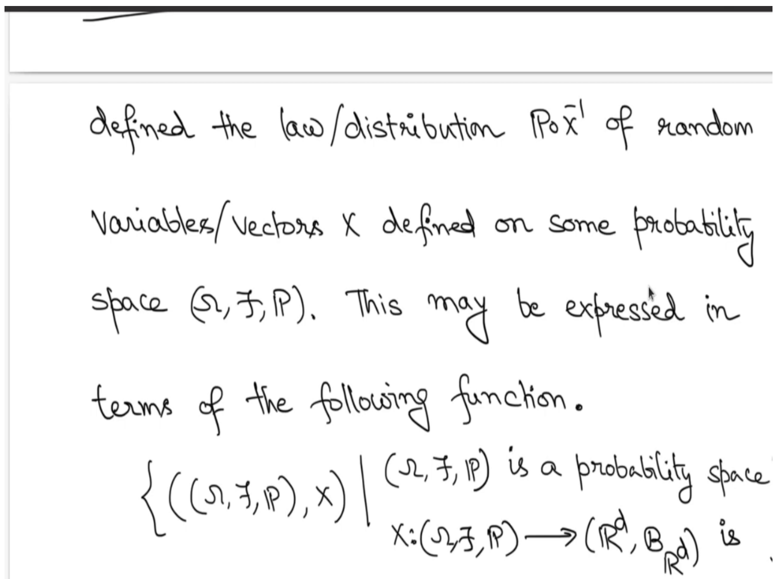
Note (4): In the above discussion, we have

You also have this nice probability measure on the domain side, which you have mentioned. Now, I claim that you can now try to compute the laws of these, $\mathbb{P} \circ X^{-1}$. So, just look at all possible pre images, evaluate the probability p of those events. And you will get, exactly get back certain nice combinations of δ_0s and δ_1s.

And for the function, when you are looking at the subtraction of these two indicators, you are going to get the combinations of δ_{-1} and δ_{-1}. So, this is a very, very interesting observation. You are getting back the convex linear combinations of certain Dirac masses.

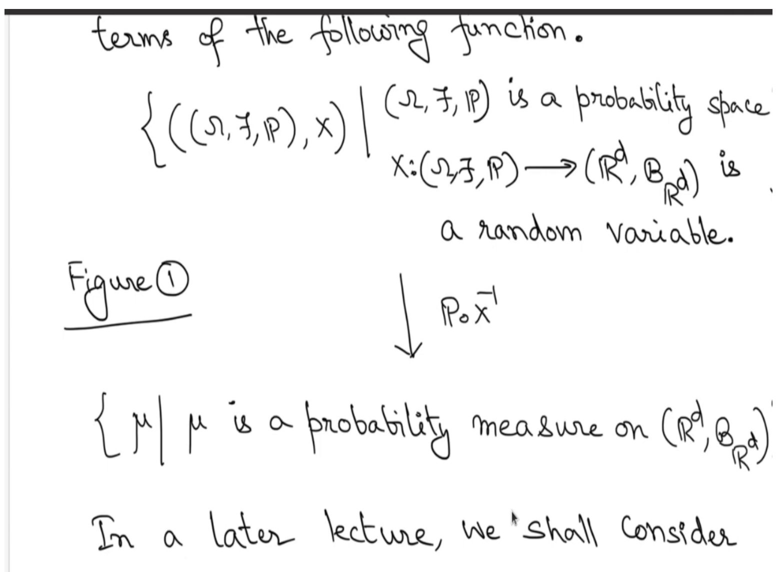
So please work this out. This gives you a very nice connection between the calculations involving indicators and the corresponding laws. So, please try to figure out the law of these random variables.

(Refer Slide Time: 23:57)



So now, let us consolidate whatever we have done in the d dimensional case itself.

(Refer Slide Time: 24:04)



So now, look at this situation. What we have done is that for arbitrary probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable X defined on top of it, we have considered or constructed this probability measure $\mathbb{P} \circ X^{-1}$. And that gave you a example of a probability measure on the range side.

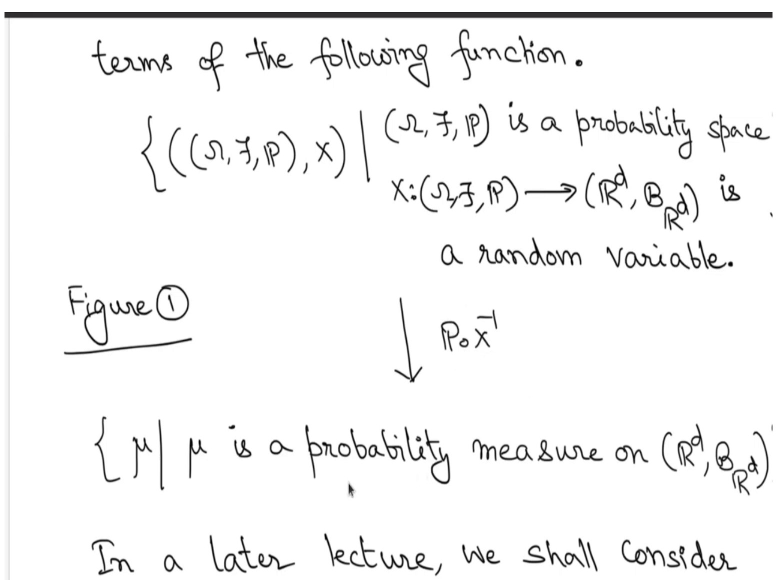
So, look at the first type of things here. $(\Omega, \mathcal{F}, \mathbb{P})$, which is a probability space and random variables, or random vectors X defined on top of it. So, that is the first type of collections. And then, the second type of collections that you are going to get is the collection of all probability measures on Borel σ-field on $\mathbb{R}^d$. These measurable space.

And we are saying that in figure 1, we are obtaining a function from the first collection to the second collection. So, from the first collection, you are looking at all these information, $(\Omega, \mathcal{F}, \mathbb{P})$ and X and using all these, you are constructing this $\mathbb{P} \circ X^{-1}$ which gives you a example of a probability measure.

So, for all these choices $(\Omega, \mathcal{F}, \mathbb{P})$ and X, you have now associated these probability measure $\mathbb{P} \circ X^{-1}$ on the range side. So, you get a probability measure in this collection.

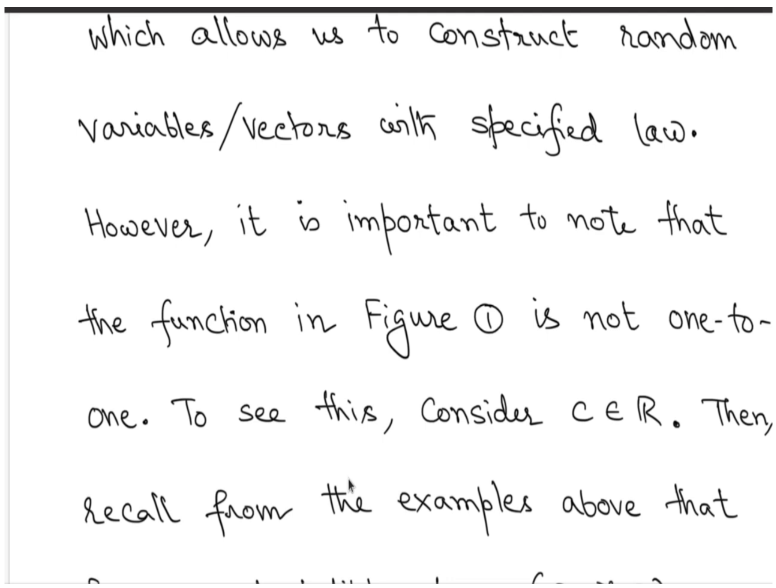
So, from this set, original collection of probability spaces and random vectors, so that is your first set contents. And the second set contents are probability measures. So, you have now made a association, you have now made a function which goes from the first set, or the first collection to the second set.

(Refer Slide Time: 25:44)



But then, we are going to, later, we are to talk about a function going in the opposite direction. So, what we are going to say is that given probability measures, we would like to construct certain probability spaces and random vectors such that the corresponding law $\mathbb{P} \circ X^{-1}$ turn out to be the exactly the probability measure given to us. This, we are going to see later on.

(Refer Slide Time: 26:05)



But this will allow us to construct random variables or random vectors with the specified law. So, you are going to start with specific probability measure on the range side and you are going to construct back examples of random variables or random vectors with that law. This, we will do later.

(Refer Slide Time: 26:32)

the function in Figure 1 is not one-to-one. To see this, Consider $c \in \mathbb{R}$. Then, recall from the examples above that for any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we have a Borel measurable function $X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ defined by $X(\omega) := c$

However, it is important to note that the association, or the function that we have done in figure 1, that from the probability spaces and random vectors you have gone to the probability measures, this is not one to one. How do you show this? So consider the case of constants C and look at the corresponding degenerate random variables. So, go back to simple one dimensional case.

(Refer Slide Time: 26:47)

for any probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we have a Borel measurable function $X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ defined by $X(\omega) := c$

$\forall \omega$. For these choices of $(\Omega, \mathcal{F}, \mathbb{P})$ and X, we have $\mathbb{P} \circ X^{-1} = \delta_c$.

space $(\Omega, \mathcal{F}, \mathbb{P})$. This may be expressed in terms of the following function.

$$\left\{ ((\Omega, \mathcal{F}, \mathbb{P}), X) \,\middle|\, \begin{array}{l} (\Omega, \mathcal{F}, \mathbb{P}) \text{ is a probability space} \\ X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}) \text{ is} \\ \text{a random variable.} \end{array} \right.$$

Figure ①

$$\Big\downarrow \; \mathbb{P} \circ X^{-1}$$

$$\left\{ \mu \,\middle|\, \mu \text{ is a probability measure on } (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}) \right.$$

Figure ①

$$\Big\downarrow \; \mathbb{P} \circ X^{-1}$$

$$\left\{ \mu \,\middle|\, \mu \text{ is a probability measure on } (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}) \right.$$

In a later lecture, we shall consider a function going in the opposite direction, which allows us to construct random variables/vectors with specified law.

So now, in any probability space, you can talk about this Borel measurable function which is this constant or degenerate random variable. And we have seen that for all such choices of probability spaces and random variables defined on top of this, the law will be exactly the Dirac measure supported at c.

So, no matter what probability space you choose, as long as you choose this random variable defined on top of such a probability space as the constant, that constant, then you are going to get this exact law, this exact probability measure on the range side, on the Borel σ-field.

So therefore, you are saying that for all such choices on the domain side of that function that we have now mentioned, so for all such choices, where the random variable is degenerate, that constant, then you are going to get exactly the same probability measure, the Dirac measure supported at the constant.

So therefore, this is not a one-to-one function. So, therefore, when you are talking about the functions that are going in the opposite direction of figure 1, you have to be careful with that. We will see what does it mean later on when we talk about it. But in this lecture, we have constructed the law or distribution of random vectors. So, we are going to continue this discussion in the next lecture. We stop here.