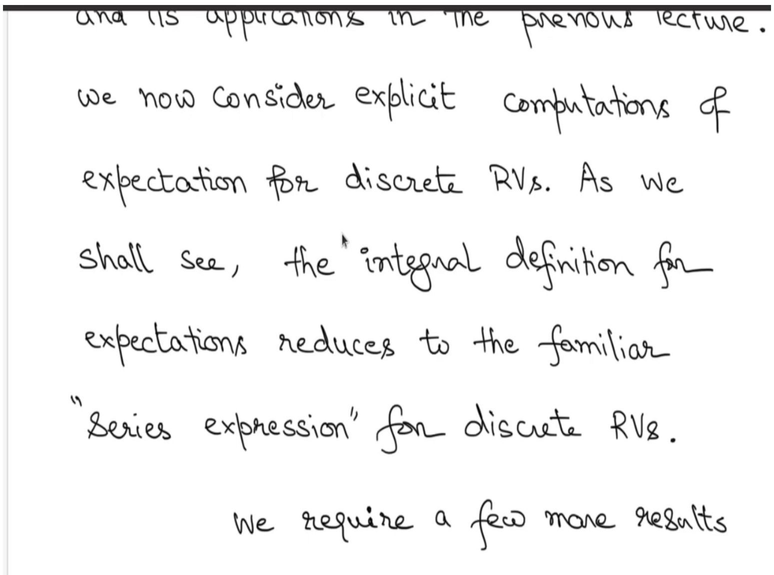
Measure Theoretic Probability -1
Professor Suprio Bhar
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur
Lecture 30
Computation of Expectation for Discrete RVs

Welcome to this lecture, this is the final lecture of week 6. So, before proceeding with the discussion of this lecture, let us first quickly recall what we have been discussing in this week. So, in this week, we have looked at the integration of measurable functions with respect to the given measure on the domain side. We have also seen some major properties of search integration.

And in particular, in the last lecture, we have finally discussed the linear structure of such an integration. In this lecture, we are going to use all these properties and restrict our attention to the special class of random variables, which is the class of discrete random variables, and compute their expectations.

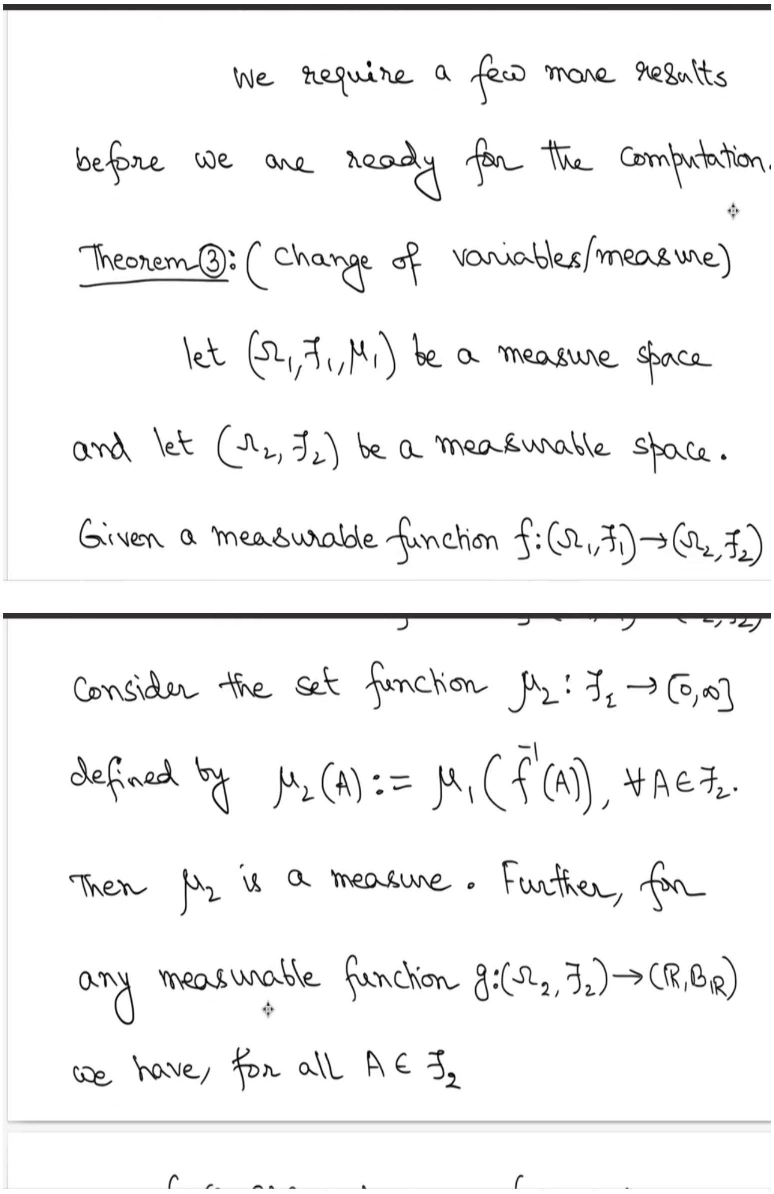
So, we are going to see that the expressions that come out while computing these expectations of discrete random variables will be the same as considered in the case when we discussed it in the basic probability courses. So, let us move on to the slides and discuss things in more concrete detail.

(Refer Slide Time: 01:30)



So, let us start with this idea that the expectation of a random variable as defined in this course, is through that measure theoretic integration. Now, we restrict our attention to these discrete random variables, and we would like to simplify that definition to a more manageable level and we would like to see how it comes out. And we would like to say that it will at the end reduce to the familiar series expression for discrete random variables.

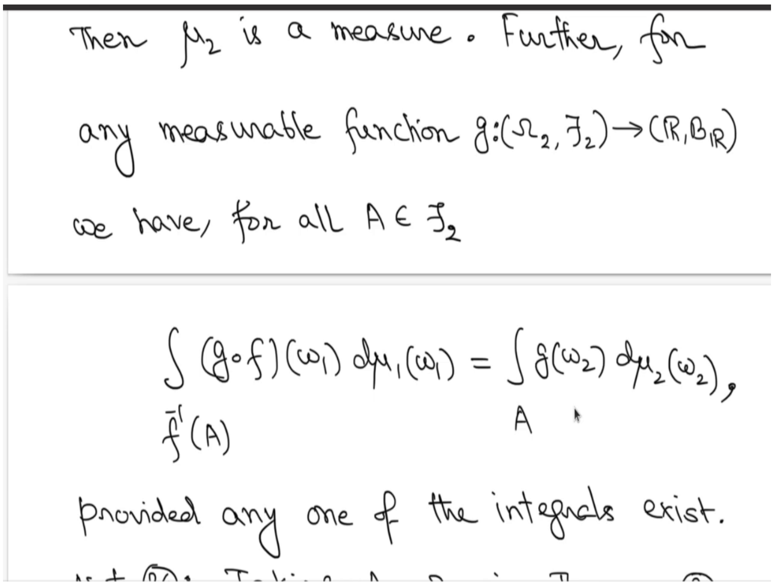
(Refer Slide Time: 02:17)



So, before going to the explicit computations, we require a few more results. So, the first result in this reduction is called the change of variables or change of measure. So, you start off with one measure space $(\Omega_1, \mathcal{F}_1, \mu_1)$, so, that will be taken on the domain side and then take a measurable space on the range side. And what you do is that you look at measurable function call it f from the first measurable space to the second measurable space.

Now, you are going to construct a set function call it μ_2 if $\mathcal{F}_2$ which is on the range side. So, this should be a set function taking values between $[0, \infty]$. So, we define it through the pre-images of all arbitrary sets on the range side under the function f. So, you look at these pre-images and this set is now on the domain side σ-field $\mathcal{F}_1$ by the measurable structure of the function f and you can consider the measure with respect to μ_1. So, what is this measure, that values get associated to this set function. So, we first claim that this set function does define is a measure.

(Refer Slide Time: 03:32)



Further, we want to claim that for any measurable function, now, on the range side, defined on the range side taking values in the real line, call them g, we have this integral relation. So, you are changing measures from μ_1 to μ_2 where you are considering the integrations over μ_2 for sets coming from the range side and when you are considering integration with respect to μ_1 you are of course considering sets on the domain side.

But here note that corresponding to the integration over a set A on the range side, you are going to look at the integration of the pre-image of that set. So, that is where the integration will be concentrated on when you are considering the domain side of it. So, this is the relation that will allow you to do a change of measures from μ_1 to μ_2 or μ_2 back to μ_1. This relation will hold when at least one of the integrals will exist.

(Refer Slide Time: 04:42)

provided any one of the integrals exist.

Note (24): Taking $A = \Omega_2$ in Theorem (3), we have

$$\int_{\Omega_1} g \circ f \, d\mu_1 = \int_{\Omega_2} g \, d\mu_2,$$

since $f^{-1}(\Omega_2) = \Omega_1$.

Note (25): Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be an RV. Take $(\Omega_1, \mathcal{F}_1, \mu_1) = (\Omega, \mathcal{F}, \mathbb{P})$, $(\Omega_2, \mathcal{F}_2)$

Note (25): Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be an RV. Take $(\Omega_1, \mathcal{F}_1, \mu_1) = (\Omega, \mathcal{F}, \mathbb{P})$, $(\Omega_2, \mathcal{F}_2)$
$= (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $f = X$. Then $\mu_2 = \mu_1 \circ f^{-1} = \mathbb{P} \circ X^{-1}$ is the law of X. The fact that $\mathbb{P} \circ X^{-1}$ is a measure has been proved in Proposition (1) of Week 4. Following the

So, before proving this, let us first see what this result suggests. So, take the set A to be the whole set on the range side. Then, what does it say? It says that you have this relation between $\int_{\Omega_1} g \circ f \, d\mu_1$ and $\int_{\Omega_2} g \, d\mu_2$. So, here we are using the fact that since f is a function with range, which is possibly a subset of Ω_2, so the pre-image of Ω_2 is nothing but the whole domain. So, that is what is being used here. So, that is all, that is all we will allow you to write down this equality.

Now, let us focus our attention to the case of random variables for random variables on the domain side, we are given a probability space and on the range side, we are given this familiar measurable space, which is the real line together with the Borel σ-field. Now, consider the previous case, the general case as given in that theorem. So, what you do is that you take the domain side measure space to be $(\Omega,\ \mathcal{F},\ \mathbb{P})$ the probability space and the range side measurable space to be the real line together with the Borel σ-field.

And take the function measureable function f to be the random variable X, then you enter the familiar picture that the measure μ_2 that you have thus defined is nothing but $\mathbb{P} \circ X^{-1}$, which is the law of X. So, we have extensively discussed about the measure theoretic properties of this $\mathbb{P} \circ X^{-1}$ and the fact that $\mathbb{P} \circ X^{-1}$ is a measure that also has been discussed earlier.

(Refer Slide Time: 06:25)

$= \mathbb{P} \circ X^{-1}$ is the law of X. The fact that $\mathbb{P} \circ X^{-1}$ is a measure has been proved in Proposition ① of Week 4. Following the same argument, prove the assertion in Theorem ③ above that $\mu_1 \circ f^{-1}$ is a measure. (Exercise)

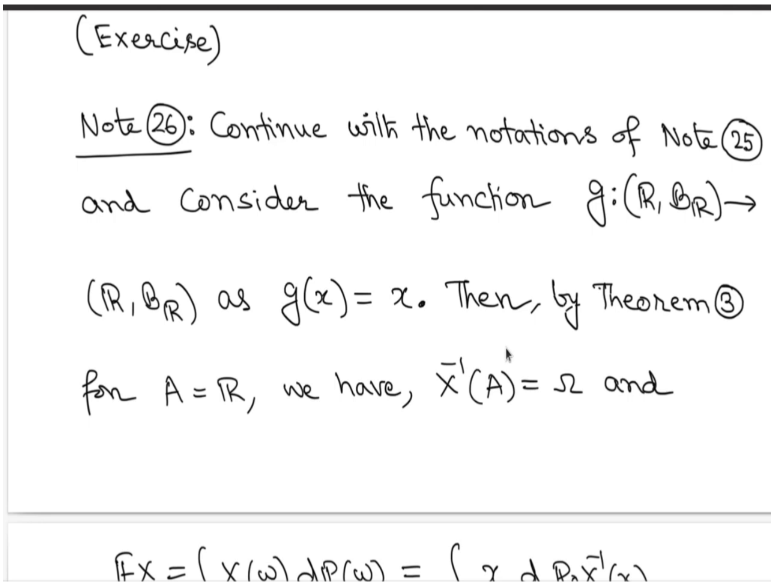
$= (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and $f = X$. Then $\mu_2 = \mu_1 \circ f^{-1}$

$= \mathbb{P} \circ X^{-1}$ is the law of X. The fact that $\mathbb{P} \circ X^{-1}$ is a measure has been proved in Proposition ① of Week 4. Following the same argument, prove the assertion in Theorem ③ above that $\mu_1 \circ f^{-1}$ is a measure.

So, following the similar argument, you can now try to prove that this $\mu_1 \circ f^{-1}$ now, this is a more general structure as stated in theorem 3 above. So, this is the more general structure, but it follows the same proof same argument that will allow you to show that μ_2 will become a measure. So, for the case of random variables, you actually had that $\mathbb{P} \circ X^{-1}$ is a probability measure, but in general, for general measures μ_1 and general measurable functions f you will only end up with a measure.

So, the push forward of a measure by a measurable function remains a measure. So, this argument is left as exercise, but as mentioned earlier, this discussion simply is the same as done for the case of $\mathbb{P} \circ X^{-1}$ the law of a random variable.

(Refer Slide Time: 07:20)



$(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ as $g(x) = x$. Then, by Theorem 3

for $A = \mathbb{R}$, we have, $X^{-1}(A) = \Omega$ and

$$\mathbb{E}X = \int_{\Omega} X(\omega)\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} x \, d\, \mathbb{P}\circ X^{-1}(x),$$

provided one of the integrals exist. This allows us to write $\mathbb{E}X$ in terms of $\mathbb{P}\circ X^{-1}$.

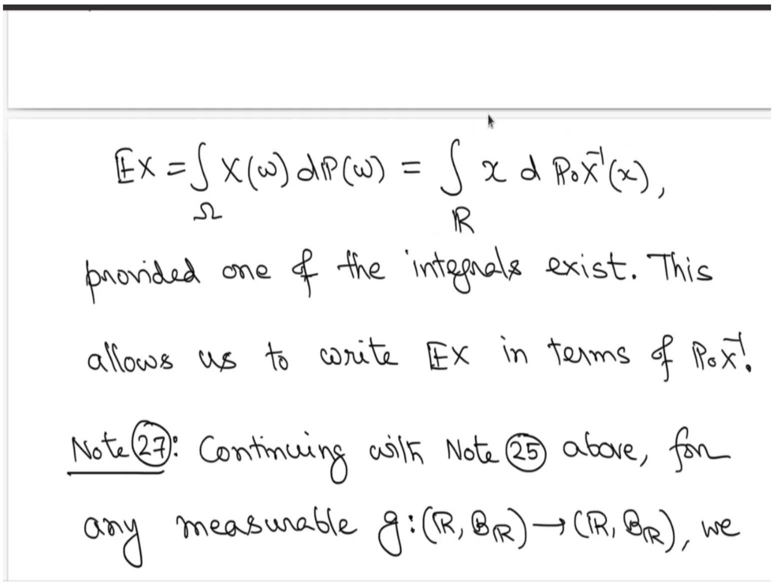
Now continue with this situation. Now, choose this g function from $\mathbb{R}$ to $\mathbb{R}$ itself and assume that it is Borel measurable. So, choose this function $g(x) = x$ the identity function. Then by theorem 3, what you can now say that for the whole range side, take the set A to be the whole range side, then what is a pre-image?

Pre-image is the whole domain. And therefore, you have this change of measure relation that

$\int_{\Omega} X(\omega)\, d\mathbb{P}(\omega)$ that we have been considering while talking about the $E(X)$ can now be written as an integration over $\mathbb{R}$ by changing the measure from $\mathbb{P}$ to $\mathbb{P} \circ X^{-1}$.

So, that is all is the application of theorem 3, when you are talking about random variable. So, you are changing the integration from over Ω to the over $\mathbb{R}$ and you have this integration now, that will come from the choice of the function g as the identity function, which takes a real number x to itself.

(Refer Slide Time: 08:30)



So, therefore, you have now written $E(X)$ as integration over the real line with respect to the law $\mathbb{P}\circ X^{-1}$, where $\mathbb{P}\circ X^{-1}$ is a genuine probability measure on $\mathbb{R}$. So, therefore, your original definition that was via this probability measure $\mathbb{P}$ on the domain side now gets transferred to the range side in terms of the law $\mathbb{P}\circ X^{-1}$, and this relation will hold provided one of the integral exists.

So, this is as per the theorem 3 above. Now, this is allowing us to write $E(X)$ in terms of the law, $\mathbb{P}\circ X^{-1}$. So, this is, therefore, saying that fact that expected value is purely dependent on the law, $\mathbb{P}\circ X^{-1}$. So, you do not have to worry about the actual probability measure $\mathbb{P}$, or the actual random variable X. As long as you know the law $\mathbb{P}\circ X^{-1}$, just look at the integration over the real line $\int_{\mathbb{R}} x\, d\mathbb{P}\circ X^{-1}(x)$. And if you can compute this, you can get $E(X)$

(Refer Slide Time: 09:42)

Note (27): Continuing with Note (25) above, for any measurable $g:(\mathbb{R},\mathcal{B}_{\mathbb{R}}) \to (\mathbb{R},\mathcal{B}_{\mathbb{R}})$, we have

$$\mathbb{E}\, g(X) = \int_{\Omega} g(X(\omega))\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x)\, d\mathbb{P}\circ X^{-1}(x),$$

provided one of the integrals exist. For example, to check the integrability of X, we need $\mathbb{E}|X| < \infty$. Here, we can take

Say again continuing with this, but let us now choose a more general function g here. So, instead of the function $g(x) = x$, now choose a more general function. So, what do you do here? Now here what will happen is that you can now consider the expected value of the random variable g composed with X or $g(X)$. And that as per definition is $\int_{\Omega} g(X(\omega))\, d\mathbb{P}(\omega)$. But again, if you change variables, you again end up with this integration, which is over the real line now. So, you are trying to integrate this measureable function g on the real line with respect to the property measure $\mathbb{P} \circ X^{-1}$. So, again, if you can make sense of one of the integrations, then this equality holds.

And moreover, if you can compute the integration over the real line, as given here, you do not have to compute the integration with respect to the probability measure $\mathbb{P}$, you can just take that same value and consider it as $E(g(X))$. So, this is the advantage that you are transferring all the information in terms of $\mathbb{P} \circ X^{-1}$.

(Refer Slide Time: 10:54)

$$\mathbb{E}\, g(X) = \int_{\Omega} g(X(\omega))\, d\mathbb{P}(\omega) = \int_{\mathbb{R}} g(x)\, d\mathbb{P}\circ X^{-1}(x),$$

provided one of the integrals exist. For example, to check the integrability of X, we need $\mathbb{E}|X| < \infty$. Here, we can take $g(x) = |x|, \forall x$ and instead of checking $\int_{\Omega} |X(\omega)|\, d\mathbb{P}(\omega) < \infty$, we may alternatively check $\int |x|\, d\mathbb{P}\circ X^{-1}(x) < \infty$.

we need $\mathbb{E}|X| < \infty$. Here, we can take $g(x) = |x|, \forall x$ and instead of checking $\int_{\Omega} |X(\omega)|\, d\mathbb{P}(\omega) < \infty$, we may alternatively check $\int_{\mathbb{R}} |x|\, d\mathbb{P}\circ X^{-1}(x) < \infty$.

Proof of Theorem 3:

Note that $g\circ f : (\Omega_1, \mathcal{F}_1) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is

Now here the, here is an interesting observation, choose the function $g(x) = |x|$. What will happen here is that then you are getting the fact that $E(|g(x)|)$ or $E(|x|)$ that is exactly equal to $\int_{\mathbb{R}} |x|\, d\mathbb{P} \circ X^{-1}(x)$. So, instead of checking these complicated integration of $|x|$ with respect to the probability measure $\mathbb{P}$, you may alternatively check this condition $\int_{\mathbb{R}} |x|\, d\mathbb{P} \circ X^{-1}(x)$.

So, instead of doing the integration over the domain side Ω, you do the integration over the range side which is on $\mathbb{R}$. So, this is the alternative condition, and may be easier to work out. Great, so, with these observations at hand, let us now go to the proof of theorem 3.

(Refer Slide Time: 11:56)

Note that $g \circ f : (\Omega_1, \mathcal{F}_1) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is measurable. First, consider the case when g is an indicator function, i.e., $g = 1_B$ for some $B \in \mathcal{F}_2$. Then

$$\int_{f^{-1}(A)} (g \circ f)(\omega_1)\, d\mu_1(\omega_1) = \int_{f^{-1}(A)} 1_B(f(\omega_1))\, d\mu_1(\omega_1)$$

$$\int_{f^{-1}(A)} (g \circ f)(\omega_1)\, d\mu_1(\omega_1) = \int_{f^{-1}(A)} 1_B(f(\omega_1))\, d\mu_1(\omega_1)$$

$$= \int_{f^{-1}(A)} 1_{f^{-1}(B)}(\omega_1)\, d\mu_1(\omega_1) = \int_{\Omega_1} 1_{f^{-1}(B)}\, 1_{f^{-1}(A)}\, d\mu_1$$

$$= \int_{\Omega_1} 1_{f^{-1}(B) \cap f^{-1}(A)}\, d\mu_1 = \int_{\Omega_1} 1_{f^{-1}(A \cap B)}\, d\mu_1$$

So, again, we take our familiar approach here, $(\Omega_1, \mathcal{F}_1, \mu_1)$ these are fixed, the function f is fixed, and what you want to do is to show that for any measurable function g, g will have this relation appropriate relation. So, that is what we want to show. First comment is that $g \circ f$ that will be measurable, because composition preserves measurability, so, as long as the composition is defined, you will get a measurable function.

So, $g \circ f$ here is defined on the domain, Ω_1 with the σ-field $\mathcal{F}_1$, and the range as the real line together with the Borel σ-field. Now, what do you want to show is that you want to show that appropriate change of variable formula or change of measure formula holds. To do that we follow our familiar approach, we want to prove it for indicator functions first, and then we would like to take it over by linearity to simple functions by limiting arguments to non-negative measurable functions.

And then finally, by linearity to general measurable functions. So, let us verify the equality when g is an indicator. Choose a set B on the domain side, for g. Here, what is happening is that you want to verify this integration equals something else with respect to the integration with respect to μ_2. So, let us start with this expression and put the fact that g is nothing but this 1_B.

So, when you put that in, put in the definition, so this is $f(\omega_1)$, we are checking whether it is in the set B or not. But then that can be rewritten in terms of $f^{-1}(B)$, and you can just check whether the point sample point small omega1 belongs to the set $f^{-1}(B)$. So, observe that $f(\omega_1) \in B$ if and only if $\omega_1 \in f^{-1}(B)$.

So, that is all we are using here. But then the integration is over the set $f^{-1}(A)$. So, if you put in the definition for integrations like that, then all you have to do is to look at the integration about the whole set and multiply by the appropriate indicator of that set. So, earlier, the integration of was over $f^{-1}(A)$, you just brought in this indicator of $f^{-1}(A)$, and you could now do the integration over the whole set omega1.

(Refer Slide Time: 14:32)

$$= \int_{f^{-1}(A)} 1_{f^{-1}(B)}(\omega_1)\, d\mu_1(\omega_1) = \int_{\Omega_1} 1_{f^{-1}(B)}\, 1_{f^{-1}(A)}\, d\mu_1$$

$$= \int_{\Omega_1} 1_{f^{-1}(B)\cap f^{-1}(A)}\, d\mu_1 = \int_{\Omega_1} 1_{f^{-1}(A\cap B)}\, d\mu_1$$

$$= \mu_1\left(f^{-1}(A\cap B)\right) = \mu_2(A\cap B) = \int_A 1_B\, d\mu_2 .$$

$$f^{-1}(A) \quad f^{-1}(B) \qquad \Omega_1 \quad f^{-1}(B) \quad f^{-1}(A)$$

$$= \int_{\Omega_1} 1_{f^{-1}(B)\cap f^{-1}(A)} \, d\mu_1 = \int_{\Omega_1} 1_{f^{-1}(A\cap B)} \, d\mu_1$$

$$= \mu_1(f^{-1}(A\cap B)) = \mu_2(A\cap B) = \int_A 1_B \, d\mu_2 .$$

The proof can now be completed by using linearity and limiting behaviour of the

Now, this is the product of two indicators. So, therefore, as observed earlier, this is nothing but the indicator of the intersection of the two sets, we will just write that down. But then you are now saying that you are going to integrate the indicator of some set with respect to the measure μ_1. Now, using the properties of the pre-image you can also rewrite it as $f^{-1}(A \cap B)$. Now since you are just doing the integration of this indicator function then this is nothing but $\mu_1(f^{-1}(A \cap B))$.

So, that is all you are doing here. Great but that as per definition, μ_2 is nothing but $\mu_1(f^{-1})$, and therefore, that is all you write that this is $\mu_2(A \cap B)$, but then you can easily write it as $\int_A 1_B \, d\mu_2$. So, this is the integration of this indicator function with respect to the measure μ_2 over the set A. So, that is just rewriting this integration procedure.

(Refer Slide Time: 15:39)

$$\Omega_1 \quad f^{-1}(B)\cap f^{-1}(A) \qquad \Omega_1 \quad f^{-1}(A\cap B)$$

$$= \mu_1\left(f^{-1}(A\cap B)\right) = \mu_2(A\cap B) = \int_A 1_B \, d\mu_2 .$$

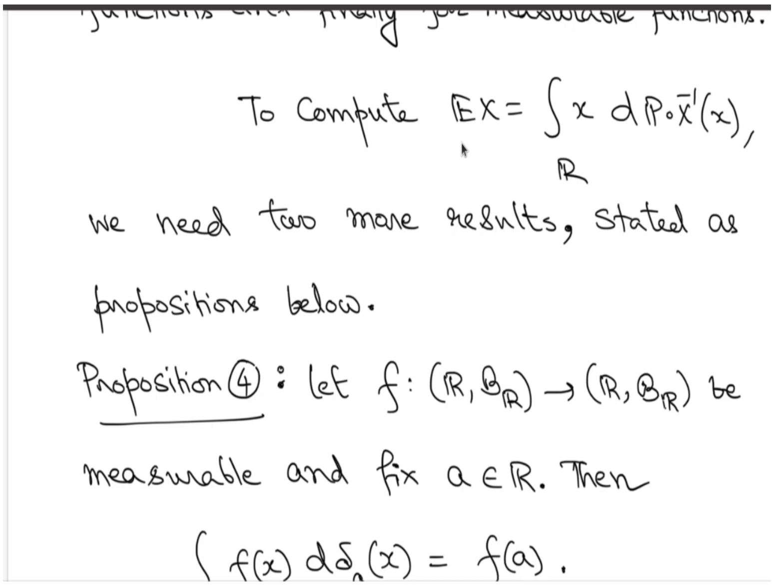
The proof can now be completed by using linearity and limiting behaviour of the integration. Using these properties, the statement is proved for simple functions, and then for non-negative measurable

linearity and limiting behaviour of the integration. Using these properties, the statement is proved for simple functions, and then for non-negative measurable
functions and finally for measurable functions.

To compute $\mathbb{E}X = \int_{\mathbb{R}} x \, d\,\mathbb{P}\circ X^{-1}(x)$,

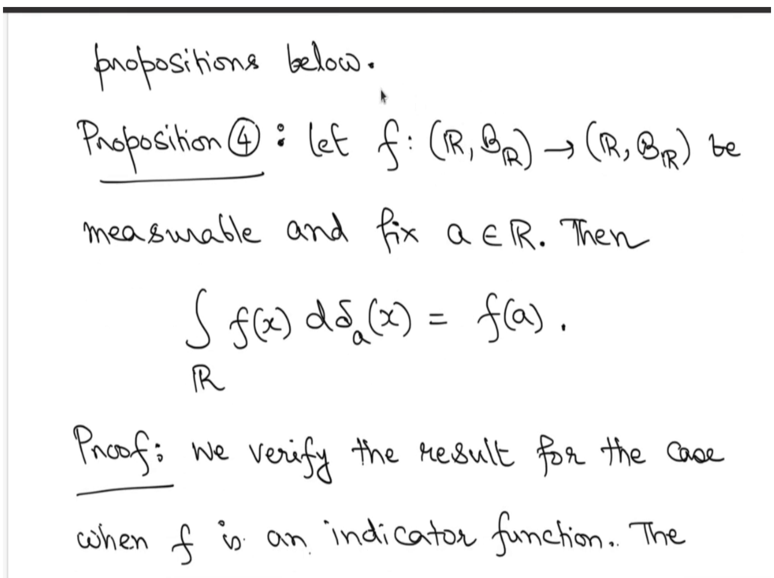
So, therefore, you have now proved the required equality, when g is a indicator. Now follow the standard approach, use linearity and limiting behavior to extend it to any general measurable function provided the integrations makes sense. So, that completes the proof for theorem 3, and this allows us to state this result that change a variable or change of measure is possible when you are talking about this measure theoretic integration.

(Refer Slide Time: 16:09)



And in particular, we have already commented earlier that one of the useful implications of this theorem is the rewriting of the expression for $E(x)$, which can be written as now, $\int_{\mathbb{R}} x \, d\mathbb{P} \circ X^{-1}(x)$. So, now, before doing these explicit computations, we need a couple of more results, and which are stated below. So, these will help us to do the simplification in our arguments.

(Refer Slide Time: 16:42)



So, let us start with any measurable function defined on the real line. So, here, we are talking about real valued measurable functions defined on the real line. So, what do you want to do we want to first look at the integration of any such measurable function f with respect to a

Dirac measure. So, choose a point a in the real line and look at the Dirac measure supported at the point a. $\int_{\mathbb{R}} f(x)\, d\delta_a(x)$.

Then we claim that the integration of this measurable function with respect to the Dirac measure is nothing but the evaluation of the function at the point, i.e. $f(a)$. So, this is a very simple relation. This says that once you try to integrate a function with respect to the Dirac measure, all you have to do is to evaluate the function at that point. So, how do you verify this again, follow the standard approach, first verify it for the case when ff is an indicator.

(Refer Slide Time: 17:35)

when f is an indicator function. The

general case follows by standard arguments.

let $f = 1_B$ for some $B \in \mathcal{B}_{\mathbb{R}}$.

Then $\int f(x)\, d\delta_a(x) = \int 1_B(x)\, d\delta_a(x)$

$= \delta_a(B) = \begin{cases} 1 & \text{if } a \in B \\ 0 & \text{otherwise} \end{cases}$

Then $\int f(x)\, d\delta_a(x) = \int 1_B(x)\, d\delta_a(x)$

$= \delta_a(B) = \begin{cases} 1 & \text{if } a \in B \\ 0 & \text{otherwise} \end{cases}$

$= 1_B(a) = f(a)$.

This completes the proof.

Recall from Exercise 5 of week 2

$$\int_{\mathbb{R}} f(x)\, d\delta_a(x) = f(a).$$

Proof: We verify the result for the case when f is an indicator function. The general case follows by standard arguments.

let $f = 1_B$ for some $B \in \mathcal{B}_{\mathbb{R}}$.

Then $\int f(x)\, d\delta_a(x) = \int 1_B(x)\, d\delta_a(x)$

So, let us see this. So, if f is 1_B for some Borel set B then put in the definition. So, $1_B(x)$. But now, this is nothing but the Dirac measure applied to the set B, this is as per definition so if you integrate indicator function, all you have to do is to look at the size of the set with respect to the given measure. So, that is all you are doing, but observe that Dirac measure associates these values, that if $a \in B$ then you associate the value 1, otherwise you associate the value 0.

But you can rewrite this in terms of the indicator in this way that this is the evaluation of the indicator function at the point a, this is just rewriting the relation or the values of Dirac measure in terms of an indicator now. But now, what was 1_B? So that was your choice for f in this simple case. So, you have verified the equality when f is an integrator, but then what do you do for the general case, you apply the appropriate linearity and limiting arguments to go to the general case.

So, as soon as you verify it for indicator functions, and then standard procedure takes over, as long as all the integration makes sense, the equalities will continue to hold on. So, this is one of the main results that will help us in computation of expectations for discrete random variables.

(Refer Slide Time: 19:08)

Recall from Exercise 5 of week 2 that addition of measures gives more examples of measures. We had also seen that multiplication of a measure by positive scalars will also produce measures.

Proposition 5: let $\mu_1, \mu_2, \ldots$ be measures

positive scalars will also produce measures.

Proposition 5: let $\mu_1, \mu_2, \ldots$ be measures
on a measurable space $(\Omega, \mathcal{F})$. Then for any measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ we have

(i) $\int f \, d(\mu_1 + \mu_2) = \int f \, d\mu_1 + \int f \, d\mu_2$

(ii) $\int f \, d\left(\sum_{n=1}^{\infty} \mu_n\right) = \sum_{n=1}^{\infty} \int f \, d\mu_n$

for any measurable function $f:(\Omega,\mathcal{F})\to(\mathbb{R},\mathcal{B}_{\mathbb{R}})$ we have

(i) $\int f \, d(\mu_1+\mu_2) = \int f \, d\mu_1 + \int f \, d\mu_2$

(ii) $\int f \, d\left(\sum_{n=1}^{\infty}\mu_n\right) = \sum_{n=1}^{\infty}\int f \, d\mu_n$,

(iii) $\int f \, d(\alpha\mu) = \alpha\int f \, d\mu, \ \forall \alpha > 0$

provided the integrals on the right hand

But then, there is one more result that we need to recall from our earlier discussions in week 2. So, recall from this exercise 5, that addition of measures gives other measures. So, what was this that if you have a sequence of measures, let us say, then, what you will get is that $\mu_1 + \mu_2 + \mu_3 + . . .$, will give you a measure, you can also state the same result for finite such number of measures.

So, if you have μ_1 up to μ_n R measures, then $\mu_1 + \mu_2 + \mu_3 + . . . + \mu_n$, that finite sum will also be a measure. But then there is also another operation which was scalar multiplication. So, if you choose positive scalars and multiply it to a measure, then you get other measures. So, in fact, we had use such operations to scale finite measures to probability measures. But of course, given any general measure, you can multiply by a positive scalar and get other examples of measures.

So, with that result in mind, we now look at the integration of measurable functions with the measures thus constructed. So, in particular, let us take two measures μ_1 and μ_2. And we claim that integration of this measurable function over this measurable space omega f, where we are assuming that all the measures are defined. So, if you are going to integrate the function f with respect to the measure $\mu_1 + \mu_2$, all you have to do is to separately integrate the function with respect to μ_1, and with respect to μ_2 and then add them up.

And the same relation is true when you are talking about this infinite combination of μ_n. So, all you have to do is to individually integrate them and then add up the results. And finally, for scalar multiplications, if you multiply this measure by this positive scalar α, then this α

can simply come out of the integration and all you have to do is to integrate the function with respect to the given measure μ.

So, these results will continue to hold provided the integrations that you are looking at or the terms that you are looking at on the right-hand side makes sense. In particular, if you are looking at this infinite linear combinations, you cannot have $\infty - \infty$ situations anywhere.

So, again, as long as these summations make sense, you can get these equalities. Expression on the right hand side hold, as long as they exist, you can talk about the appropriate equalities. So, this is one other result that is going to help us in computing the expectation for discrete random variables.

(Refer Slide Time: 21:53)

integrals can be defined.

Proof: We discuss the proofs of (i) and (iii). Proof of part (ii) is left as an exercise.

Let $f = 1_B$ for some $B \in \mathcal{F}$. Then

$$\int f \, d(\mu_1+\mu_2) = \int 1_B \, d(\mu_1+\mu_2) = (\mu_1+\mu_2)(B)$$

$$= \mu_1(B) + \mu_2(B) = \int 1_B \, d\mu_1 + \int 1_B \, d\mu_2$$

for any measurable function $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ we have

(i) $\int f \, d(\mu_1+\mu_2) = \int f \, d\mu_1 + \int f \, d\mu_2$

(ii) $\int f \, d\left(\sum_{n=1}^{\infty} \mu_n\right) = \sum_{n=1}^{\infty} \int f \, d\mu_n$,

(iii) $\int f \, d(\alpha\mu) = \alpha \int f \, d\mu, \ \forall \alpha > 0$

provided the integrals on the right hand

So, how do you prove this? So, let us go back to the statements 1, 2, and 3. So, the statement 2 is just a generalization of finite addition to infinite additions. So, once you prove it, for the finite additions case, the argument will follow by some appropriate limiting procedure so that we will be left as an exercise. So, let us first focus our attention to I and III. So, in I, we are talking about finite additions, finite number of additions, and in III, we are talking about scalar multiplication.

So, let us try to work them out. And again, we are going to use this familiar procedure that we are going to work with indicator functions. So, let us choose this function f to be 1_B, where B is coming from the domain side. So, let us look at the integration now. So, f is chosen to be 1_B, and you want to integrate it with respect to the measure, $\mu_1 + \mu_2$, but that is nothing but the measure of the set B with respect to the measure $\mu_1 + \mu_2$.

Now, as per definition, this is measure of the set B with respect to μ_1, and you are going to add to it the measure of the set B with respect to the measure μ_2. So, that is as per definition of the set function $\mu_1 + \mu_2$. But then, individually, they are nothing but $\int 1_B \, d\mu_1$ and $\int 1_B \, d\mu_2$ and therefore you get the required relation when f is the indicator function.

(Refer Slide Time: 23:32)

$$= \mu_1(B) + \mu_2(B) = \int 1_B \, d\mu_1 + \int 1_B \, d\mu_2.$$

and

$$\int f \, d(\alpha\mu) = \int 1_B \, d(\alpha\mu) = (\alpha\mu)(B)$$

$$= \alpha\mu(B) = \alpha\int 1_B \, d\mu = \alpha \int f \, d\mu.$$

The general case follows by standard arguments

Note (28): We now compute the expectations of

Let us try to verify same irrelevant equalities when f is an integrator, and the measure is getting multiplied by a positive scalar. So, let us try to look at this. So, let us again put in that

f as 1_B, and then again, put in the definition. So, this is the measure of the set B with respect to the measure $\alpha\mu$, but again, as per definition of the set function $\alpha\mu$, this is nothing but $\alpha\mu(B)$.

But then you write $\mu(B)$ as $\int 1_B\, d\mu$. And that is nothing but $\alpha \int f\, d\mu$. So, therefore, you have verified the required relations when f is an indicator, and the general case will follow for by the standard arguments, all you have to check is that the necessary integrals exist. If the integrals exist, you can ensure that the equality holds by this standard procedure.

(Refer Slide Time: 24:38)

Note (28): We now compute the expectations of discrete RVs. Let $X:(\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be a discrete RV with jumps $\mathbb{P}(X = x_j) = \mathbb{P}\circ X^{-1}(\{x_j\})$ on the points x_j in the support S. Recall from Note (22) of Week 4 that $\mathbb{P}\circ X^{-1} = \sum_{x_j \in S} \mathbb{P}\circ X^{-1}(\{x_j\})\, \delta_{x_j}$.

a discrete RV with jumps $\mathbb{P}(X = x_j) = \mathbb{P}\circ X^{-1}(\{x_j\})$ on the points x_j in the support S. Recall from Note (22) of Week 4 that $\mathbb{P}\circ X^{-1} = \sum_{x_j \in S} \mathbb{P}\circ X^{-1}(\{x_j\})\, \delta_{x_j}$.

Then by Propositions (4) and (5),

$$\int_{\mathbb{R}} x \, d\,\mathbb{P}\circ X^{-1}(x) = \sum_{x_j \in S} \mathbb{P}\circ X^{-1}(\{x_j\}) \int x \, d\delta_{x_j}(x)$$

$$= \sum_{x_j \in S} x_j \, \mathbb{P} \circ X^{-1}(\{x_j\}),$$

provided the series on the right hand side makes sense. Thus, we get back the usual formula* for expectation of discrete RVs, as seen in basic probability courses.

So, we are now ready to compute the expectations of discrete random variables. So, this is the culmination of all the things that we have been discussing so far. So, let us try to describe the setting and then we will do the computation. So, let us start with this random variable X, defined on a probability space $(\Omega, \mathbb{P})$ and taking values in the real line. So, the measureable structure will be taken with respect to the σ-field $\mathcal{F}$ on the domain side and the Borel σ-field on the range side.

So, suppose this is a discrete random variable, so, you can now specify where the jumps are occurring. So, suppose the jump points are x_j, which you put in the support S and then what is the jump size so, that is nothing but $\mathbb{P}(X) = x_j$. And another way of writing these jumps size in terms of the law is nothing but $\mathbb{P} \circ X^{-1}(\{x_j\})$. So, this is the size of the singleton set $\{x_j\}$ with respect to the law of $\mathbb{P} \circ X^{-1}$.

So, this we had discussed earlier. Now, recall from note 22 of week 4 that, in this case the law can be written as a linear combination, a convex linear combination of the Dirac masses situated at this jump points x_j. And the sum is over all points in the support. And the weights that you put are exactly the jump sizes for x_j for the point x_j, you just look at the size of the singleton set $\{x_j\}$ with respect to the law $\mathbb{P} \circ X^{-1}$.

And that as we have already mentioned, this is $\mathbb{P}(X) = x_j$. So, this is the linear combination, the convex linear combination of the Dirac masses situated at the points in the support, and that is going to give you back the law of $\mathbb{P} \circ X^{-1}$. So, $\mathbb{P} \circ X^{-1}$ is nothing but a convex linear combination of these Dirac masses. Now, remember, we have done integration with respect to the Dirac masses, we have also learned how to do how to split the integration of over linear combinations of measures.

So, this was discussed in proportions 4, and 5. So, use that results and what do you get, you have to now compute these kind of an integration that appeared in the expectations. So, here, you first use the change of variable formula to move the integration over the domain side Ω_2, integration over the range side with respect to the measure $\mathbb{P} \circ X^{-1}$. So, that is our first step. Now, you rewrite the $\mathbb{P} \circ X^{-1}$ the law as a linear combination of the Dirac masses.

And as for propositions earlier, this scalar simply come out the summation simply come out and you only have to do integrate this function which is the identity function x here with respect to the Dirac mass. And for Dirac masses, if you want to integrate any measurable function, all you have to do is to evaluate that function at that point where the Dirac mass is situated.

And therefore, integration of the identity function over this point over this Dirac mass x_j is nothing but x_j. This is just the evolution of the identity function at the point x_j that is all therefore, you get back your familiar series expression for expectation of discrete random variables.

So, this is the exact same expression that you have already seen in your basic probability course. So, here are what we have just done, we have started off with the measure theoretic definition of $\mathbf{E}\,X$, which was defined to be integration of the measurable function X with respect to the probability measure $\mathbb{P}$ and the integration was over the domain set Ω, then we use change of variable formula to move to integration over the law $\mathbb{P} \circ X^{-1}$.

And once you change this variable or measure, you end up with the integration over the real line and you are integrating the identity function x here. But then you observe that $\mathbb{P} \circ X^{-1}$ is nothing but a convex linear combination of the Dirac masses. So, therefore, you first bring out that summation and then bring out the appropriate corresponding scalars these are all again non-negative scalars positive scalars because jumps are non-trivial jumps and then all you have to do is to integrate this identity function with respect to the Dirac mass.

So, these are all the things that you are using to get back your familiar expression that you had seen in your basic probability courses. So, the general measure theoretic integration that we have done already gives you your familiar expression for discrete random variables. Later on, we are going to see that this general definition while restricted to other special cases give you back your familiar expressions.

So, therefore, the measure theoretic integration is a general structure and once you prove any result in this general structure, all these particular cases will follow and you can apply all these general results to all these particular cases. So, that is the power of measure theory.

(Refer Slide Time: 30:02)

side makes sense. Thus, we get back the usual formula for expectation of discrete RVs, as seen in basic probability courses.

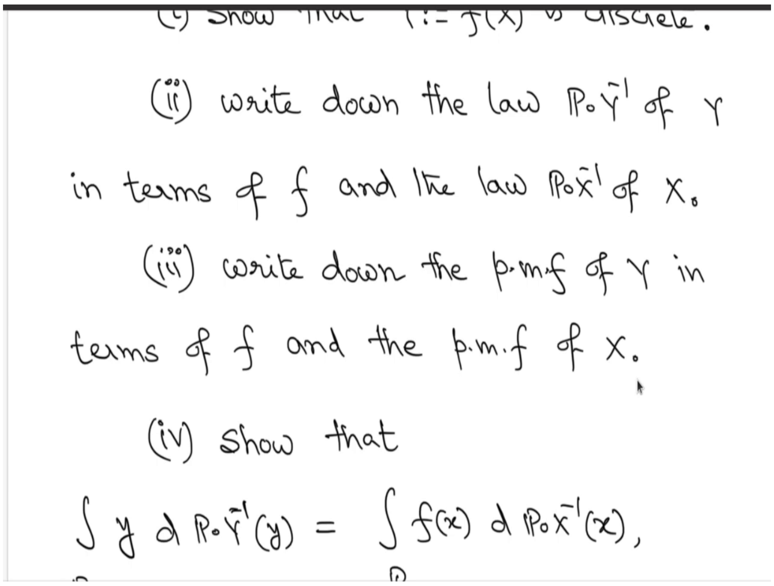
Exercise ⑤: Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be a discrete RV and let $f : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be measurable

Exercise ⑤: Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be a discrete RV and let $f : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be measurable.

(i) Show that $Y := f(X)$ is discrete.

(ii) Write down the law $\mathbb{P} \circ Y^{-1}$ of Y in terms of f and the law $\mathbb{P} \circ X^{-1}$ of X.

(iii) Write down the p.m.f of Y in



So, with this in hand, once you know how to compute these expectations, you can now make some interesting comments. So, this is these are listed in this exercise. So, what do you do? You start by looking at discrete random variable and let f be measurable. So, now look at the composition of f with X, so, that is nothing but $f(X)$. So, you can immediately claim that $f(X)$ is a random variable, call it Y. Now, what the first thing that you need to show is that $f(X)$ which we are now calling as Y is a discrete random variable. So, that is the first statement.

Now, once you have a discrete random variable, you can now try to look at its law. So, first thing that you will do is that you can try to rewrite the law $\mathbb{P} \circ Y^{-1}$ in terms of the function f and the law $\mathbb{P} \circ X^{-1}(x)$. So, use these two quantities, the function f and the law of X to write down the law of Y, so, that is a second expression.

The third thing is that once you have managed to identify the law, you can also try to write down the probability mass function of Y, since Y is given to be discrete, you can talk about the probability mass function. So, therefore, the property mass function of Y, you can try to write it down in terms of the function f and the PMF of X. So, that is good.

(Refer Slide Time: 31:32)

(iv) Show that

$$\int_{\mathbb{R}} y \, d\, \mathbb{P}\circ Y^{-1}(y) = \int_{\mathbb{R}} f(x) \, d\, \mathbb{P}\circ X^{-1}(x),$$

provided one of the integrals is defined.

(v) Using (iv), justify the following.

$$\mathbb{E}Y = \sum_{y \in S_Y} y \, \mathbb{P}\circ Y^{-1}(\{y\}) = \sum_{x \in S_X} f(x) \, \mathbb{P}\circ X^{-1}(\{x\})$$

And the interesting comment here is that you can now talk about the expectation of the random variable y and as for our understanding, you have to do integration of the identity function with respect to the law of y which is $\mathbb{P} \circ Y^{-1}$ and the integration has to take place over the real line.

But now, we are saying this integration if it exists, it will be equal to an integration over the law of x $\mathbb{P} \circ X^{-1}$ and this integration is still over the real line, but you are now supposed to integrate the measurable function f. So, this is another way to compute $\mathbf{E}\, Y$that is through the law of X. So, provided one of the integrals exists the other one will also exist. So, this is again application of the change of variable formula, try to work this out.

But then using this observation, you can now justify the following for this special case that we are discussing. So, here everything is discrete, so, $\mathbf{E}\, Y$ is nothing but this linear combination over the support of Y. So, S_Y denote the support of Y but then you have this series expression, which can be rewritten in terms of the summation over the support of the original random variable X.

So, what do you do here, you look at $f(x)$ and weight it with or factor it with these jumps sizes. So, if you combine these things together, you will get back $\mathbf{E}\, Y$. So, this is simply another familiar expression to you, that will allow you to compute expectations of functions of random variables. So, all you have to do is to choose the right measureable function f and just do these computations.

So, again here we are doing the summation over S_Y on the left hand side here and here we are doing the summation over the support of the original random variable X on the right hand side.

(Refer Slide Time: 33:38)

side makes sense. Here, S_X and S_Y denote the supports of X and Y respectively.

Note (29): As mentioned in Note (6), by choosing the appropriate functions f, we consider the moments $\mathbb{E}(X-c)^n$ for RVs X. Compute these moments for the cases when X is a degenerate,

choosing the appropriate functions f, we consider the moments $\mathbb{E}(X-c)^n$ for RVs X. Compute these moments for the cases when X is a degenerate,
Bernoulli, Binomial or Poisson RV.

And as mentioned earlier, you can choose this appropriate functions f to talk about moments of the random variable X. So, you can talk about this n-th moments of X about any point c. So, all you have to do is to choose this function f to be $(x - c)^n$ defined for any real numbers x. So, if you choose such appropriate functions, you can combine or compose with X and you can get these moments.

So, you can now try to compute these kind of moments for the cases when X is some known standard discrete random variable. For example, you can choose X to be a degenerate random variable or from Bernoulli binomial or Poisson random variables. So, in all of these cases, try to work out what is the expression for the moments following this measure theoretic discussion.

So, try to work out the values and you will expect that you will get back your familiar expressions that you have already seen in your basic probability theory. So, this completes the discussion for computation of expectations for discrete random variables. And we have now seen these as an application of the measure theoretic integration.

For the other type of random variables, including the absolutely continuous random variables, we have to develop this measure theoretic structure a little more, but then we would like to make similar comments about absolutely continuous random variables and their expectations and moments. This we are going to see in later lectures. We shall continue this discussion. In the next week. We stop here.