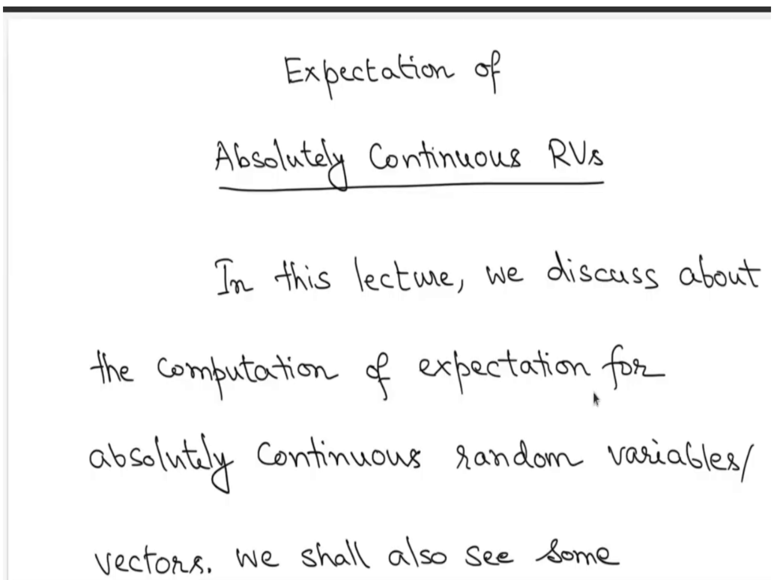
Measure Theoretic Probability 1
Prof. Suprio Bhar
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Lecture 38
Expectation of Absolutely Continous RVs

Welcome to this lecture, before proceeding forward let us first quickly recall what we have done in this week. In this week, we have concentrated on looking at the structural properties of measures. In particular, we have seen these two very important results called the Lebesgue decomposition theorem and the Radon-Nikodyne theorem. Using these results, we have obtained important decomposition results for general probability measures on the real line together with the Borel σ-field.

Using that decomposition, we have also talked about the corresponding decompositions about the distribution functions, and using these decompositions, we have talked about the corresponding classes of random variables or random vectors. In particular, we have concentrated our attention on these absolutely continuous random variables, which correspond to the random variables with law being absolutely continuous with respect to the underlying Lebesgue measure.

With these absolutely continuous random variables at hand, we are now moving on to their properties and analysis involving these random variables. We are going to look at expectations for these random variables. Let us now move on to the slides for this lecture.

(Refer Slide Time: 01:49)



vectors. We shall also see some inequalities involving the moments of general RVs.

<u>Note</u> (17): Given an absolutely continuous RV $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and a Borel measurable function $g : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,

measurable function $g : (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$,

the RV $g(X) : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is not necessarily absolutely continuous. For example, taking g to be a constant function, we get $g(X)$ is a constant/degenerate RV.

degenerate RV.

To ensure that $g(X)$ is absolutely continuous, we need additional regularity of g. We recall a result, usually discussed in basic probability courses. We state this without proof.

In this lecture we are going to discuss about the computation of expectation for these absolutely continuous random variables that we have discussed, but then we will also comment about the higher dimensional versions of this, so we will also consider absolutely continuous random vectors. We are going to see some inequalities involving the moments after we discuss these expectations.

We have earlier discussed about expectation for general random variables then for the special case of discrete random variables, now we finish the discussion about expectations for absolutely continuous random variables and putting everything together we are going to look at certain inequalities that are applicable to all these situations.

Let us first start with the notations, so given some absolutely continuous random variable call it X defined on some probability space $(\Omega, \mathcal{F}, P)$ and a Borel measurable function g which is taking values on the real line and defined on the real line, so if you take these things X and g then remember g composition with X so that is $g(X)$ that is a random variable again defined on the probability space.

However, this is not necessarily absolutely continuous. For an example, take g to be a constant function, so then $g(X)$ is just a constant random variable you can treat it as a degenerate random variable and this is not absolutely continuous. Therefore, you have to be careful with this choice of g. You may have seen these kind of results certain sufficient conditions on the function g which will allow you to claim that $g(X)$ is also absolutely continuous.

These sufficient conditions are usually in terms of regularity of this function g all, so let us recall this result. However, please note that we are going to state this without proof assuming that you have already seen this result in your basic probability theory courses, and also, we are not going to state this version in higher dimensional case again this result can be generalized to the higher dimensions.

(Refer Slide Time: 04:11)

<u>Theorem ③</u>: Continue with the notations of Note ⑰. Let $g : \mathbb{R} \to \mathbb{R}$ be differentiable with $g'(x) > 0$ $\forall x \in \mathbb{R}$. Then the RV $Y = g(X)$ is also an absolutely continuous RV with p.d.f f_Y given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \end{cases}$$

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) & \text{if } y \in (g(-\infty), g(\infty)) \\ 0, & \text{otherwise}, \end{cases}$$

where f_X is the p.d.f of X, $g(\infty) := \lim_{x \to \infty} g(x)$ and $g(-\infty) := \lim_{x \to -\infty} g(x)$.

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

<u>Computation of expectation for absolutely continuous random variables/vectors:</u>

let $X : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ be an absolutely continuous random vector

Let us see this result, so we continue with this setup that we have our absolutely continuous random variable X and we have a nice measurable function. Let us now look at what are these conditions we impose on this function g. Suppose you take g to be differentiable in particular notice that g must be continuous and hence it is Borel measurable, so this is a special type of a measurable function.

Now, in addition you also assume that g' is strictly positive for all points that will ensure that your function g is one-to-one and strictly increasing. In this case what you claim is that the random variable $Y = g(X)$ is also an absolutely continuous random variable and you can in fact write down its PDF given by some such formula.

Remember if the function g is strictly increasing and one-to-one, you can consider the inverse function and that is the inverse function that appears in this formula, and then here just to note there are certain range of values for this random variable capital y which is given as this range of values. This is g of minus infinity to g of infinity. So, what are these values?

These values are nothing but $\lim_{X \to \infty} g$ and limit of the function $\lim_{X \to -\infty} g$. Here f_X denotes the usual notation that is the PDF of the given random variable X. With respect to all of these terms you can write down the PDF of this random variable y which is $g(X)$. There are these other sufficient conditions for example you can take $g'(X)$ to be negative for all X and you will get a similar result for this.

For a higher dimensional case the result usually involves this Jacobians of these functions g, let us not go into to that detail but this is a standard result that you may have seen in your basic probability courses and we are going to assume this. Since this is just for the record that we can figure out certain sufficient conditions under which $g(X)$ becomes absolutely continuous. But now our main interest is in the computation of expectation for absolutely continuous random variables or random vectors. Here we are doing the setup in higher dimensional at one go.

(Refer Slide Time: 06:43)

an absolutely Continuous random vector

with p.d.f f_X. By definition,

$$\frac{d\, \mathbb{P}\circ X^{-1}}{d\, \lambda^{(d)}} = f_X,$$

where $\lambda^{(d)}$ is the Lebesgue measure on $\mathbb{R}^d$.

By Exercise ④(i),

where $\lambda^{(d)}$ is the Lebesgue measure on $\mathbb{R}^d$.

By Exercise ④(i),

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

$$= \int_{\mathbb{R}^d} x\, f_X(x)\, d\lambda^{(d)}(x) = \int_{\mathbb{R}^d} x\, f_X(x)\, dx,$$

provided one of the integrals exist. Note provided one of the integrals exist. Note that "x" and "$\mathbb{E}X$" in the above Computation is $\mathbb{R}^d$-valued and the equality can be interpreted Component-wise. Another way to interpret the above equalities is through the integration of $\mathbb{R}^d$-valued

Suppose you start off with a random vector X taking values in $\mathbb{R}^d$, here d could be anything 1, 2, 3, 4 any such dimensions, suppose you write the PDF as a subscript X then what is the definition? By definition the Radon-Nikodyne derivative of the law of the random variable which is $\mathbb{P} \circ X^{-1}$ that has this density function with respect to the Lebesgue measure in this dimension.

That is the idea that you take the Lebesgue measure on $\mathbb{R}^d$ and look at this Radon-Nikodyne derivative of the law that will give you the density function and that is your probability density function. So, this result we have already seen, but we have also identified that we can transfer integrations with respect to $\mathbb{P} \circ X^{-1}$ into an integration involving the Lebesgue measure just by multiplying by this function.

This was mentioned in exercise 4 part 1. Then what do you do? You apply this exercise and observe this set of equalities. Expected value of the random vector X as defined it is the value of the random vector integrated against the probability measure. And here if you do the change of variable you will end up with this structure that it is the integration over $\mathbb{R}^d$.

Here $x \in \mathbb{R}^d$, so we will make a clarification about this d dimensional integration in a minute, but formally let us accept this equality is true. This is just a change of variable formula or change of measure formula that we have been using so far, but now if you change measures from $\mathbb{P} \circ X^{-1}$ to the Lebesgue measure by using that Radon-Nikodyne derivative you can bring in the density function here, that is what we have done, that is what this exercise suggested.

Then what you do is that you observe that this is the usual integration with respect to the Lebesgue measure and using the standard notational terminology that we have introduced this is just for notational convenience that we will simply write it as usual dx, so that is what it is. Now these equalities will hold provided one of the integrations exist but now we have to clarify this issue about integration in $\mathbb{R}^d$.

Here this x or $X(\omega)$ all of these are taking values in $\mathbb{R}^d$ and we have only explained about integration of extended real value functions or real value functions. Here these terms x and $E(X)$ such terms are now $\mathbb{R}^d$ valued. So let us make a clarification about this. What do we mean by these equalities is that the equality is interpreted component wise.

What do I mean? You take i −th component of X that is what you put it in so this will be X^i let us say that denotes the ith component of X and if you do the change of variables appropriately you will go to here. That is the idea so this equality holds component wise so that is the idea, but again you can also do this change of variable in the $\mathbb{R}^d$ dimensional setup. That will give you this formula.

Again, if you define the integration of $\mathbb{R}^d$ valued functions exactly in the way we have defined the real valued integrations then we will get the similar results, so again one way is to define it component wise, another way is to define the integration directly in $\mathbb{R}^d$ and do this in change of variable formula in this d dimensional setup, either way works. That is this meaning of these equalities provided one of these integrations exists.

(Refer Slide Time: 10:25)

through the integration of $\mathbb{R}^d$-valued measurable functions with respect to measures on $\mathbb{R}^d$. This can be defined in a manner similar to the discussion for $\mathbb{R}$-valued functions in week 6.

Note (18): Continue with the notations of Note (17) and write $Y = g \circ X = g(X)$.

Then, show that (Exercise)

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

provided one of the integrals is defined.

By the above discussion,

By the above discussion,

$$\mathbb{E}Y = \mathbb{E}g(X) = \int_{\mathbb{R}} g(x) f_X(x)\, dx,$$

provided the integral exists. Note that the expression is valid for any Borel measurable g and the RV Y is not necessarily absolutely continuous. By provided the integral exists. Note that the expression is valid for any Borel measurable g and the RV Y is not necessarily absolutely continuous. By choosing appropriate functions g, we consider the moments $\mathbb{E}(X-c)^n$ for X.

Observe that the above expressions for

Now, what do you do after you interpret this you can do a usual analysis involving your expectation and do computations. Continue with these notations that X is a absolutely continuous random variable and g is a Borel measurable function defined on R and take $Y = g(X)$. Then what you can try to show this is a usual result that we have discussed earlier for general random variables and also for discrete random variables and this is again true for absolutely continuous random variables.

That $E(Y)$ is given as this expression provided the integrations exist, but then this Y is nothing but the $E(g(X))$ because Y is $g(X)$ itself and then $g(X)$ you can compute in terms of the law of X. This is the integration that appears here, so you have gone from integrations with respect to the law of Y to the integrations involving the law of X, so that is the meaning of this equality.

Provided one of the integrations is defined if it exists you can make sense of these equalities, but then look at this last expression. Here the law of X appears and then if you know that X for you is absolutely continuous just put in the probability density function there so you will get this expression the familiar expression for the expectation of $g(X)$, where you are multiplying the function g by the probability density function and then taking the integration.

Again, in all of this we have to first ensure that the integrations exist but it is important to note that in these computations we are not assuming any structural properties of the function g other than that it is Borel measurable. So, this is valid for any Borel measurable function. In this case remember the random variable $Y = g(X)$ need not be absolutely continuous even then you get these expressions for $E(Y)$ which you can compute in terms of the probability density function of X or the law of X. So this is the power of measure theory.

(Refer Slide Time: 12:47)

necessarily absolutely continuous. By choosing appropriate functions g, we consider the moments $\mathbb{E}(X-c)^n$ for X.

Observe that the above expressions for expectation matches with those discussed in basic probability courses.

Exercise 5: Compute the moments of X

in basic probability courses.

Exercise 5: Compute the moments of X
when $X \sim \text{Uniform}(0,1)$, $\text{Exp}(\alpha)$ or $N(\mu,\sigma^2)$.

Note 19: We may now repeat the usual analysis done in basic probability courses involving variance and courses involving variance and

covariance etc.

Exercise 6: Check that $\mu = EX$ and $\sigma^2 = \text{Variance}(X) = \mathbb{E}(X-\mu)^2$ when $X \sim N(\mu, \sigma^2)$.

Note 20: Following the discussion in Note 26 of Week 7, we look at the

Then by choosing appropriate functions g you can talk about the moments for these absolutely continuous random variables X. So, these are these moments of the random variable X about the point C. Now what do you observe so far is that this above expressions for the expectations is matching with whatever you have seen in your basic probability theory.

What we have done in an earlier week when you have been discussing about these discrete random variables, we have got the usual expression for the expectation of discrete random variables, now we are talking about absolutely continuous random variables again we have got the familiar expression, but all of these can be derived from the general theory of measure theoretic integration.

So, that is the power of measure theory that all of these special cases for discrete random variables or absolutely continuous random variables are appearing as special cases but the general theory is allowing you to do all of these computations in pretty general setup. You can redo these exercises from your basic probability theory that for standard random variables you can try to compute the moments.

This is left as an exercise, **but then what you have** done in your basic probability theory is that you have also looked at all of these properties involving variances, covariances and other things and you can now repeat these same arguments as soon as you have achieved the same quantities. We have gone to these same moments but we have now been able to define them in a more concrete fashion than what you have seen in your basic probability theory.

But now you can repeat all these usual analysis as you have seen in your basic probability theory, so you are not repeating that part but you can already fit it together with these measure theoretic arguments. For these normal random variables, we had mentioned this parameters μ and σ squared so you can now identify it with the standard identification that μ becomes $E(X)$ and the σ^2 is the variance of X, which is nothing but this second moment of X about this mean.

(Refer Slide Time: 15:03)

Note 20: Following the discussion in Note 26 of Week 7, we look at the characteristic functions of absolutely continuous RVs. For such RVs X,

$$\phi_X(u) = \mathbb{E}\, e^{iuX}, \quad u \in \mathbb{R}$$

$$= \int \cos(ux)\, d\mathbb{P}\circ X^{-1}(x)$$

characteristic functions of absolutely continuous RVs. For such RVs X,

$$\phi_X(u) = \mathbb{E}\, e^{iuX}, \quad u \in \mathbb{R}$$

$$= \int_{\mathbb{R}} \cos(ux)\, d\mathbb{P}\circ X^{-1}(x)$$

$$+ i \int_{\mathbb{R}} \sin(ux)\, d\mathbb{P}\circ X^{-1}(x)$$

$$= \int_{\mathbb{R}} \cos(ux)\, d\mathbb{P}\circ X^{-1}(x)$$

$$+ i \int_{\mathbb{R}} \sin(ux)\, d\mathbb{P}\circ X^{-1}(x)$$

$$= \int_{\mathbb{R}} e^{iux} f_X(x)\, dx.$$

when X is an $\mathbb{R}^d$-valued random vector,

when X is an $\mathbb{R}^d$-valued random vector,

the characteristic function is defined in the following way: for $u \in \mathbb{R}^d$,

$$\phi_X(u) = \mathbb{E}\, e^{iu.X}$$

$$= \mathbb{E} \exp\Big(\sum_{j=1}^{d} i\, u_j X_j\Big)$$

Recall this discussion from note 26 of week 7, so there we had mentioned some things about characteristic functions, in particular we had mentioned these characteristic functions for general random variables and then in specifically we have mentioned that we can do these computations for discrete random variables, but then for absolutely continuous random variables from our earlier understanding once more, you can now go and obtain some nice expressions for the characteristic functions.

If you have an absolutely continuous random variable X look at this characteristic function which is this $E\, e^{iu.X}$. Here u is a real number, so u varies and it is giving you this function, but now if you split this real part and the imaginary part you get these cosines and sines and that you integrate against the $\mathbb{P} \circ X^{-1}$ so that is the usual definition for any random variable X.

If you know that your random variable X is absolutely continuous you put in the density function, and if you put in the density function and club these integrations together you get back this familiar expression that you have to integrate $e^{iu.X}$ multiplied by the density function. What we are using here is that you are changing the usual integration in terms of the law of X to the integration involving the Lebesgue measure and we have used the standard notational terminology that we have introduced in this course.

We are not writing down this Lebesgue measure explicitly but it is understood that this is interpreted as Lebesgue integration. Of course, if you can do the integration in the Riemann sense then you can get the same value, maybe the Riemann integrations will be easier to compute so then you do that and take it as the value for the Lebesgue integration. Here it is the powerful use of this Radon-Nikodyne theorem that is allowing you to go from this law of X to the integration involving Lebesgue measure.

This is for the one-dimensional setup that when X is a real valued random variable but now what happens if you consider any $\mathbb{R}^d$ valued random vector, here again you can go via the standard definitions that you may have seen in your basic probability theory. Now what you have to do is to interpret all the things accordingly in the measure theoretic setup.

(Refer Slide Time: 17:46)

$\mathbb{R}$

when X is an $\mathbb{R}^d$- valued random vector, the characteristic function is defined in the following way: for $u \in \mathbb{R}^d$,

$$\phi_X(u) = \mathbb{E}\, e^{iu.X}$$

$$= \mathbb{E} \exp\left(\sum_{j=1}^{d} i\, u_j X_j\right)$$

$$= \mathbb{E} \exp\left(\sum_{j=1}^{d} i\, u_j X_j\right)$$

$$= \int_{\Omega} e^{iu.X(\omega)}\, d\mathbb{P}(\omega)$$

$$= \int_{\mathbb{R}^d} e^{iu.x}\, d\mathbb{P}\circ X^{-1}(x).$$

As mentioned in Note 26 of Week 7,

Ω

$$= \int_{\mathbb{R}^d} e^{iu.x} \, d\mathbb{P}\circ X^{-1}(x).$$

As mentioned in Note 26 of Week 7, ϕ_X uniquely determines the law of X.

Note 21: Recall from Note 25 of week 3 that $X = (X_1, \dots, X_d)^t$ is a random vector

Here for any $u \in \mathbb{R}^d$ what do you do, you look at $\phi_X(u) = e^{iu.X}$

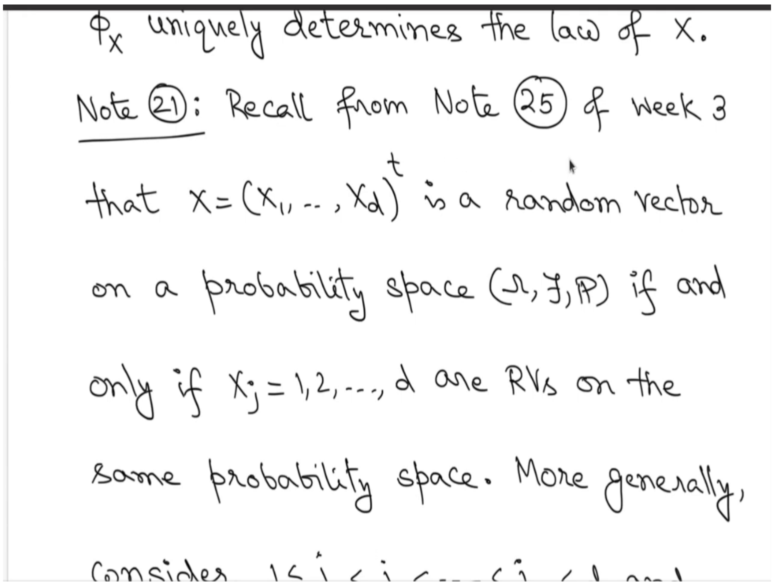
the value is given as the $e^{i.uX}$, so $u.X$ is the usual dot product which is given as this i times the sum of these component wise products. If u_j's are the components of u, X_j's are the components of X then you look at this componentwise products and add them up and this complex number i is just in front acting as a scalar, but now as per our understanding these integration is nothing but the integration with respect to the underlying probability measure and if you change variables you just go from this integration to an integration involving the law.

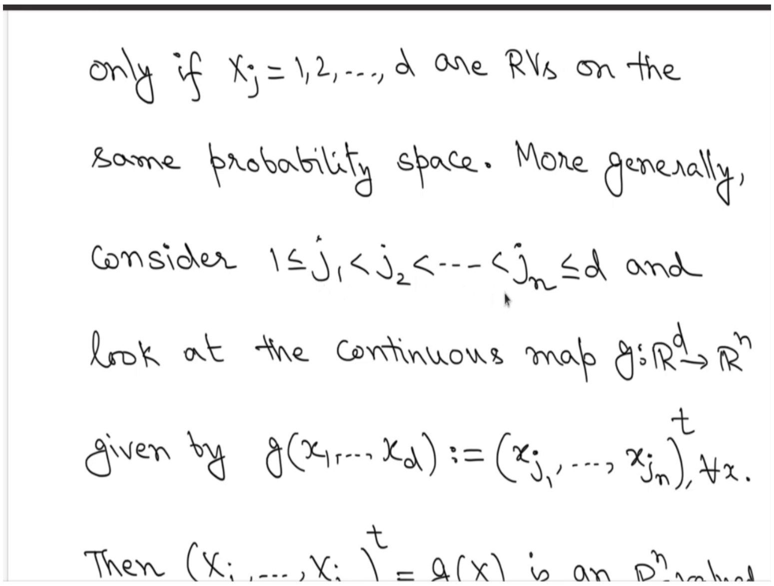
This is integration with respect to the law of X, again you can go back to the usual ideas that you can replace this law for an absolutely continuous random variable and just bring in the density function and write it as Lebesgue integration. That is the idea great so this here again you can do this usual computation for characteristic functions in d dimensional setup.

For discrete random variables what do you do? You write this $\mathbb{P} \circ X^{-1}$ as this convex linear combinations of Dirac masses and that is it you do the usual analysis and get this finite sum or a countable sum accordingly.

We again recall an important comment as mentioned in note 26 of week 7 that this characteristic function uniquely determines the law of this random vector X, so in week 7 we had mentioned this fact when we were talking about one dimensional case, but now we are saying that this is just the generalization to higher dimensions and here again if you know the characteristic function then the law of X which is a random vector now is also uniquely determined.

(Refer Slide Time: 19:51)





Consider $1 \le j_1 < j_2 < \cdots < j_n \le d$ and look at the continuous map $g : \mathbb{R}^d \to \mathbb{R}^n$ given by $g(x_1, \ldots, x_d) := (x_{j_1}, \ldots, x_{j_n})^t, \forall x$.

Then $(X_{j_1}, \ldots, X_{j_n})^t = g(X)$ is an $\mathbb{R}^n$-valued random vector.

Exercise 7: Continue with the notation

So, let us move forward, you can recall from note 25 of week 3 that you can identify the components of random vectors as random variables. Let us put in the notational terminology as appropriately so suppose X is a random vector with these components X_1 up to X_d then what we have said at the time is that this is a random vector defined on this probability space if and only if the components are also random variables on the same probability space.

Then what we have said is that if you are given this one dimensional random variables you can put them together and obtain the random vector. This was the identification through this coordinate wise projection maps and compositions of the random vector X, so this was the idea that was used to prove this connection. More generally what you can now do is that you can choose n such coordinates call them $j_1, j_2, \ldots, j_n$.

We are putting it in the increasing order so these are all distinct. What do you do? You that you choose these values $j_1, j_2, \ldots, j_n$ within 1 to d, and then what you can do is that you can look at this continuous map which is from $\mathbb{R}^d$ to $\mathbb{R}^n$ which maps any vector small $x_1, \ldots, x_d$ to a n vector with these components. You ignore all the other components of the original vector x and just write down the components corresponding to this $j_1, j_2, \ldots, j_n$.

Just write them down that will give you this map from $\mathbb{R}^d$ to $\mathbb{R}^n$, but observe that this is a nice continuous map and hence this must be Borel measurable. Therefore, as a composition with the random vector X it will now give you this random vector which is now n dimensional. So, from the d dimensional random vector by this appropriate projection maps you have now obtained n valued random vector.

(Refer Slide Time: 22:05)

Exercise ⑦: Continue with the notations of Note (21). Show that $(X_{j_1}, \ldots, X_{j_n})^t$ is discrete if $(X_1, \ldots, X_d)^t$ is discrete. Can you make a similar statement for absolutely continuous random vectors?

Note (22): Continue with the notations of Note (21). As mentioned above, the law/distributions of $(X_{j_1}, \ldots, X_{j_n})^t$ for $1 \leq n < d$ with $1 \leq j_1 < \cdots < j_n \leq d$ can be obtained from the law/distribution of

$1 \le n < d$ with $1 \le j_1 < \cdots < j_n \le d$ can be obtained from the law/distribution of $X = (X_1, \ldots, X_d)^t$. These n-dimensional distributions are referred to as the n-dimensional marginal distributions of the d-dimensional random vector X.

n-dimensional marginal distributions of the d-dimensional random vector X.

Given X, the marginal distributions are uniquely determined. However, the converse is not true.

Exercise (8): Find an example showing

If you continue with these notations that we have just discussed you can show is that this random vector n dimensional random vector must be discrete provided the original random vector was discrete. If you start with a discrete random vector, apply appropriate functions on this, you will get a discrete random vector once more.

So, that is the idea that is being used here, but here is a question can you make a similar statement for absolutely continuous random vectors? So, if you take X to be absolutely continuous random vectors can you make a similar statement, try to check this. Now we continue with the same notation, so we are still interested in these random vectors and this individual collection of components which again give you some lower dimensional random vectors.

As mentioned above you can compute this law or distributions of these n dimensional random vectors from the given random vector X you can do that. So, this is just a appropriate change of variable formula, you apply that function j and by this appropriate transformations you can obtain this law for this n dimensional random vector, but these n dimensional distributions that you obtain are usually referred to as the n dimensional marginal distributions of the original d dimensional random vector X.

Just to repeat we are saying that you choose these n many components out of the original d many components and look at the law or distribution of these components random vectors so this is the joint distribution of these n many random vectors. You can obtain this law or distributions and for each such combination of n many components you get one such law or distribution.

These collections of n dimensional distributions we shall refer to as the n dimensional marginal distributions of X. It is important to note that given this X you can immediately obtain the marginal distributions; however, the converse is not true. If you are given all these marginal distributions and you are asked to figure out the law of X it is not uniquely determined, so this converse is not true.

(Refer Slide Time: 24:42)

are uniquely determined. However, the converse is not true.

<u>Exercise (8):</u> Find an example showing that the marginal distributions do not uniquely determine the distribution of a random vector.

What do you do? You should try to find an example which shows this that given these marginal distributions you can possibly get two different random vectors with different laws such that the marginal distributions remain the same, but the laws are different. That is the idea, so please try to figure out such an example.

With this analysis we have managed to wrap up all the ideas that you have seen in your basic probability theory, and have connected the computations of discrete random variables and computations for absolutely continuous random variables in terms of the expectations and moments to the usual formulas.

We have used this measure theoretic structures but we have obtained these formulas that you have seen in your basic probability theory. We also have seen this connection with Riemann integrations and Lebesgue integrations, so that also came up as a special case of this measure theoretic integration with this we can understand the power of measure theory that can cover all of these distinct situations under the same umbrella.

Therefore, you can apply all these general results that you may have obtained in your measure theoretic setup and individually they will give you results in these special cases. So, we are going to see much more analysis about these moments and variances and covariances of these random variables and vectors in the next lecture. This is what we are going to discuss in the next lecture, we stop here.