Measure Theoretic Probability 1
Professor Suprio Bhar
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Lecture 18
Decomposition of Distribution functions

Welcome to this lecture. Before we proceed with the discussions of this lecture, let us first quickly recall what we have been doing in this week. So, in this week, we have looked at the measurable structure of random variables or random vectors and put them together with the domain side probability measure. And when you put them together, we obtain the law or distribution of the corresponding random variable or random vector.

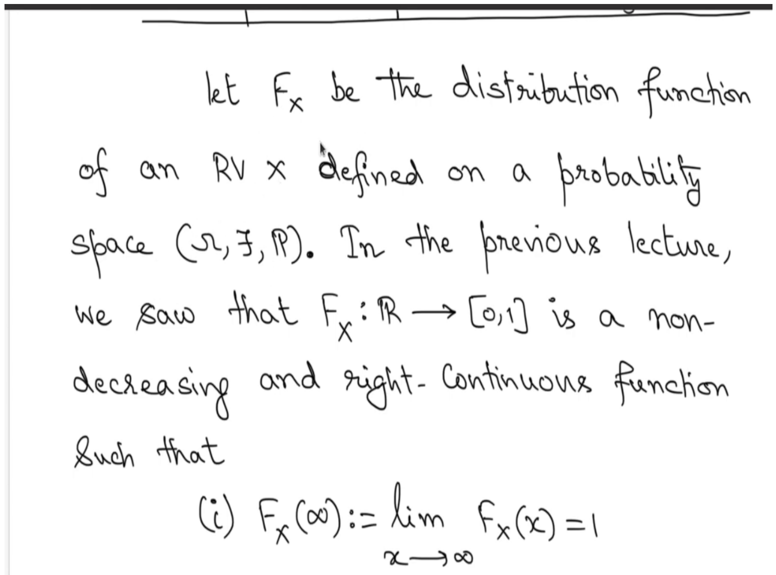
Now, using that law or the distribution, which is a probability measure on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, in the case of random variables and on $\mathbb{R}^d$ with the Borel σ- field of $\mathbb{R}^d$ in the case of $\mathbb{R}^d$ valued vectors, so using that probability measure, we have also defined the corresponding distribution functions.

Now, what we have done in the previous lecture was that we looked at properties of such distribution functions. The thing to note is that we have defined distribution functions, corresponding to probability measures on these Euclidean spaces together with the Borel σ- field.

And then, for random variables or random vectors, we looked at the corresponding law and we said that the distribution function corresponding to the law is the distribution function of the random variable or random vector.

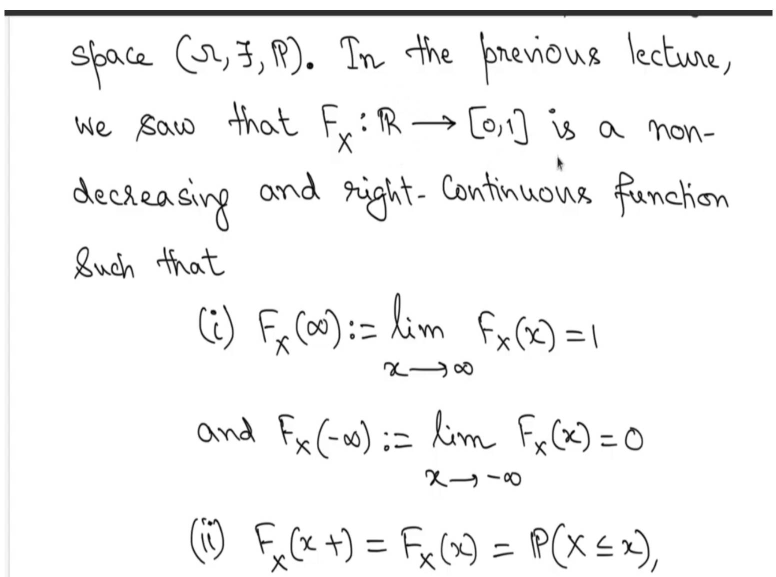
And what we have seen is that the basic properties of probability measures immediately gives you the well-known properties of distribution function as you know about them, as you have studied in your basic probability course. So alright, let us quickly recall these properties that we have already seen in the previous lecture. And we can continue with the discussions of this lecture.

(Refer Slide Time: 02:10)



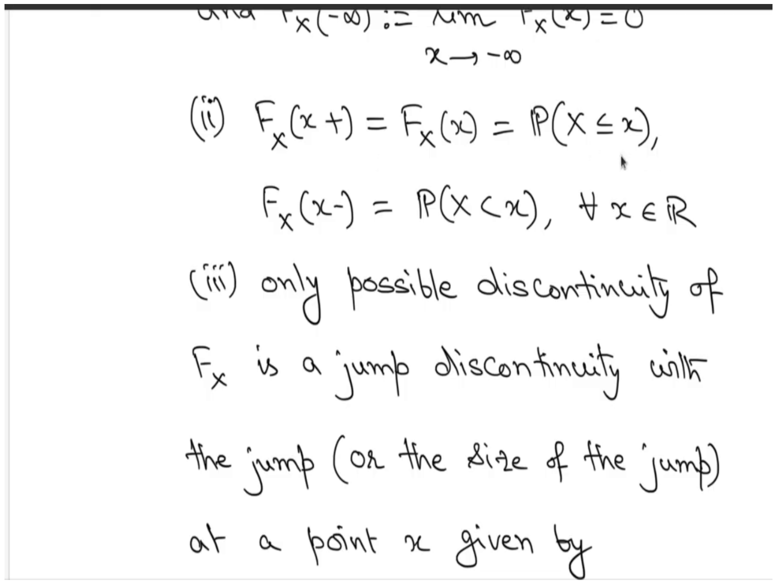
So, start with a random variable X, so just for simplicity, let us work in dimension 1. So, this random variable, it is a real-valued function on this probability space and it is a measurable function. So, I look at its distribution function.

(Refer Slide Time: 02:26)



And what we have seen is that distribution function is the function defined on the real line and it takes values between 0 and 1. It is non-decreasing and right continuous function with some nice properties.

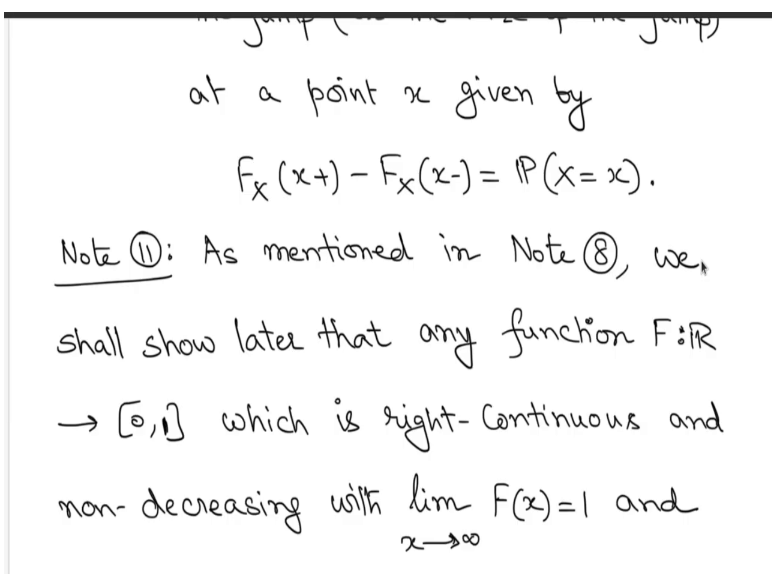
(Refer Slide Time: 02:38)



Like the limit at ∞ being 1, limit at $-\infty$ is 0, the left limit is exactly this, the $\mathbb{P}(X < x)$, and the right limit matches with the actual function value so the function is right continuous and that is exactly $\mathbb{P}(X = x)$. So, here, x_i is any arbitrary real number.

So, once you have identified the left limits and right limits, the existence of those are confirmed and then only possible discontinuity that can happen for the function, for the distribution function is the jump discontinuity.

(Refer Slide Time: 03:18)

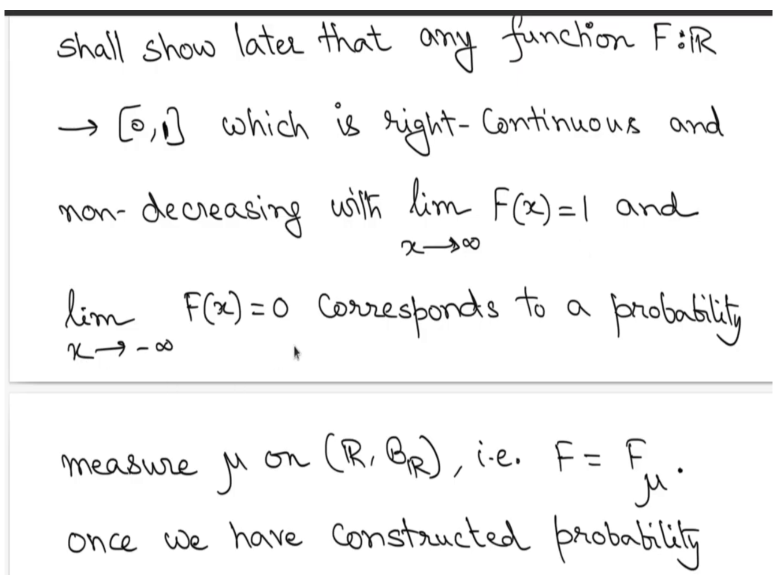


And then you can also compute the size of the jump, and that is exactly given by the difference of these values, and that is computed as $\mathbb{P}(X = x)$. So, these are the properties we have seen.

And, what we have also remarked in Note 8 earlier is that given probability measures, we can now transform it into a function. So, from the collection of all probability measures on the real line with the Borel σ-field, we can obtained this function, which is non decreasing right continuous, limit at ∞ is 1, limit at $-\infty$ is 0.

So, you have obtained the association from the collection of probability measures of that type of collection to the collections of functions of that type. So, you have obtained the association. Now, what do you do later on is to obtain function going in the opposite direction. So, we would like to associate or make an association going in the opposite direction.

(Refer Slide Time: 04:18)



And what it is going to show is that given any function on the real line with values between 0 and 1, which is right continuous, non-decreasing, limit at ∞ is 1 and limit at $-\infty$ is 0, so with these properties if you get a function, then it will correspond to a probability measure μ on the real line together with the Borel σ-field.

(Refer Slide Time: 04:36)

$\lim_{x \to -\infty} F(x) = 0$ Corresponds to a probability

measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, i.e. $F = F_\mu$.
once we have constructed probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and an RV $X : \Omega \to \mathbb{R}$ with law $\mathbb{P} \circ X^{-1} = \mu$, we can extend the above identification to $F = F_\mu = F_{\mathbb{P} \circ X^{-1}} = F_X$,

And then, for that, what will happen is that that given function will not correspond to the distribution function of that probability measure which you have just identified. So, these we will do later on.

(Refer Slide Time: 04:49)

measure μ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, i.e. $F = F_\mu$.
once we have constructed probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and an RV $X : \Omega \to \mathbb{R}$ with law $\mathbb{P} \circ X^{-1} = \mu$, we can extend the above identification to $F = F_\mu = F_{\mathbb{P} \circ X^{-1}} = F_X$,
i.e., any function F satisfying the required properties is the distribution function of some RV. We are going to

Now, recall that we have also mentioned this connection that if you start with the probability space and you are given the random variable, then you can construct the law. But we have also claimed that we shall construct an association going in the opposite direction.

So, once you have completed that justification, that once you have constructed probability spaces and the random variable such that $\mathbb{P} \circ X^{-1}$ is equal to μ for a given probability measure, then we can actually extend this identification, with this general collection of functions.

So, what we will do? So, we will first look at a general function with these nice properties, you first identify it as a distribution function of that probability measure and once you have identified that probability space and the random variable, μ becomes $\mathbb{P} \circ X^{-1}$ and that is what you have identified as the distribution function of that random variable. So, we will do this connection once we have done the association going in the opposite direction. So, this is what we are supposed to do.

(Refer Slide Time: 05:55)

with law $\mathbb{P} \circ X^{-1} = \mu$, we can extend the above identification to $F = F_\mu = F_{\mathbb{P} \circ X^{-1}} = F_X$, i.e., any function F satisfying the required properties is the distribution function of some RV. We are going to assume this fact.

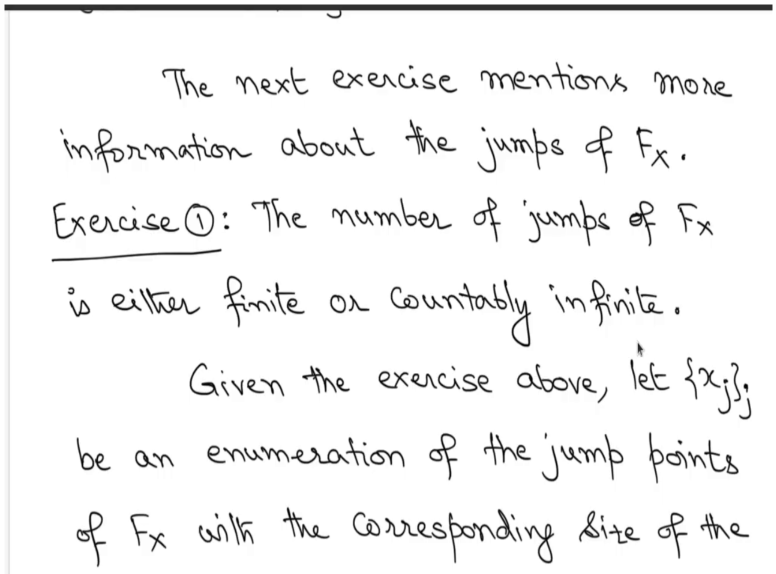
The next exercise mentions more information about the jumps of F_X.

But for the purpose of this discussion, what we are going to assume is that this thing can be done that this identification can be proved. So, then, any function, satisfying these required properties.

By that I mean, that it is a function defined on the real line, takes values between 0 and 1, is non-decreasing right continuous, limit at ∞ is 1, limit at $-\infty$ is 0, then for such a function, you can construct such probability measures and such random variables.

Therefore, such a function will become the distribution function of some random variable. So, we, we are going to assume this fact in this lecture. So, these are the basic facts about distribution function that we have already discussed.

(Refer Slide Time: 06:38)

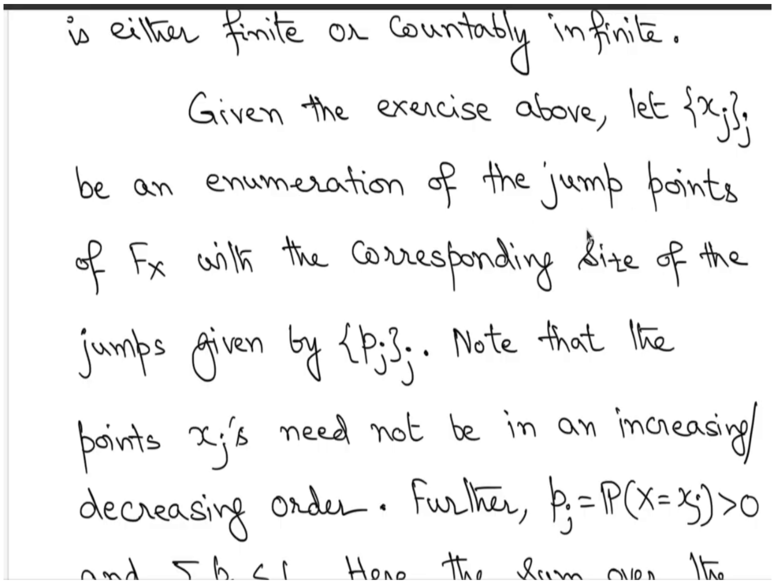


But now, we are going talk about other nice properties about the distribution function. But in the previous lecture, we have already mentioned that we are getting back the usual properties of the distribution function from the identification through the probability measures.

So, whatever properties you have already studied in your basic probability course about distribution functions, you are getting back and you are actually proving them through the properties of the probability measure in this course.

So similarly, you can try out this very well-known property of distribution function which you have seen in your basic probability course that the number of jumps of a distribution function is either finite in number or countable infinite in number. So, please try to work this out.

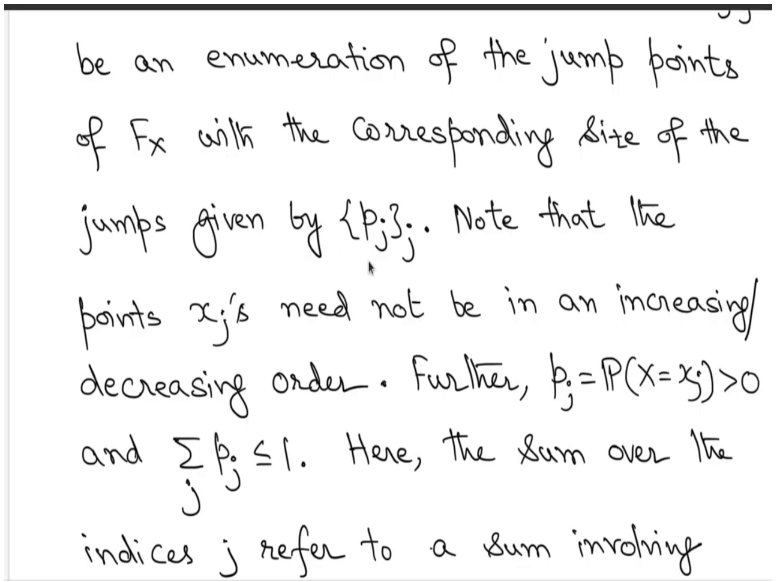
(Refer Slide Time: 07:26)



But then once you assume this fact that there are certain number of jumps of a distribution function, so since their number is finite or countable infinite, you can index the points where the jumps occurred.

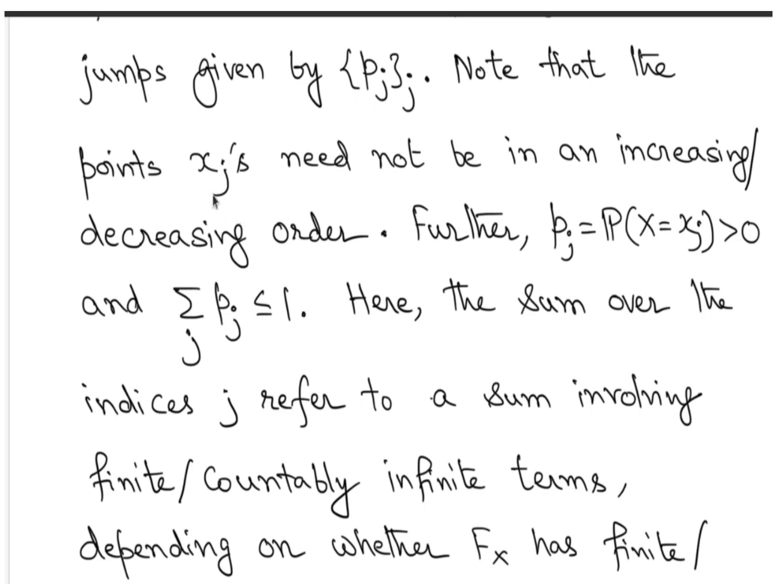
So, let us x_j denotes the enumeration of the jump points. So, you just number them x_1, x_2, x_3, and so on. So, if it is finite, you will stop at a finite stage, if it is countable infinite, you will get a infinite sequence. So, this is the list of jump points wherer the jumps are occurring.

(Refer Slide Time: 07:55)



But then correspondingly, you will also get the jump sizes and let us called them as p_j. So, jump at the point x_j is exactly given by p_j.

(Refer Slide Time: 08:05)



But then, this point x_j, need not be in a increasing or decreasing or decreasing order. So, just be careful with that. It is just a list, it is just an enumeration of the points where the jumps are occurring.

But now these p_j's, as identified earlier p is, these p_j's are nothing but $\mathbb{P}(X = x_j)$. That is the jump size. And that is quantity, that is strictly positive because that is where the jumps are occurring. But then contribution from all the jumps, if you add them together, that will be less or equal to 1. So, this is a standard fact that you can also identify, that you can always prove.

(Refer Slide Time: 08:44)

points x_j's need not be in an increasing/decreasing order. Further, $p_j = \mathbb{P}(X = x_j) > 0$ and $\sum_j p_j \leq 1$. Here, the sum over the indices j refer to a sum involving finite/countably infinite terms, depending on whether F_X has finite/

countably infinite number of jumps.

But here, this sum that you are seeing, this sum is over the indices *j*, and this refers to a sum involving finite number of terms in the case of finite number of jumps of distribution function, otherwise you will get countably infinite number of terms when you are, you are dealing with infinite number of jumps for the distribution function.

So, this is simply a basic fact. Whether the function, distribution function has finite number of jumps or a countably infinite number of jumps, accordingly you will get a summation which is finite or countable infinite. This is a simple fact.

(Refer Slide Time 09:17)

Countably infinite number of jumps.

Note (12): The exercise above is a familiar property of distribution functions, covered in basic Probability courses. We now state more of these familar properties/ Statements from basic probability.

But then, this, familiar property for the distribution functions that you have just mentioned in this exercise, then it will also give you some ideas that the properties coming from the jumps, they should also follow.

So, this is the idea, that once you identify the basic properties of a distribution function, from the properties of probability measure, then you are expecting that all the other well-known properties of distribution functions that you know about should also be proved. So, one of the things is left as exercise there as the properties of the jumps.

(Refer slide Time: 09:58)

Consider the function $F_X^{(d)} : \mathbb{R} \to [0,1]$ defined by

$$F_X^{(d)}(x) := \sum_j p_j \, 1_{[x_j, \infty)}(x), \; \forall x \in \mathbb{R}.$$

Since $1_{[x_j,\infty)}(x) = 1$ if and only if $x_j \leq x$, the sum on the right-hand side above is the sum of all jump-sites for the defined by

$$F_X^{(d)}(x) := \sum_j p_j \, 1_{[x_j,\infty)}(x), \ \forall x \in \mathbb{R}.$$

Since $1_{[x_j,\infty)}(x) = 1$ if and only if $x_j \leq x$,

the sum on the right-hand side above is the sum of all jump-sizes for the jumps occuring in $(-\infty, x]$.

So now, let us look at more of these familiar properties. And we are just recalling them. Look at this function F_X^d. So, what do we do? We define this function on the real line, taking values between 0 and 1 and we look at this quantity on, defined on the right-hand side.

So, for any point x on the real line I define this function as the summation value. So, what is this? So you check, whether the point x is in this interval $[x_j, \infty)$ or not.

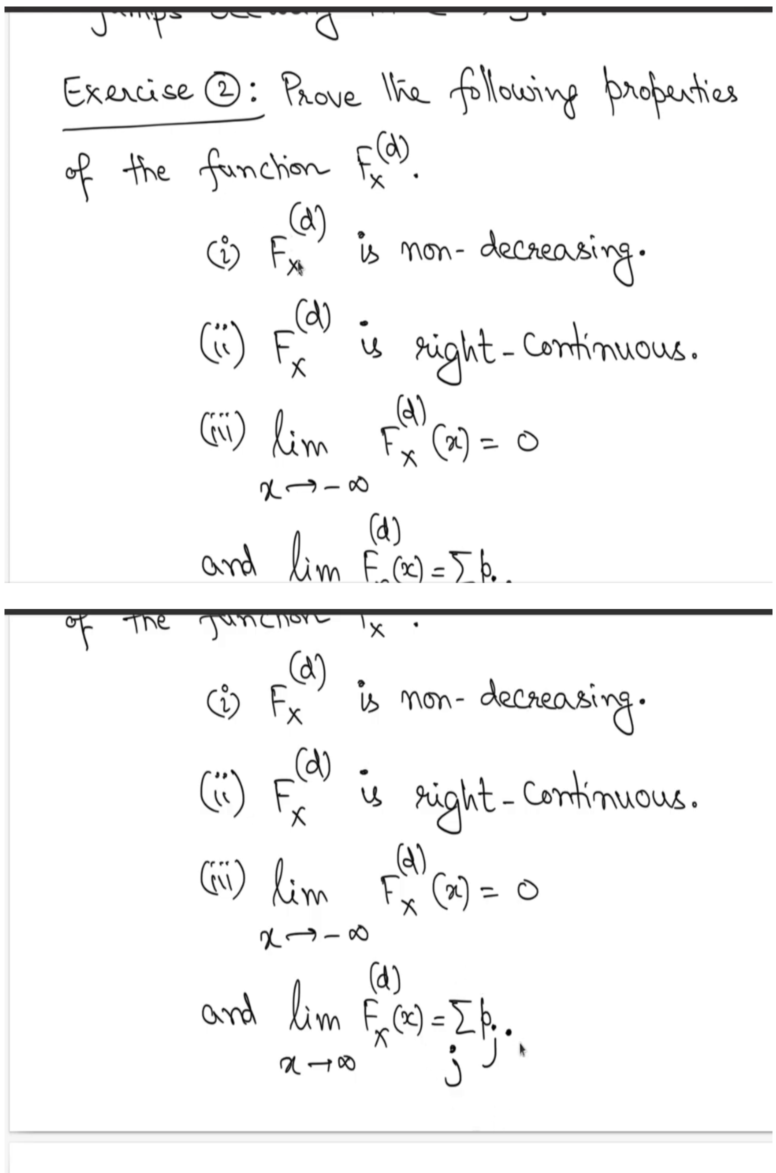
So, if the point x is in this interval, you will get the value 1. But this is, this happens if and only if x_j, the value $x_j \leq x$. So, you are basically saying that you fix point x and look at all the jumps that have occurred before the point x.

So, you have looked at the real line up to and including the point small x and you are just looking at all the possible jumps that have occurred within that location, within that interval $(-\infty, x]$. So, you are just looking at those jump points, and corresponding to those things you are just adding up the jump sizes.

So, that is exactly what you are doing in this summation. So, this sum on the right hand side is exactly the sum of all the jumps that you find in this interval $(-\infty, x]$. So, that is how this function is defined.

So, this function F_X^d simply looks at the number of jumps, first in the interval $(-\infty, x]$ and then you simply add up the corresponding jump sizes. So, there is just a contribution from the jumps. So, that is what you are looking at in F_X^d.

(Refer Slide Time: 11:40)

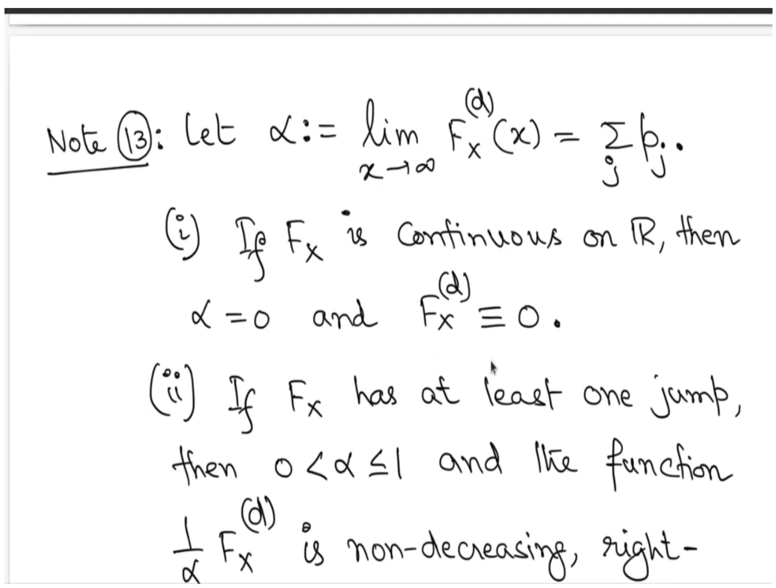


But then, there are some very nice properties about this function that you have just obtained from the contributions from the jumps. So, this function is non-decreasing, it is right continuous, limit at $-\infty$ is 0 and limit at ∞ is exactly the contribution from all the jumps.

So, remember we have mentioned that given a distribution function the contribution from (summa) sum of all these jumps is less or equal to 1. So, this quantity that you get, this quantity is less or equal to 1. So, the limit at ∞ is something less or equal to 1.

So, you now see that the F_X^d which is purely considered from the jumps contribution, this is a non-decreasing function, it is right continuous and you can also identify the limits at ∞ and $-\infty$. So, these are very standard facts that you can prove easily and this you must have seen in your basic probability course. If not, please take this as an exercise. Please work them out.

(Refer Slide Time: 12:43)

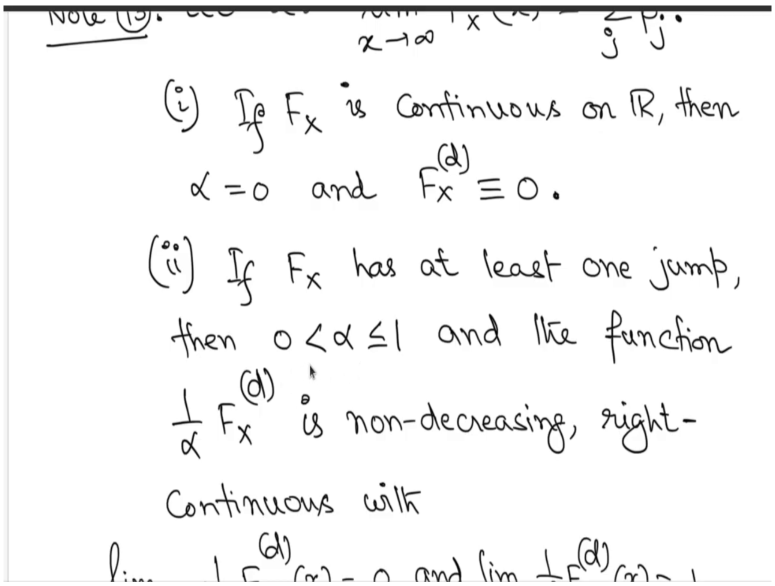


But now, let us focus on that limit at ∞. So, I just look at that summation or the contribution from all the jumps, all the jump sizes. So, now, this quantity, as I have remarked, that this is less or equal to 1. So, call that limit at α. So, it is just a short hand notation that is going to be used now. So, I am just looking at the contribution from all the jumps. So, that, I call it as α.

Now, if F_X is continuous, that means F_X does not have a jump, so there is no contribution from the jumps, then α must be 0. And accordingly, you will also have no contribution from the jumps anywhere. So, F_X^d, that function which you have just considered, will also be identically 0.

So, if the distribution function is continuous throughout the real line, then both these quantities α and the function F_X^{d} must be 0. So, F_X^{d} is a function which becomes identically 0 if the original distribution function is continuous.

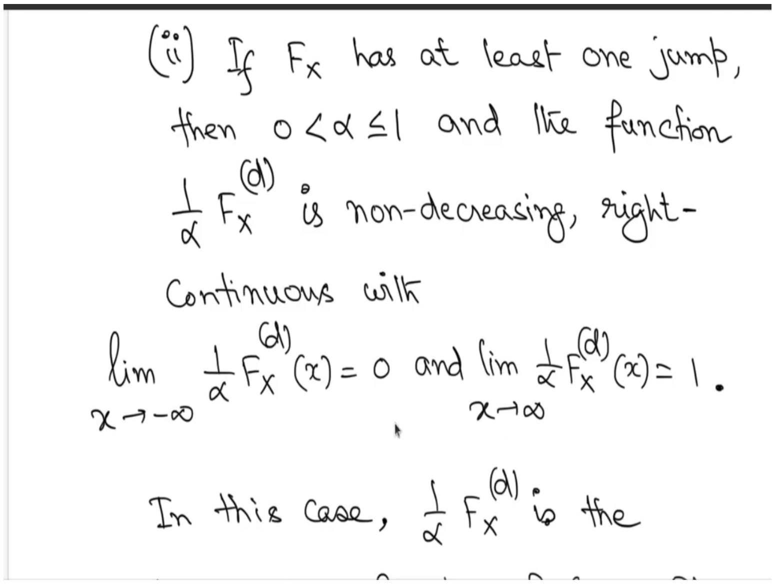
(Refer Slide Time: 13:48)



But then, what happens if F_X has at least one jump. So, you just move away from continuity, allow at least one jump. Then, there will be some contribution coming from the jumps and that will be contained in that quantity α . So now, α will be in $(0,\ 1]$. So, there will be some non-trivial contribution from this at least 1 jump quantity.

And now, what you can choose to look at is this scaling factor. So, α is now strictly positive, so you can now look at $\frac{1}{\alpha}$, divide the actual contribution from the jump, so divide it by α . So, $\frac{1}{\alpha}F_X^{d}$. So, look at that, that function.

(Refer Slide Time: 14:31)



So, that function has these properties following from the properties of F_X^{d}. So, what are these properties? Since α is strictly positive, the $\frac{1}{\alpha} F_X^{d}$, that function will remain a non-decreasing right continuous function with limit at $-\infty$ being 0, limit at ∞ being 1.

So, just to repeat, you are looking at this scaled version of F_X^{d} and it turns out to be non-decreasing right continuous with some nice limits at $-\infty$ and $+\infty$. The limit value being 0 and 1, respectively.

(Refer Slide Time: 15:10)

$$\lim_{x \to -\infty} \frac{1}{\alpha} F_X^{(d)}(x) = 0 \text{ and } \lim_{x \to \infty} \frac{1}{\alpha} F_X^{(d)}(x) = 1.$$

In this case, $\frac{1}{\alpha} F_X^{(d)}$ is the distribution function of some RV.

(See Note 11)

Now, consider the function $F_X^{(c)} : \mathbb{R} \to \mathbb{R}$ defined by $F_X^{(c)}(x) := F_X(x) - F_X^{(d)}(x), \forall x$.

So, therefore, what you can immediately now claim is that $\frac{1}{\alpha} F_X^{d}$, in this case, when there is at least one jump is a genuine distribution function of some random variable. So, here, we are just making that connection from Note 11, which we have mentioned at the start of the discussion.

So, we are saying that if we have a function with all these nice properties, it must be the distribution function of some random variable. Again, just to recall, we have mentioned in Note 11, that we are going to prove this fact later on. But for now, we are going to assume this as a fact.

So, this is the contribution from the jumps and we have identified these things as a distribution function when α is something non-trivial, when $\alpha > 0$. Great.

(Refer Slide Time: 15:56)

(See Note (11))

Now, consider the function $F_X^{(c)} : \mathbb{R} \to \mathbb{R}$ defined by $F_X^{(c)}(x) := F_X(x) - F_X^{(d)}(x), \forall x$.

Exercise (3): Prove the following properties of the function $F_X^{(c)}$.

(i) $0 \leq F_X^{(c)}(x) \leq F_X(x) \leq 1 \ \forall x$

(ii) $F_X^{(c)}$ is non-decreasing

But then, you can now choose to look at the remaining portion that takes away the jump parts. So, you take away all the contribution from the jumps. So, you are looking at the distribution function, given distribution function and subtracting out the contribution from the jumps.

And whatever you get, call it F_X^c. So, that is a now, function, which is difference of two function values. So, therefore, it is a real-valued function taking some real numbers as its possible values.

So, F_X^c is defined as the original function, original distribution function, but you are taking out all the contribution from the jumps. But then again, these will also be some familiar properties to you, so I am just recalling them for the sake of the record. Please work out these exercises, if you have not seen this before.

(Refer Slide Time: 16:45)

defined by $F_X \; (x) := F_X(x) - F_X \; (x), \forall x$.

Exercise ③: Prove the following properties of the function $F_X^{(c)}$.

(i) $0 \leq F_X^{(c)}(x) \leq F_X(x) \leq 1 \;\; \forall x$

(ii) $F_X^{(c)}$ is non-decreasing

(iii) $F_X^{(c)}$ is continuous.

(iv) $\lim_{x \to -\infty} F_X^{(c)}(x) = 0$.

Exercise ③: Prove the following properties of the function $F_X^{(c)}$.

(i) $0 \leq F_X^{(c)}(x) \leq F_X(x) \leq 1 \;\; \forall x$

(ii) $F_X^{(c)}$ is non-decreasing

(iii) $F_X^{(c)}$ is continuous.

(iv) $\lim_{x \to -\infty} F_X^{(c)}(x) = 0$.

So, F_X^c has some nice properties. So, first of all we have mentioned that F_X^c is a function, defined on the real line and takes real values. But then the first property itself says that it actually takes values between 0 and 1, and moreover, the F_X^c is dominated from above by the actual distribution function.

So, in particular, F_X^c, it is a non-negative function and it takes value at most 1. So, that is the first property. So, F_X^c now becomes a function from the real line to $[0, 1]$. Great.

But then, it has also some other nice properties like, it is non-decreasing and it is continuous. So, F_X^c is non-decreasing and it is continuous. And you can also identify its limit at $-\infty$, so, which is 0. So, please try to work out these properties of the function F_X^c.

(Refer Slide Time: 18:03)

Note (14): let $\beta := \lim_{x \to \infty} F_X^{(c)}(x)$. From our discussion above, $0 \le \beta \le 1$. Recall that $\alpha = \lim_{x \to \infty} F_X^{(d)}(x)$ (See Note (13) above).

(i) $\beta = 0$ if and only if $F_X^{(c)} \equiv 0$

if and only if $F_X \equiv F_X^{(d)}$.

(ii) $\alpha = 0$ if and only if $F_X^{(d)} \equiv 0$

if and only if $F_X \equiv F_X^{(c)}$.

discussion above, $0 \le \beta \le 1$. Recall that

$\alpha = \lim_{x \to \infty} F_X^{(d)}(x)$ (See Note (13) above).

(i) $\beta = 0$ if and only if $F_X^{(c)} \equiv 0$

if and only if $F_X \equiv F_X^{(d)}$.

(ii) $\alpha = 0$ if and only if $F_X^{(d)} \equiv 0$

if and only if $F_X \equiv F_X^{(c)}$.

(iii) If $\beta > 0$, then $\frac{1}{\beta} F_X^{(c)}$ is the distribution function of some RV.

But then, you will be wondering about what happens at ∞. So, F_X^c is a non-decreasing function, so the limit at ∞ will exist. So, call that limit at β. So, remember, for the contribution from the jumps, we have looked at F_X^d. There we looked at the limit at ∞, we called it α.

Here, we are looking at the remaining portion, we are taking away all the contribution from the jumps, we have constructed this function F_X^c. It turned out to be non-decreasing. I am now looking at the limit at ∞ and calling it β. But now, since F_X^c is bounded between 0 and 1, this quantity, this limit, this β should be between 0 and 1. So, that is not a problem. So, it is a, follows easy.

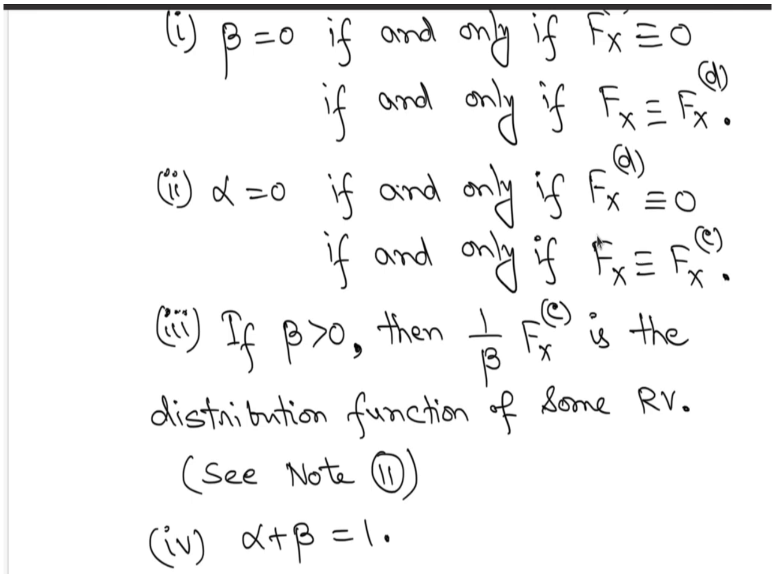
But then, recalling the fact that the contribution from the jumps that F_X^d has a limit at ∞ to be α, so that is basically sum of all the jump sizes, then you can make some nice comments about α and β. When you put α and β together, you can make some very nice comments.

So, if β is 0, then what you are saying is that this limit at ∞ for the F_X^c is 0. But then, this is the, one of the bounds, upper bounds for F_X^c, so therefore, what you will not identify that there should be no contribution from this part.

So, β is 0, if and only if F_X^c is identically 0. And this can only happen if F_X, the original distribution function is identically equal to the contribution from the jumps, so F_X^d.

So, this can only happen if an only if $\beta = 0$. So, that is exactly what the first statement says. So, what basically is said now is that all the contributions are coming from the jump parts. So, that is the first statement.

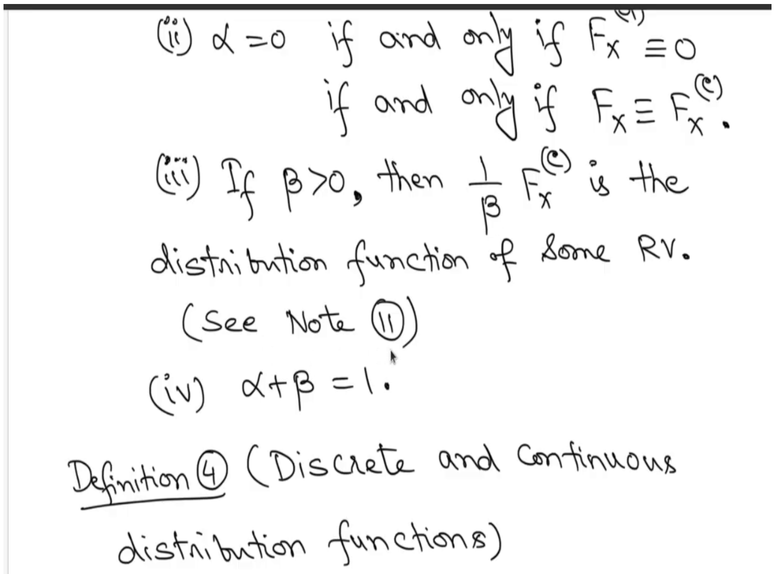
(Refer Slide Time: 20:00)



So, the second statement says, assume $\alpha = 0$. So, take $\alpha = 0$. So, that means there is no contribution from the jump. So, it is exact opposite situation compared to the (())(20:09) one. But this can only happen if and only if, there is no contribution from the jump, so therefore this function F_X^{d} is identically 0.

But then as per definition, F_X^{c} must be the difference $F_X - F_X^{d}$, therefore F_X^{c} under this condition must be exactly be F_X, the original distribution function. So, $\alpha = 0$ if and only if there is no contribution from the jumps and F_X^{c} equals the original distribution function F_X . So, these are two extreme cases, $\alpha = 0$ or $\beta = 0$.

(Refer Slide Time: 20:48)



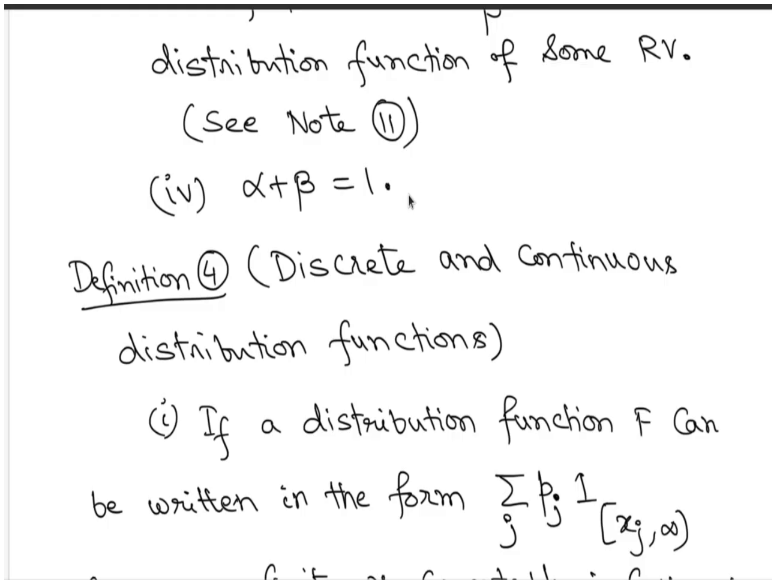
But then, if β has some non-trivial contribution. If $\beta > 0$, then you can now scale this F_X^c by $\frac{1}{\beta}$. So, recall we had done a similar argument involving F_X^d. So, we looked at the limit value α. If $\alpha > 0$, if there was some non-trivial contribution from the jumps, then we scaled it, scaled the function F_X^d by the α and we observed some nice properties.

We do the same here. If β is positive, look at $\frac{1}{\beta}$, the scaled version of F_X^c. I say that this is now a distribution function of some random variable. Why? Again, let us try to check whether F_X^c satisfies all those nice properties.

So, remember F_X^c is taking values between 0 and β. β is the upper bound for F_X^c. So, therefore $\frac{1}{\beta} F_X^c$ takes values between 0 and 1. That is the first point.

Next is, F_X^c is non-decreasing and β is positive so therefore $\frac{1}{\beta} F_X^c$ is still non-decreasing. You can also show by the similar argument that it is also continuous. And you can also identify the limits at $+\infty$ and $-\infty$ as 1 and 0, respectively. So, it satisfies all the required properties and therefore by the discussion in note 11 earlier, this must be the distribution function of some random variable. So, that is the third statement.

(Refer Slide Time 22:24)



distribution function of some RV.

(See Note 11)

(iv) $\alpha + \beta = 1$.

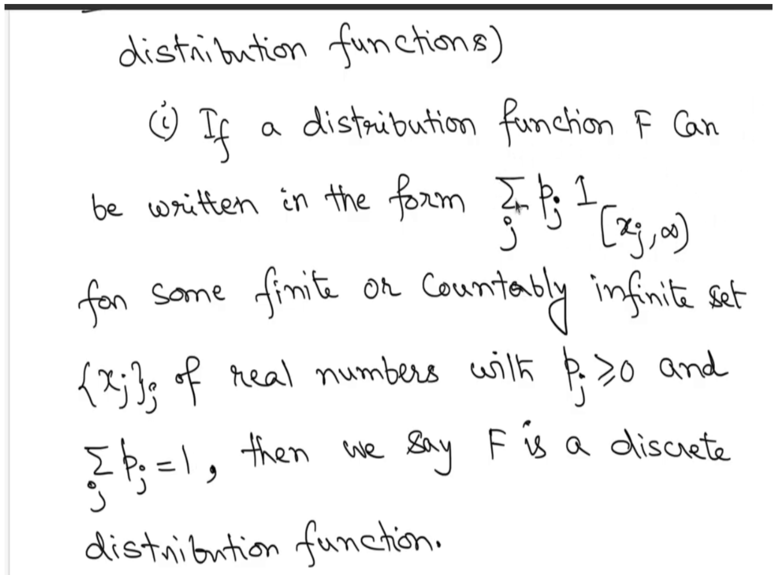
Definition 4 (Discrete and Continuous distribution functions)

(i) If a distribution function F can be written in the form $\sum_j p_j 1_{[x_j, \infty)}$

And then, put α and β together. What you can now try to show is that $\alpha + \beta$ actually equals to 1. So, therefore, α basically looks at contribution from all the jumps, β looks at the remaining part. What you are saying is that put them together, you are get the total quantity is exactly equal to 1. Please try to check this.

So, with that identification in mind, what you are now going to do is to look at certain types of distribution functions. So, what we have observed so far is that the original distribution function F_X has a part coming from the jumps, and there is a remaining part which is continuous. And we have proved some nice properties about these separate parts.

(Refer Slide Time: 23:16)



And with, this gives us some motivation for looking at the Definition 4, and this tells us that we look at certain special type of distribution functions. So, the first thing that we are looking at here is a distribution function which can be written in this form. So, remember this was the form that we have looked at, when we looked at the contribution from the jumps only.

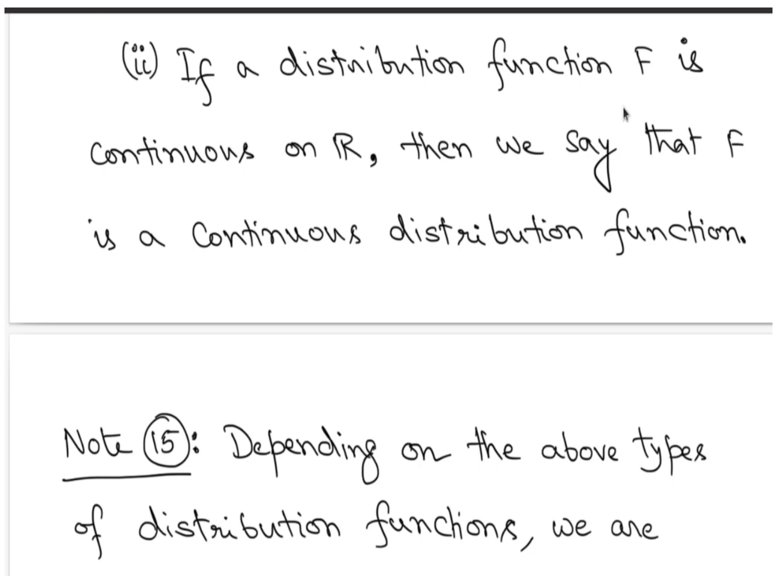
So, you are just looking at all the jumps that occur within points small x and then you just add up the corresponding jump sizes. So, all the jumps that can, have occurred up to that point. So, that is how that jump contribution was considered.

But if a distribution function is purely of this type for some finite or countable infinite set of x_j's, with these p js being non-negative and sum ups to 1, then you are going to say that F is a discrete distribution function. So, it is purely made up of jumps.

So, without loss of generality you can always assume p_j is strictly positive, it is greater than 0, you can ignore the equality case because if $p_j = 0$, then there will be no contribution coming from this summation. So therefore, without loss of generality you can also take p_j to be greater than 0. You can ignore the equality case. But just for simplicity, let us continue with these assumptions.

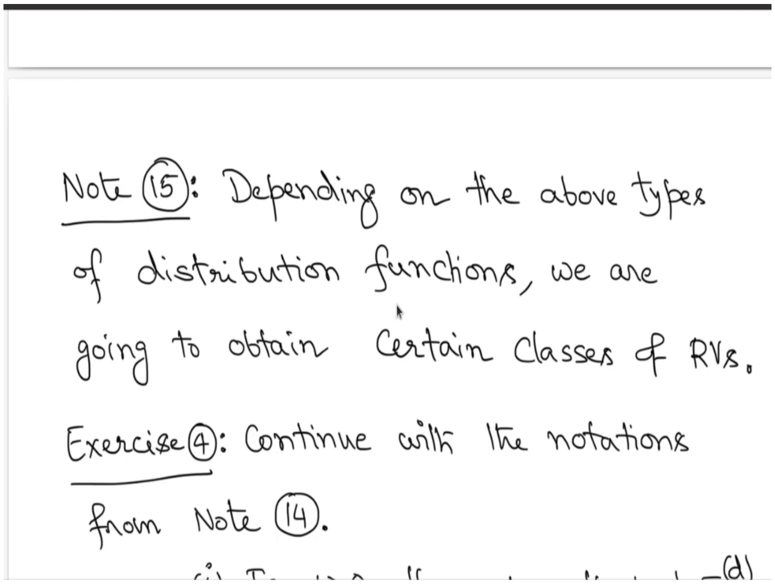
So, in this case we are looking at a discrete distribution function which is defined as the, made up of purely of jumps. So, the total contribution from the jumps must be 1. And the distribution function has exactly this form, made up of a, some (comment) kind of a combination of indicators, for some appropriate finite or countable infinite set of x_js. So that is where exactly the jumps are occurring for that distribution function. So, this is what a discrete distribution function means.

(Refer Slide Time: 25:02)



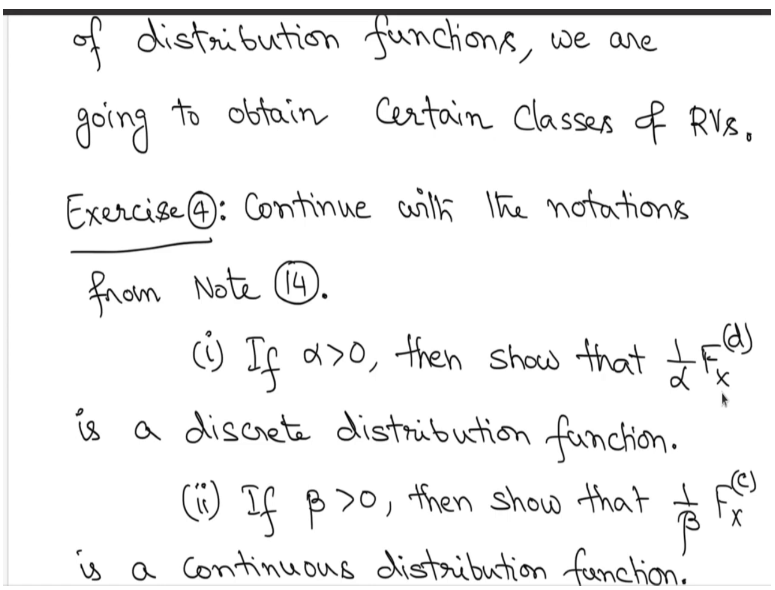
And the exact counterpart for the continuous case is this. So, if you get that the distribution function F is continuous on R, then you say that F is a continuous distribution function. So, there is no contribution from the jumps. There are, there are no jumps. So, you just get a purely continuous function, and this is the case you are going to say that F, the distribution function is a continuous distribution function. So, this is a very simple definition.

(Refer Slide Time: 25:33)



But then, depending on these two types of distribution functions, we are now going to obtain certain nice classes of random variables. So, this is what we are going to discuss later on.

(Refer Slide Time: 25:45)

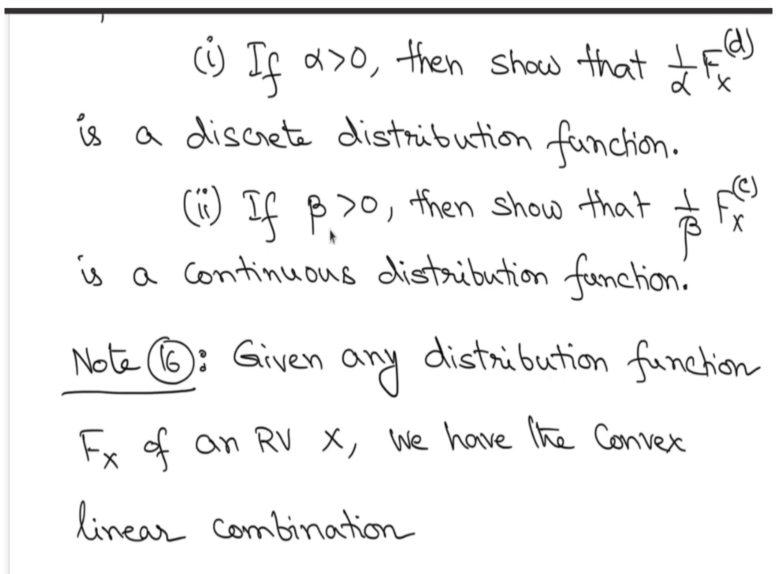


But now, continue with these notations that you have discussed in Note 14. So, just to recall, we have considered the contribution from the jumps, we have defined the F_X^d function. That function has a limit at ∞ which we denoted as α.

Then we got rid of all the contribution from the jumps. We looked at the remaining part and that we called as F_X^c and that was also non-decreasing function. So, we looked at the limit at ∞, called it β. So, remember all those α and β.

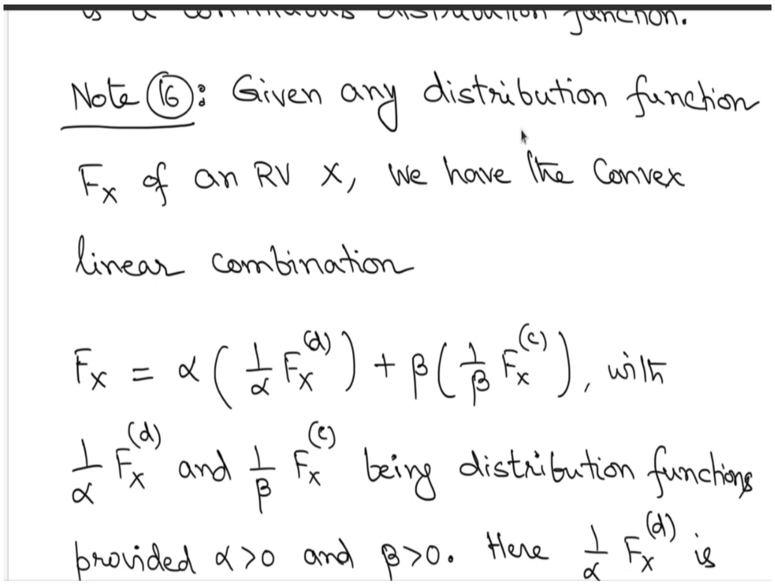
So, remember, if $\alpha = 0$, then there is some non-trivial contribution from jumps. So, then, you can try to show that $\frac{1}{\alpha}F_X^d$ is a discrete distribution function. So, you have already mentioned this fact that $\frac{1}{\alpha}F_X^d$ in this case, when $\alpha > 0$, in this case it is a distribution function. Now, you can try to check that it is exactly of the relevant form and it is a discrete distribution function.

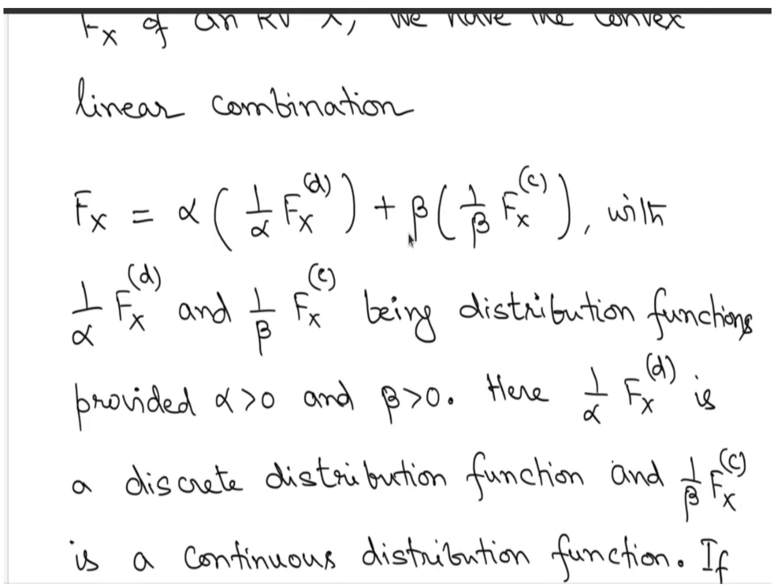
(Refer Slide Time: 26:45)



Similarly, you get a statement for the continuous part. So, you look at $\beta > 0$. So, there is non-trivial contribution outside the jumps. So, there is some contribution coming from continuous parts. So, in this case, you can try to show that $\frac{1}{\beta}F_X^c$, which we already identified as some distribution function, this has a continuous distribution function. So, please try to write down these proofs. It is a very simple argument. You just have to verify the corresponding structures. Great.

(Refer Slide Time: 27:20)





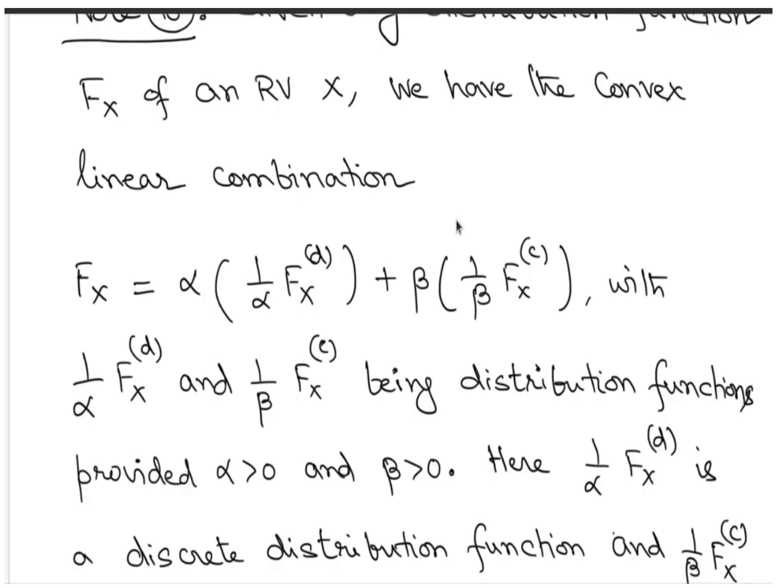
But then, with these identifications at hand, what do you get? You get that given any distribution function F_X, of a random variable X, we have this convex linear combination.

So, remember, in the case when α is positive and β is also positive, consider that case, then $\frac{1}{\alpha}F_X^{\ d}$, this part within the brackets, this is a genuine distribution function and it is a discrete distribution function.

Look at this part $\frac{1}{\beta}F_X^{\ c}$, this is also a genuine distribution function and it is a continuous distribution function. So, the terms, the function within brackets are distribution functions. But they have been scaled according to α and β.

So, these scalars that I have put outside this α and β, they are non negative and they add up to 1. So, remember the, all those properties. So here, on the right hand side, you have a convex linear combination of distribution functions, α and β being the scalars for the convex linear combination.

(Refer Slide Time: 28:24)



But then, what you have basically, you are saying is that given a distribution function, you can split it as a convex linear combination of a discrete distribution function and the continuous distribution function.

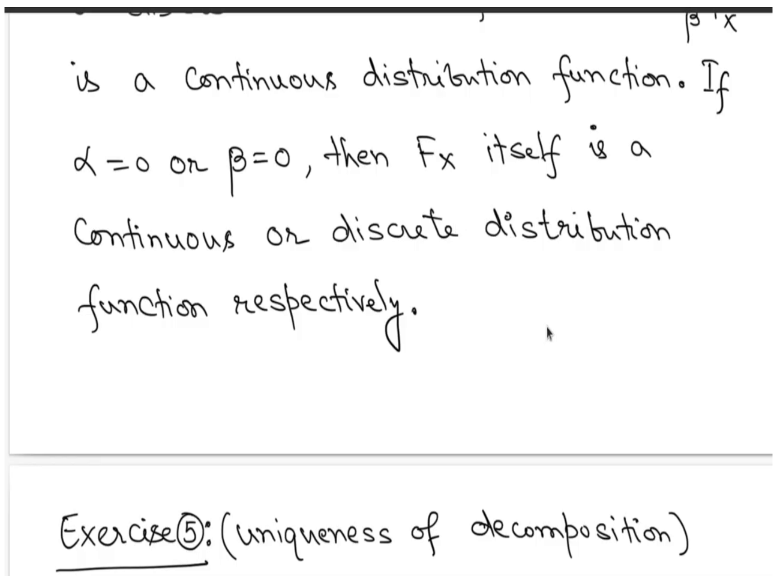
But you will ask, what happens to this convex linear combination if $\alpha = 0$ or $\beta = 0$. Then of course, you cannot divide them by these 0 quantities. So, if $\alpha = 0$, $\frac{1}{\alpha}$ does not make sense.

But then, ignore this first term there. Remember, if $\alpha = 0$, then β must be 1. And then, $F_X = F_X^{\ c}$. So, this equality still makes sense when you ignore the contribution from the jumps as $\alpha = 0$. So, ignore this first term. So, it is just rewriting $F_X = F_X^{\ c} = \beta\left(\frac{1}{\beta}F_X^{\ c}\right)$.

Exact counterpart of this is the case when β is 0. So, then again, ignore this contribution coming from the continuous part. So, there is no contribution in fact, so ignore this term here. So, purely contributions are coming from the jumps.

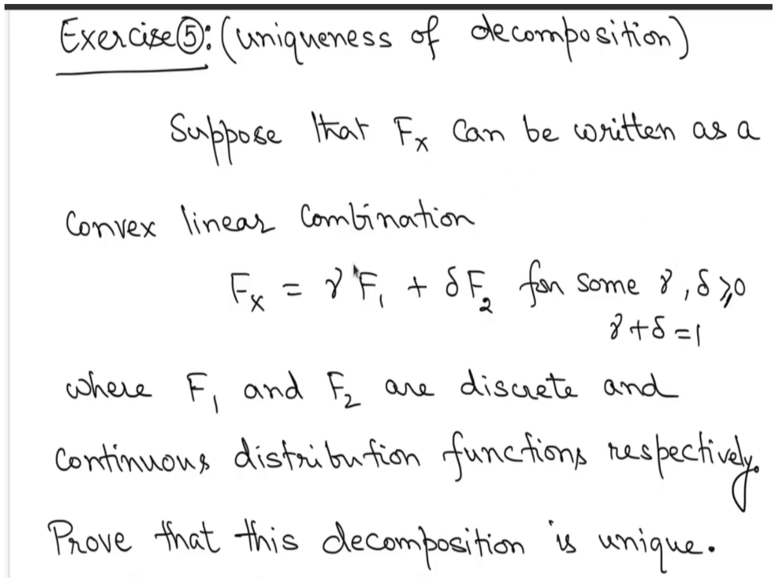
So F_X is exactly equal to $F_X^{\ d}$, but in this case $\alpha = 1$, and then what you are just rewriting this as $F_X = F_X^{\ d} = \alpha\left(\frac{1}{\alpha}F_X^{\ d}\right)$. So, these equalities still make sense as long as you interpret it right.

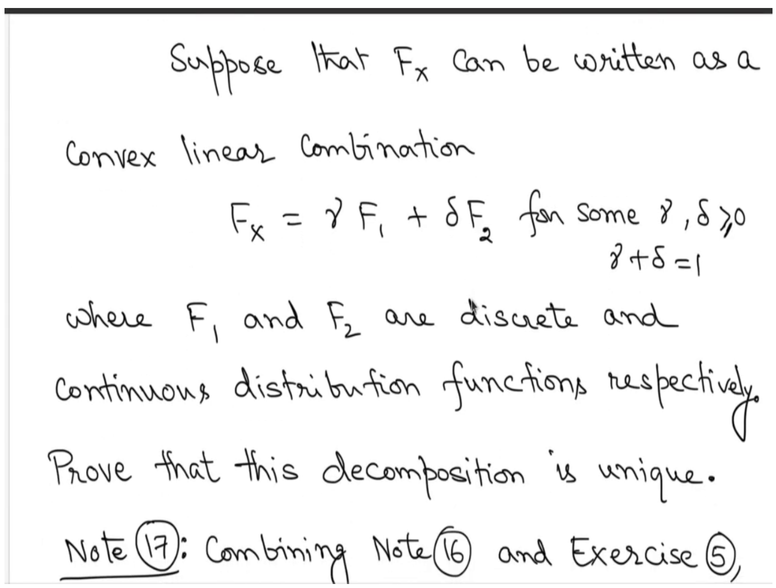
(Refer Slide Time: 29:51)



So, you can still continue to say that F_X itself becomes a convex linear combinations of distribution functions in all of these situations, that $\alpha = 0$, or $\beta = 0$ or α and β, both are positive and adds up to 1.

(Refer Slide Time: 30:10)





But then, you can now ask, you can now ask is the decomposition that we have obtained, is this unique? And the answer to that is yes. So, why? Because F_X, when you are writing it, if you can write it in terms of such a convex linear combination with gamma delta non-negative, adding up to 1, for F_1, which is a discrete distribution function, F_2, which is a continuous distribution function, then you can show that this must be unique.

So, if you can write a general distribution function as a convex linear combination over discrete and continuous distribution functions, then this identification is unique. So, that is the exercise. Please try to prove this.

In particular, the identification that we have done in the previous page, splitting it in terms of the appropriate contributions coming from the jumps, the $F_X^{\ d}$ and the $F_X^{\ c}$, that was one example of a convex linear combination but this exercise is saying that such a combination must be unique.

(Refer Slide Time: 31:16)

continuous distribution functions respectively.

Prove that this decomposition is unique.

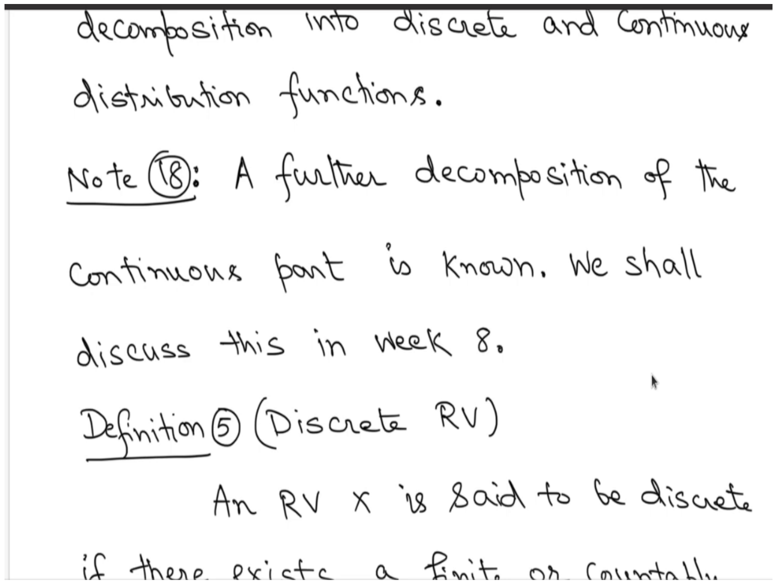
Note (17): Combining Note (16) and Exercise (5), we have the complete description of the decomposition into discrete and continuous distribution functions.

Note (18): A further decomposition of the continuous part is known. We shall

So, if you combine these two things together, so one hand, in Note 16, we have the existence of such a combination and Exercise 5 says, such a combination, if it is there, then it must be unique. So, put them together, we have a complete description of the decomposition into discrete and continuous parts. So, given a distribution function, you can split it.

So, either it is purely discrete or purely continuous, or it is the mix of this with a explicit convex linear combination coming from the jump parts and the remaining part, the continuous part.

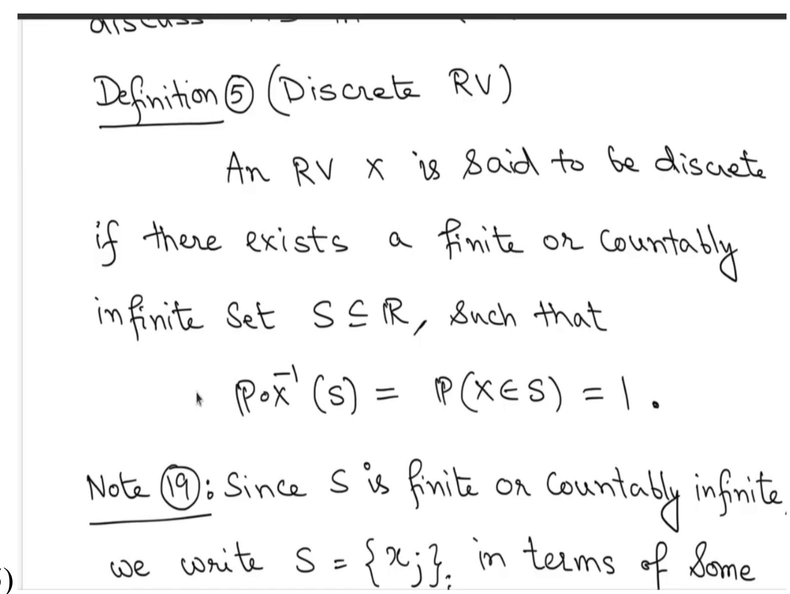
(Refer Slide Time: 31:54)



But, if you look at the continuous parts separately, a further decomposition is known. We are going to discuss that part in Week 8, later on. So, we do not go into this, we will require much more measure theoretic setup before we can discuss about this further decomposition of the continuous part. We will come back to this.

So there, remember, we have looked at the F_X^d function which is purely made up of the jump parts. And with appropriate scaling, we can get it to be a discrete distribution function.

(Refer Slide Time: 32:35)



Now, with that at hand, we are now going to consider certain class of random variables which exactly correspond to these kind of distribution functions. So, what are these? A random variable X is said to be discrete if there exists a finite or a countably infinite set S, which is a subset of the real line, such that $\mathbb{P} \circ X^{-1}$, the law of X associates full marks to this. So, the probability that the random variables takes values in this set is 1.

So, remember, any singleton set is a Borel subset of the real line, so therefore, if you consider finite unions or countable unions of those, you will get any finite or countably infinite set like S as a Borel subset of the real line.

And $\mathbb{P} \circ X^{-1}$ which is defined on $\mathcal{B}_R$, can therefore associate certain values, certain weight, certain size to this set S, and we are saying that, that size must be the full quantity and that is 1. So, if this happens, then you say your random variable X is discrete.

Now, we are just recalling your basic definition of a discrete random variable that you have seen in basic probability courses. But then, we are using this measure theoretic knowledge and putting the definition in terms of the corresponding law. And you will see the advantage of this later on.

(Refer Slide Time: 34:03)

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

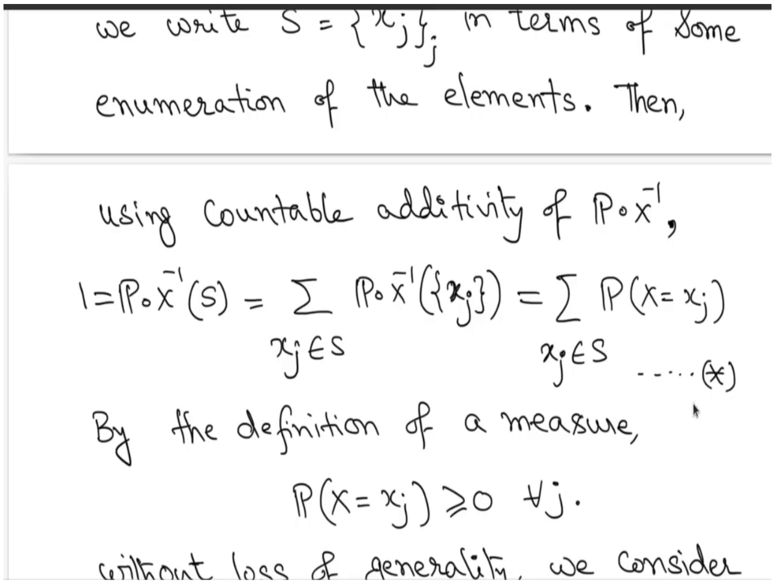
Note 19: Since S is finite or countably infinite, we write $S = \{x_j\}_j$ in terms of some enumeration of the elements. Then,

using countable additivity of $\mathbb{P}\circ X^{-1}$,

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

But then since S is finite, or countably infinite, you can enumerate the elements in the set S. So, if it is finite, you will get x 1, x 2, x n, after, up to a finite number n. And if it is countably infinite, you will get a sequence of points. So, just enumerate them.

(Refer Slide Time: 34:21)



Then, using this countable additivity of the probability measure $\mathbb{P} \circ X^{-1}$, look at this quantity. So, this is as per the definition for discrete random variable, $\mathbb{P} \circ X^{-1}(S) = 1$.

But then the set S is a countable disjoint union of this singletons. So, split it according to the countable additivity of the probability measure $\mathbb{P} \circ X^{-1}$. Then, what you get is this summation over all the points on the set S. Just rewrite this summation using the notations that we already learned about. And it is nothing but the addition of all the probabilities that $x = x_j$.

So, we add up all the things where x_j falls in the set S. So, this is a relation that sum of all these probabilities adds up to 1. So, this is the relation star which we shall recall in a few minutes once more.

But remember, since $\mathbb{P}$ is a probability measure, $\mathbb{P} \circ X^{-1}$ is a probability measure, these quantities, these contributions coming from the jumps here, these must be non-negative. So, you are just looking at addition of certain non-negative quantities and they must be 1. So, that is what this star relation suggests.

(Refer Slide Time: 35:37)

Using countable additivity of $\mathbb{P}\circ X^{-1}$,

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

By the definition of a measure,

$$\mathbb{P}(X = x_j) \geq 0 \quad \forall j.$$

without loss of generality, we consider the set S such that in $(*)$

$$\mathbb{P}(X = x_j) > 0, \quad \forall x_j \in S.$$

But without loss of generality, you can only consider the points x_j, when these quantities are positive, when these quantities are not 0. So, without loss of generality, ignore those terms for which the x_j gives you 0 probability. So, $\mathbb{P}(x = x_j) = 0$, such terms we will ignore. It is not a big deal, they will not contribute to the summation here in the star relation.

(Refer Slide Time: 36:07)

$$\mathbb{P}(X = x_j) \geq 0 \quad \forall j.$$

without loss of generality, we consider the set S such that in $(*)$

$$\mathbb{P}(X = x_j) > 0, \quad \forall x_j \in S.$$

We refer to the set S as the support of X

Note 20: $\mathbb{P}(X = x) = 0 \quad \forall x \in S^c$.

Note 21: If S is finite, then the values

So, without loss of generality, this set S can be considered to be the set of points x_j, where the probability associated is positive, where there is a genuine jump. So, the set S basically is collecting these jump points. So, whenever there is a jump, you just collect it in the set S.

And for the case of this discrete random variables, you get this set S, and we are going to call it as a support of this random variable X. And, just as you expect, whenever you look outside this set S, if your point x is outside, then the probability associated to this must be 0 because there is no jump there. You have listed all possible jumps in the set S.

Just note this that if S is finite, you can now rearrange these x_j's, so that you can get a increasing order for all these points. So, if you have finitely any points x_1 up to x_n, you can easily sort them and put them in a increasing order. So, this is a basic observation. This, you can use it sometimes. But you be careful whenever S is countably infinite. Sometimes it may not be possible to rearrange all these jump points. So, let us ignore that for now.

(Refer Slide Time: 37:25)

for all indices j.

Note 22: (Law of a discrete RV X)

Since $\mathbb{P}\circ X^{-1}(S)=1$, for any $A\in\mathcal{B}_{\mathbb{R}}$,

we have $\mathbb{P}\circ X^{-1}(A)$

$$= \mathbb{P}\circ X^{-1}(A\cap S)$$

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

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

Since $\mathbb{P}\circ X^{-1}(S)=1$, for any $A\in\mathcal{B}_{\mathbb{R}}$,

We have $\mathbb{P}\circ X^{-1}(A)$

$$= \mathbb{P}\circ X^{-1}(A\cap S)$$

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

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

$$= \sum_{x_j\in S} p_j\,\delta_{x_j}(A),$$

So now, let us go back to the question about the law of this discrete random variables. So, now, what we are going to look at is this quantity $\mathbb{P}\circ X^{-1}$ associated to general sets.

So, for any set A on the Borel σ- field, if A is a Borel subset, it is there, so you look at this quantity. But then if $\mathbb{P}\circ X^{-1}$ associates full mass to the set S, then $\mathbb{P}\circ X^{-1}(A)$ is nothing but $\mathbb{P}\circ X^{-1}(A\cap S)$. So, this is a easily verifiable property. So, please check this.

But then, you split this according to the usual countable additivity that we have used a few minutes back. So, S, the set S, the support is made up of all these points, singleton sets where the jumps are occurring. So, you write it as a summation over x_j's. So, that is nothing but probability of X belonging to this set. Great.

(Refer Slide Time: 38:31)

$$= \mathbb{P}\circ X^{-1}(A\cap S)$$

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

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

$$= \sum_{x_j\in S} p_j\,\delta_{x_j}(A),$$

where $p_j = \mathbb{P}\circ X^{-1}(\{x_j\}) = \mathbb{P}(X=x_j)$. Hence,

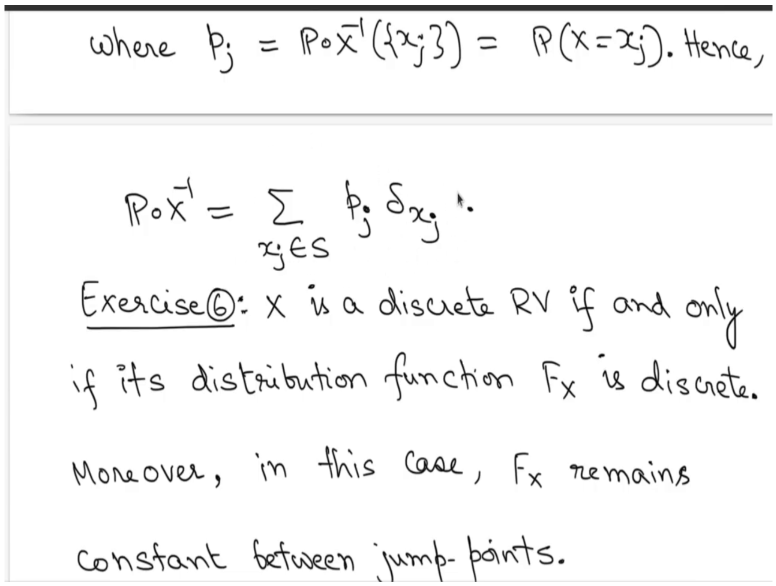
But then, you rewrite it using this notation. So, you just check whether the point x_j is in the set or not. So, if the x_j is not in the set, then this intersection is empty, there will be no contribution. If the point x_j is in the set, then this intersection is exactly this singleton set x_j.

So, in either case, what you can do? You can write it as, this quantity, this contribution coming from the jump, multiplied by the Dirac measure. So, this is what we are just rewriting. We have just observed that in either of the cases when x_j is in the set A or not, you can write it using this notation.

So therefore, you just rewrote this expression in terms of the Dirac measures. But remember, we use these notations p_js for the jump sizes, and just rewrite it using this p_js, so therefore what we have started off with, $\mathbb{P} \circ X^{-1}(A)$ is nothing but this summation $\sum_{x_j \in S} p_j \delta_{x_j}(A)$. So, Dirac measure, then you are considering certain combinations.

But remember, you are considering a discrete random variable. So, these $p_j > 0$, when there, whenever there is a non-trivial jump, you are looking at that. So, that is the jump size, it is non-trivial, positive quantity and they sum up to 1. So, therefore, this quantity, whatever this is, this is a convex linear combinations of the weights coming from the Dirac mass.

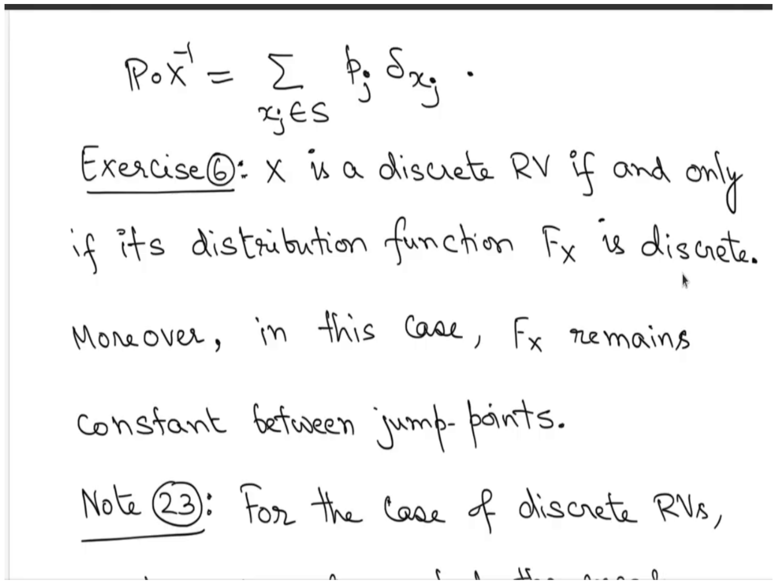
(Refer Slide Time: 39:59)



And hence, what you will immediately obtain is that $\mathbb{P} \circ X^{-1}$ is exactly this convex linear combination of the Dirac masses. Remember, these are the similar things, we have discussed in Week 2. So, when we looked at measures, we said that certain convex combinations of probability measures also gives you probability measures. Great.

So, just to recall, just to repeat, $\mathbb{P} \circ X^{-1}$ of a arbitrary set A can be written as this convex combination where p_j's are exactly the jump sizes and $\delta_{x_j}(A)$ are exactly the weightages coming from the Dirac masses. So, that is why $\mathbb{P} \circ X^{-1}$ exactly turns out to be of this form. So, whenever you get a discrete random variable, you get this quantity. Great.

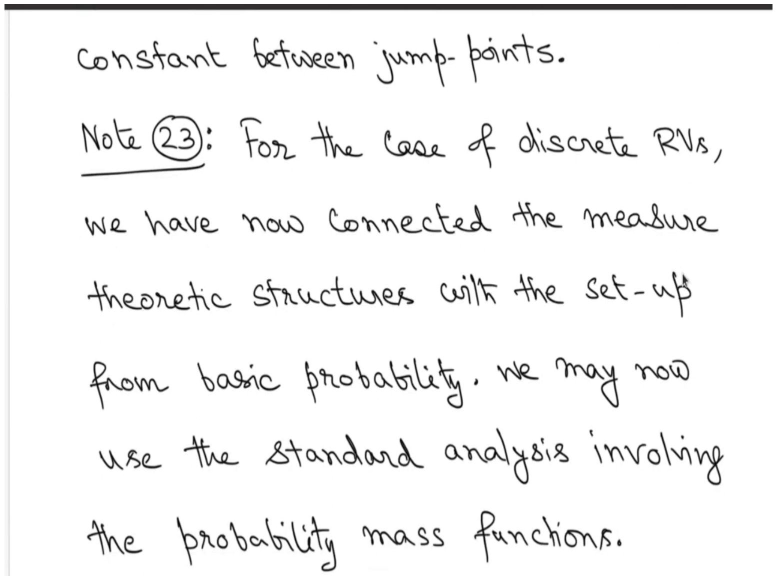
(Refer Slide Time: 40:47)



Now, you can also connect it with discrete distribution functions. So, if X is a discrete random variable, then its distribution function must be discrete. And this is an if and only if condition that if its distribution function is discrete then the original distribution function X must be discrete. This is an if and only if condition.

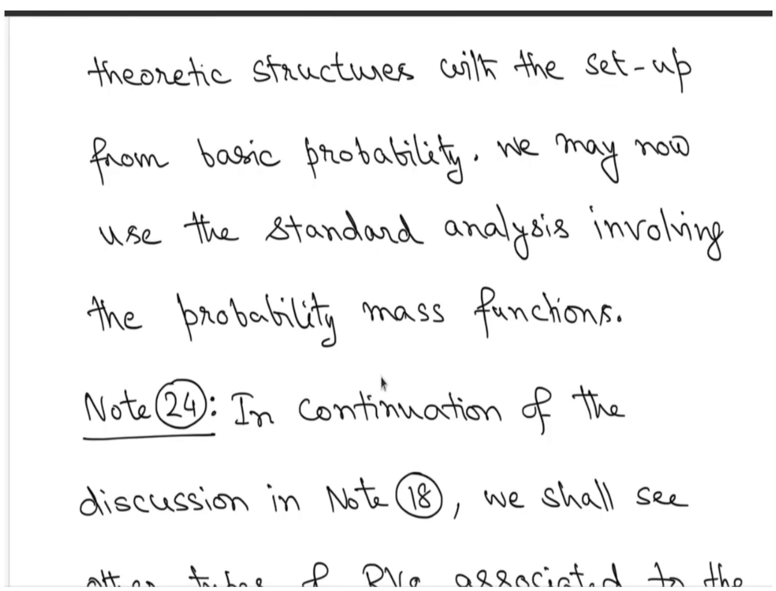
And in this case, what you can see is that between any two jump points, so if there are no jumps in between two such jump points, then F_X will remain a constant. So, if you plot the graph of the function, graph of the distribution function, it will remain flat between the consecutive jump points.

(Refer Slide Time: 41:29)



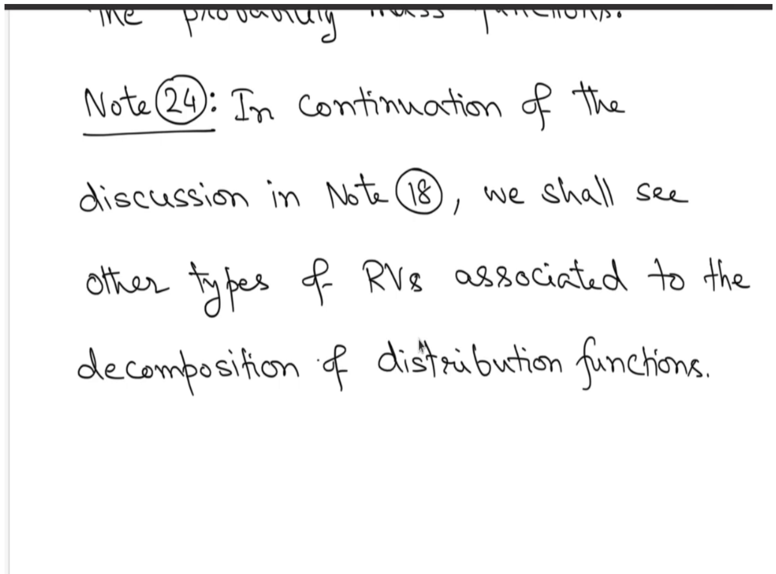
So, some final comments, that for the case of discrete random variables, we have now connected the measure, measure theoretic structures with the setup from the basic probability. So, we have proved these properties of distribution functions, identified jumps, we looked at discrete distribution functions, looked at discrete random variables, but we commented that we will come back to the continuous cases later on.

(Refer Slide Time: 41:54)



But we can now, therefore use the standard analysis that you do in basic probability. In particular, since you have already identified everything from the discrete random variable, you can now go back to the probability mass functions and talk about all the relevant properties.

(Refer Slide Time: 42:12)



And then, just to make this comment about continuous distribution functions, so remember, so you have associated the discrete random variables with discrete distribution functions. But then you will ask what happens to the continuous distribution functions? Does it correspond to some types of random variables? And what happens if the distribution function is a convex linear combinations of discrete and continuous distribution functions?

So, in all of these cases, you can identify appropriate class of random variables. So, we will do that identifications and we will identify these other types of random variables and that discussions will be done later on, when we identify more structures about continuous distribution functions in Week 8. So, there all, there we are going to talk about continuous random variables.

So, in the next lecture, we will come back to this issue about random variables, whether you can construct them according to specified law. That discussion we will do in the next lecture. We stop here.