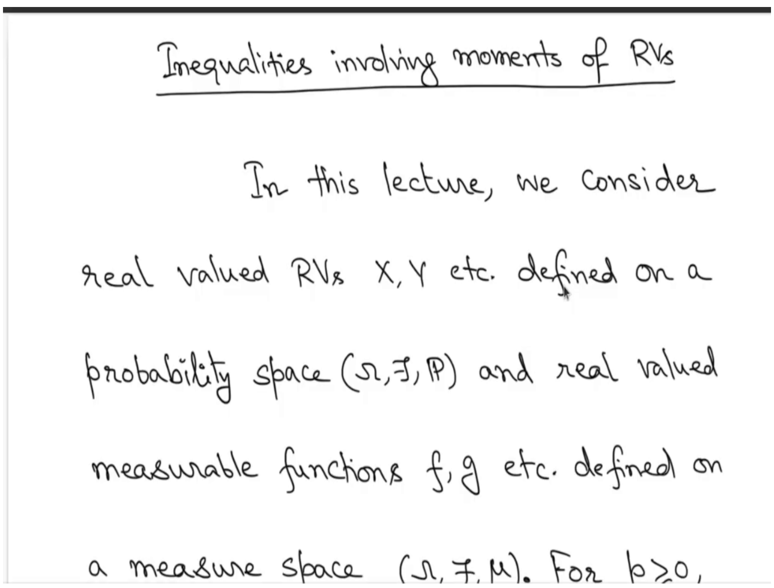
Measure Theoretic Probability 1
Prof. Suprio Bhar
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Lecture 39
Inequalities Involving Moments of RVs

Welcome to this lecture, before proceeding forward let us first quickly recall what we have done throughout this course. We have learnt about measure spaces, measurable functions, and as a special case of that we have studied probability spaces and random variables. We have also learned measured theoretic integration and as a special case we have talked about expectations of random variables. In particular, we have seen that this general definition reduces to the standard formulas that we know about for the cases when the random variables are discrete or absolutely continuous.

With this knowledge, we can also compute the moments of all these random variables, but now measure theory allows us to look at random variables in general, meaning we do not have to separately talk about discrete random variables or absolutely continuous random variables.

We can put them together at the same time; this allows us much more freedom in writing down many of the expressions or stating their proofs. We are going to see certain important inequalities involving moments of these random variables in general. Let us move forward and look at the slides of this lecture.

(Refer Slide Time: 01:46)



a measure space $(\Omega, \mathcal{F}, \mu)$. For $p \geq 0$,

we are interested in the quantities,

$\mathbb{E}|X|^p = \int_\Omega |X|^p \, d\mathbb{P}$ and $\int |f|^p \, d\mu$ etc.

Note that $|X|^p$ and $|f|^p$ are non-negative

and hence the above integrals always

exist, but may take the value ∞.

Note that $|X|^p$ and $|f|^p$ are non-negative

and hence the above integrals always

exist, but may take the value ∞.

We first recall an algebraic inequality.

This result is stated without proof.

Lemma ①: (Jensen's inequality, finite form)

In this lecture we are going to stick to these notations that X and Y will be real valued random variables defined on a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and this measurable function f, g, etc. shall refer to measurable functions real-valued defined on this measurable space $(\Omega, \mathcal{F})$ but together with some given measure μ. What we are interested in are these kinds of quantities $\int_\Omega |X|^p \, d\mathbb{P}$ or $\int |f|^p \, d\mu$.

Here p is some non-negative real number. Remember $|X|^p$ or $|f|^p$ whatever they are these are non-negative measurable functions and hence, these above integrals always will exist. However, this may take the value ∞; we allow that in our analysis. Before we start, we first recall an algebraic inequality and this we state without proof.

(Refer Slide Time: 03:03)

This result is stated without proof.

Lemma 1: (Jensen's inequality, finite form)

Let U be an open convex set and let $\phi: U \to \mathbb{R}$ be convex. Then for $x_1, \ldots, x_n \in U$ and scalars $a_1, \ldots, a_n > 0$, we have

$$\phi\left(\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i} \right) \leq \frac{\sum_{i=1}^{n} a_i \phi(x_i)}{\sum_{i=1}^{n} a_i}.$$

Note 23: If $\phi: U \to \mathbb{R}$ is twice differentiable, then $\phi''(x) > 0 \ \forall x \in U$ implies that ϕ is convex on U.

Lemma 2: (Young's inequality)

Let $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $a, b \geq 0$, we have

Lemma 2: (Young's inequality)

Let $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then for all $a, b \geq 0$, we have

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Further, the equality holds if and only if $b = a^{p-1}$.

Proof: The inequality holds if either $a = 0$ or $b = 0$.

This is referred to as the Jensen's inequality finite form. What do we do? Take any open convex set U, what do I mean? I mean that U must be an open subset of the real line, but it should also have this property that given any two points inside this set U, the average of that value also lies inside the set U, i.e., if $a, b \in U$, then $\frac{a+b}{2} \in U$, so that is the convexity of the set.

Consider a function which is defined on U taking real values and this function we want it to be convex, what we need is that the function value at a midpoint will be less equal to the average of the function value that is the convexity of the function. In this finite form we are interested in certain other type of averages; here what we do we choose n many points from the domain U and scalars $a_1, a_2, \ldots, a_n > 0$, then

$$\phi\left(\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i}\right) \leq \frac{\sum_{i=1}^{n} a_i \phi(x_i)}{\sum_{i=1}^{n} a_i}.$$

What are we looking at is that this is the weighted average of the points x_i with respect to the weights a_i and we are dividing by the total weight and if you evaluate the function we are saying this will be less or equal to the same weighted average, but of the function values. That is all we are saying in this statement.

If our given function phi is twice differentiable then there is this easy condition that allows us to check whether phi is convex or not? So what do we do we lo at the double derivative of the function and if it so happens that the double derivative is strictly positive for all points in the domain then phi will be convex. This is a sufficient condition that will imply that the given function is convex, so this is quite useful in practice to check that a given function is convex.

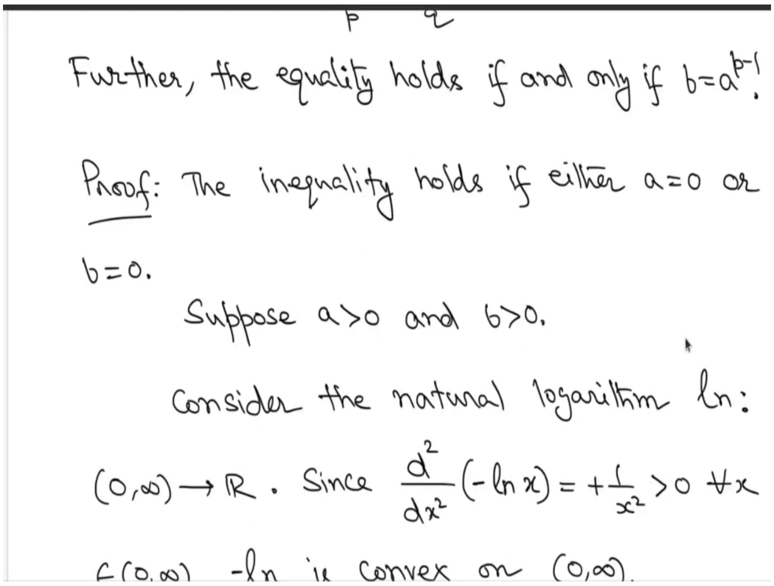
Using this we are moving on and we are going to show this important inequality called the Young's inequality. What is the statement? Here we are choosing two real numbers p and q between 1 and ∞, so we are not allowing them to be equal to the boundary points 1 or ∞. They are strictly between 1 and ∞.

With the relation that $\frac{1}{p} + \frac{1}{q} = 1$ then the statement says for all a and b non-negative the product of $ab \le \frac{a^p}{p} + \frac{b^q}{q}$. Here, note that, $\frac{1}{p} + \frac{1}{q} = 1$ is actually giving you a convex combination.

What we are seeing here is that ab is less or equals to some kind of a convex combination of these values a^p and b^q and the statement also says that this equality will hold if and only if a and b are related by this relation that $b = a^{p-1}$. Here note that $p > 1$, therefore, $p - 1 > 0$, so a^{p-1} is well defined, so there is no issue about that.

How do you show this? So first let us get rid of the boundary related conditions so if $a = 0$ or $b = 0$ then this product is 0 this product appearing on the left-hand side is 0 and therefore whatever is appearing on the right hand side that will be greater equals to 0, because that term is non-negative anyway, so this inequality is automatically satisfied provided either a or b is 0.

(Refer Slide Time: 07:02)



$\in (0,\infty)$, $-\ln$ is convex on $(0,\infty)$.

Applying lemma 1 to $-\ln : (0,\infty) \to \mathbb{R}$, we have

$$-\ln\left(\frac{a^p}{p} + \frac{b^q}{q}\right) \leq -\frac{\ln a^p}{p} - \frac{\ln b^q}{q} = -\ln(ab).$$

The inequality follows.

The case of equality is left as an exercise.

Theorem 4 (Hölder's inequality)

Theorem 4 (Hölder's inequality)

Let $f, g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be measurable. Fix $p, q \in (1,\infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\int |fg| \, d\mu \leq \left(\int |f|^p d\mu\right)^{1/p} \left(\int |g|^q d\mu\right)^{\frac{1}{q}}.$$

Proof: Exercise (Hint: the functions $|fg|$, $|f|^p$ & $|g|^q$ are non-negative. Their integrals

Proof: Exercise (Hint: the functions $|fg|$, $|f|^p$ & $|g|^q$ are non-negative. Their integrals

take values in $[0,\infty]$. Use Lemma 2 with $a = |f(\omega)| \left(\int |f|^p d\mu\right)^{-1/p}$ etc.)

Note 24 (i): The integrals appearing in Theorem 4 may be $+\infty$. In particular, if $\int |fg| \, d\mu = +\infty$, then $\int |f|^p d\mu > 0$ and $\int |g|^q d\mu > 0$ with

Now, consider the case when both a and b are positive, here we apply this trick that we are going to choose a very specific convex function and we are going to apply the Jensen's inequality the finite form of it. Here we start off with the natural logarithm function

$ln : (0, \infty) \rightarrow \mathbb{R}$. Here what happens is $\frac{d}{dx^2}(-\ ln\, x) = \frac{1}{x^2} > 0,\ \forall x$.

Therefore, the $-\ ln$ function is a convex function on $(0, \infty)$. Here we have used the fact that double derivative of this function is positive and therefore, it must be convex, so we have used this criteria here. If you apply the Jensen's inequality finite form to the minus logarithm function, here note that this is logarithm with respect to the base e.

Now, if you apply this we get this fact that minus logarithm of this convex linear combination must be less or equal to this quantity, which is the convex linear combination of the function values but then observe that this right side quantity is nothing but $-\ log\,(ab)$. If you exponentiate it the inequality will follow, by exponentiating you are going to get ab is less or equal to this convex linear combination that was appearing on the right hand side.

Therefore, the inequality is following from this argument, but the case of equality is being left as exercises please work it out. We are going to use this Young's inequality to allow us to prove a very important inequality called the Holder's inequality. Here we are finally entering the integrations that we are interested in. What is happening is that we are choosing two measurable functions exactly as stated in the introduction, so f and g are real valued measurable functions defined on this measurable space $(\Omega, \mathcal{F})$.

Fix two real numbers p and q between 1 and $+\ \infty$, we are not allowing the value 1 for this p and q and we also do not want them to take the value $+\ \infty$. This is strictly between 1 and ∞ these are real numbers, but we also want that this convex linear combination should be formed so by the time in that $\frac{1}{p} + \frac{1}{q}$ should be 1 then the claim is that if you look at

$$\int |fg|\, d\mu \leq \left(\int |f|^p\, d\mu\right)^{1/p} \left(\int |g|^q\, d\mu\right)^{1/q}.$$

Here we are looking at the product of the $\left(\int |f|^p \, d\mu\right)^{1/p}$ and $\left(\int |g|^q \, d\mu\right)^{1/q}$. So, we are looking at these two integrations and looking at their appropriate powers and then finally we are multiplying. This is what appears on the right hand side.

Now, the proof is being left as an exercise this is an important applications of the Young's inequality, but here is a hint, first observe that the functions that appear in this inequality which are $|fg|$, $|f|^p$ and $|g|^q$ all of these are non negative measurable functions. Therefore, the integrations exist but could take the value of ∞.

The first thing to note is that if it so happens that the right-hand side is 0 then at least one of the integrations that appear on the right-hand side must be 0 and that as per our understanding will imply that one of the functions f or g must be 0 μ almost everywhere but then their product will also be 0 μ almost everywhere and therefore this integration will also turn out to be 0. Therefore, the inequality will follow if at least one of the terms on the right hand side is 0.

Then we are interested in the situation where these quantities are non-trivial, so if any one of these quantities are $+\infty$, then also the inequality follows but now we are interested in the situation where these integrations that are appearing on the right hand side are some non-trivial real numbers, not $+\infty$, not 0.

Then what you can do is that you can apply the Young's inequality with these specific choices of a and b. What do you do? You look at the value of a for every fixed value of point ω, you look at $|f(\omega)|\left(\int |f|^p\right)^{-1/p}$, so this quantity appears on the right hand side, see. Here we are just dividing by that quantity and looking at this.

So, choose a like this and similarly you will choose for b the value $|g(\omega)|\left(\int |g|^q\right)^{-1/q}$. If you choose a and b that like that then if you apply the Young's inequality you will immediately get the required inequality. So, this is the Holder’s inequality and this gives estimates for integration of the product of the functions.

(Refer Slide Time: 12:57)

$a = |f(\omega)|(\int |f|^p d\mu)^{-1/p}$ etc.)

Note 24 (i): The integrals appearing in Theorem 4 may be $+\infty$. In particular, if $\int |fg| d\mu = +\infty$, then $\int |f|^p d\mu > 0$ and $\int |g|^q d\mu > 0$ with at least one of $\int |f|^p d\mu$, $\int |g|^q d\mu$ being $+\infty$. However, if the right hand side of Hölder's inequality is $+\infty$, then we can not claim the finiteness of $\int |fg| d\mu$. The next

$p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\int |fg| d\mu \leq \left(\int |f|^p d\mu\right)^{1/p} \left(\int |g|^q d\mu\right)^{\frac{1}{q}}.$$

Proof: Exercise (Hint: the functions $|fg|$, $|f|^p$ & $|g|^q$ are non-negative. Their integrals take values in $[0, \infty]$. Use Lemma 2 with $a = |f(\omega)|(\int |f|^p d\mu)^{-1/p}$ etc.)

Hölder's inequality is $+\infty$, then we can not claim the finiteness of $\int |fg| d\mu$. The next result is a version of Theorem ④, assuming finiteness of appropriate integrals.

(ii) If $|f|^p$ and $|g|^q$ are μ-integrable, then the integrals on the right hand side of the Hölder's inequality are finite. In

Now, there are several comments that appear here, so look at this integration that appear in this inequality they could be $+\infty$ and as remarked they can give some important comments or important implications when they are ∞. So, if it so happens that this term which appears on the left hand side of the inequality if this integration is ∞ then it immediately tells you that the right hand side terms which involve this $|f|^p$'s integration or $|g|^q$ integrations these cannot be 0, because remember 0 times anything is 0.

So, that is our convention therefore, both of these must be positive. That is the first observation; moreover, at least one of them should be now ∞ for the inequality to hold. Let us go back to this inequality, we are saying if the left hand side is ∞ then at least one of them must be ∞ and the other must be positive. None of them can be 0; if one of them is 0 then the product is 0.

So, you have to be careful with that. That is the first observation that we get, so if the left hand side is ∞ then it still gives you some information about the right hand side. However, if it so happens that the right hand side term is $+\infty$ meaning, this product is $+\infty$ then this inequalities automatically satisfied and this does not give you any useful information for the left hand side.

We are going to restate this result of theorem 4 in terms of certain appropriate integrations for random variables so that we are going to do in a minute but now $|f|^p$ and $|g|^q$ whose integrations we are considering, suppose they are mu integrable so their integrations will be finite. Then all these terms that appear on this Holder's inequality must be finite and this actually tells you by the Holder's inequality that the left-hand side term which is the integration of the product of the functions must be finite and that tells you that $|fg|$ is μ -integrable.

So, this is an important comment about integrability, so let us go back to that statement once more so we are saying if both the terms on the right-hand side are finite then the left-hand side is finite but this being finite will tell you that $|fg|$ has a finite integration that means that fg is μ-integrable.

(Refer Slide Time: 15:32)

Theorem 5: let X, Y, f, g be as above.

(i) For $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then,

$$\mathbb{E}|XY| \leq \left(\mathbb{E}|X|^p\right)^{\frac{1}{p}} \left(\mathbb{E}|Y|^q\right)^{\frac{1}{q}}.$$

(ii) (Cauchy-Schwarz inequality for RVs)

Taking $p = q = 2$ in (i), we have

Taking $p = q = 2$ in (i), we have

$$\mathbb{E}|XY| \leq \left(\mathbb{E}X^2\right)^{\frac{1}{2}} \left(\mathbb{E}Y^2\right)^{\frac{1}{2}}.$$

If $\mathbb{E}X$ and $\mathbb{E}Y$ exist and are real valued, then,

$$\mathbb{E}\left|(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right| \leq \left(\mathbb{E}(X - \mathbb{E}X)^2\right)^{\frac{1}{2}} \left(\mathbb{E}(Y - \mathbb{E}Y)^2\right)^{\frac{1}{2}}.$$

(iii) (Cauchy-Schwarz inequality for measurable functions) Taking $p = q = 2$ in Theorem 4,

(iii) (Cauchy-Schwarz inequality for measurable functions) Taking $p = q = 2$ in Theorem (4), we have,

$$\int |fg| \, d\mu \leq \left(\int |f|^2 d\mu \right)^{\frac{1}{2}} \left(\int |g|^2 d\mu \right)^{\frac{1}{2}}.$$

Exercise (9): When does the equality hold in Theorem (4) and Theorem (5)?

Note (25): (i) Taking $Y \equiv 1$ in Theorem (5)(i), we

We restate these results for random variables what happens here, you again continue with these notations p and q with these convex linear combinations $\frac{1}{p} + \frac{1}{q} = 1$. Then what happens? You look at $E(|XY|)$, so that is the integration of $|XY|$ with respect to the probability measure $\mathbb{P}$ and by the Holder's inequality you immediately get this bound on the right-hand side.

This is simply a special case of the Holder's inequality when you are considering random variables for the measurable functions and the probability measure appears in place of the measure mu. Therefore, you immediately get this but now consider this special case which is well known as the Cauchy Schwarz inequality what do you do you choose p and q to be 2.

Here all you need is that p and q should be between 1 and ∞ forming this convex linear combination. Now if you choose $p = q = 2$ these conditions are satisfied and then you will get this interesting inequality. That tells you that $E(|XY|)$ is bounded by this quantity on the right hand side and that involves the second moments of the random variables X and Y.

If it so happen that $E(X)$ and $E(Y)$ will exist and will be real valued, so in this case what you can do? Instead of looking at the random variables X and Y, look at the random variables $X - E(X)$ and $Y - E(Y)$. Therefore, applying this observation, this inequality that we have just observed to these new random variables, you get this interesting bound, we are going to make some comments about this in a minute.

Let us come to Cauchy-Schwarz inequality for the measurable functions this is in general statement again you chooses p and q to be 2 and then you get this required inequality as a special case of the Holder's inequality. What you can ask is that there are this Holder's inequality that you have just discussed and there are these Cauchy-Schwarz inequalities and these inequalities involving random variables.

Now, what you can ask is that when does the equality hold? In Holder's inequality which is theorem 4 and basically the applications of Holder's inequality which is theorem 5. Please check this when does the equality hold.

(Refer Slide Time: 18:10)

Theorem (4) and Theorem (5)?

Note (25): (i) Taking $Y \equiv 1$ in Theorem (5)(i), we have $\mathbb{E}|X| \leq \left(\mathbb{E}|X|^p\right)^{1/p}$ for any $p \in (1, \infty)$.

(ii) Fix $p, q \in (0, \infty)$ with $p < q$. Put $Z = |X|^p$ and $r = \frac{q}{p} > 1$. Then, by part (i),

$$\mathbb{E}|X|^p = \mathbb{E} Z \leq \left(\mathbb{E}|Z|^r\right)^{\frac{1}{r}} = \left(\mathbb{E}|X|^q\right)^{p/q}.$$

Rewriting the above inequality, we have

(i) For $p, q \in (1, \infty)$ with $\frac{1}{p} + \frac{1}{q} = 1$. Then,

$$\mathbb{E}|XY| \leq \left(\mathbb{E}|X|^p\right)^{\frac{1}{p}} \left(\mathbb{E}|Y|^q\right)^{\frac{1}{q}}.$$

(ii) (Cauchy-Schwarz inequality for RVs)

Taking $p = q = 2$ in (i), we have

$$\mathbb{E}|XY| \leq \left(\mathbb{E} X^2\right)^{\frac{1}{2}} \left(\mathbb{E} Y^2\right)^{\frac{1}{2}}.$$

$\mathbb{E}|X| = \mathbb{E}Z \le (\mathbb{E}|Z| \;\;) = (\mathbb{E}|X| \;\;)$.

Rewriting the above inequality, we have

$$\left(\mathbb{E}|X|^p\right)^{1/p} \le \left(\mathbb{E}|X|^q\right)^{1/q}.$$

(iii) If the second moments of X and Y exist finitely, then by (i), $\mathbb{E}|X| < \infty$ and $\mathbb{E}|Y| < \infty$. In particular, $\mathbb{E}X \in \mathbb{R}$ and $\mathbb{E}Y \in \mathbb{R}$. In this case, using Theorem 5 (ii), we get

(ii) (Cauchy-Schwarz inequality for RVs)

Taking $p = q = 2$ in (i), we have

$$\mathbb{E}|XY| \le \left(\mathbb{E}X^2\right)^{\frac{1}{2}} \left(\mathbb{E}Y^2\right)^{\frac{1}{2}}.$$

If $\mathbb{E}X$ and $\mathbb{E}Y$ exist and are real valued, then,

$$\mathbb{E}\left|(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right| \le \left(\mathbb{E}(X-\mathbb{E}X)^2\right)^{\frac{1}{2}} \left(\mathbb{E}(Y-\mathbb{E}Y)^2\right)^{\frac{1}{2}}.$$

(iii) (Cauchy-Schwarz inequality for measurable functions) Taking $p = q = 2$ in Theorem 4,

We are interested in special sub cases of the Holder's inequality, here what we do? We look at this random variable which is identically equal to 1 then go back to the theorem 5, 1 so what is this so go back to that statement so we have this statement that $E(|XY)$ is dominated by this quantity. However, if you put $Y = 1$ then what happens your left hand side is nothing but $E(|X|)$ and on the right hand side this moment of Y is nothing but 1.

Therefore, you get the inequality that $E(|X|) \le \left(E(|X|^p)\right)^{1/p}$. Here you are able to choose any number p between 1 and ∞. So that is as per the Holder's inequality.

We try to extend this inequality for general indices, here you choose p and q between 0 and ∞, so these are positive numbers and choose $p < q$. There is no other restrictions on p and q

these are positive numbers and $p < q$. Consider this random variable Z which is mod X to the power p and choose this number $r = \frac{q}{p}$ and by our choice of p and q, $r > 1$.

Then by this first part that we have just observed you can look at $E(Z)$ that is nothing but $E(|X|^p)$, but then look at this $E(Z)$. Z is non-negative, so it is equal to $E(Z)$. If you apply this part 1 you will immediately claim that for this number r you will get this bound, so you are looking at the r-th absolute moment of Z and looking at the r-th root of that.

That is the inequality that is stated in the part 1 of theorem, but if you write Z again back in terms of X you get this expression here. If you rewrite this inequality just for the random variable X you get this interesting inequality which says that for any p and q which are positive and if then you get this bound.

This says that the pth roots of pth absolute moments are increasing in that index p, so if you go from p to q this quantity increases. For $p = 1$ we already had this inequality in part 1. That is exactly what we have extended from the part 1 to part 2 we have just extended that. As a special case of this observation, let us consider the second moments of X and Y.

So, suppose these are finite then what will this give, this will tell you that $E(X)$ and $E(Y)$ will also be finite because this will be dominated by the square root of the second moment and square root of the second moment both for the random variables X and Y. What will happen is that in this case X and Y are integrable and therefore these values $E(X)$ and $E(Y)$ are in the real numbers.

Therefore, you go back to this Cauchy-Schwarz inequality that we stated in theorem 5 part 2, so let us go back to that statement. In theorem 5 part 2 we had stated these inequalities and we already have this situation where $E(X)$ and $E(Y)$ are real numbers. Let us use this inequality what happens?

(Refer Slide Time: 22:03)

$\mathbb{E}|Y|<\infty$. In particular, $\mathbb{E}X\in\mathbb{R}$ and $\mathbb{E}Y\in\mathbb{R}$.

In this case, using Theorem 5 (ii), we get the familiar inequality:

$$|\text{Covariance}(X,Y)| \leq \sqrt{\text{Variance}(X)}\sqrt{\text{Variance}(Y)}$$

where $\text{Covariance}(X,Y)=\mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big)$.

<u>Proposition 1</u>: Fix $p\in[1,\infty)$. If $|f|^p$ and $|g|^p$

<u>Proposition 1</u>: Fix $p\in[1,\infty)$. If $|f|^p$ and $|g|^p$

are μ-integrable, then so is $|f+g|^p$.

<u>Proof</u>: It is enough to establish an inequality of the form $|f+g|^p \leq c.\left(|f|^p+|g|^p\right)$ for some constant $c>0$ and use $\int |f|^p d\mu<\infty$ etc.

(Method 1):

for some constant $c>0$ and use $\int |f|^p d\mu<\infty$ etc.

(Method 1):

$$|f+g|^p \leq \left(|f|+|g|\right)^p \leq \left(2\max\{|f|,|g|\}\right)^p$$

$$\leq 2^p \max\{|f|^p,|g|^p\} \leq 2^p\left(|f|^p+|g|^p\right).$$

(Method 2):

Since $p\geq 1$, the function $(0,\infty)\to(0,\infty)$
$x\mapsto x^p$

You get this interesting inequality that says

$|Covariance(X, Y)| \leq \sqrt{variance(X)}\sqrt{variance(Y)}$. Here $Covariance(X, Y)$ is defined as $E((X - E(X))(Y - E(Y)))$. These are again all these familiar expressions that come from your basic probability courses, so we are not going into the exact details of covariance and variance.

We assume that you know about this but we are now proving this from the major theoretic setup. Remember here X could be discrete, Y could be absolutely continuous, there is no restriction on X and Y, X and Y could be of mixed type also. With this inequality at hand so let us now move on to other discussions.

Here what do you do choose a number $p \in [1, \infty)$ and allow the value 1 for the point p. You allow this range of values for p, if it so happens that $|f|^p$ and $|g|^p$ both of these are integrable, specifically let us consider the case $p = 1$. If we assume that $|f|$ and $|g|$ are μ integrable, so by that I mean that f and g are μ integrable then we are saying that $|f + g|$ or equivalently $f + g$ is μ integrable and this is being generalized with this power p, so that is all this statement says.

If f and g are integrable then f plus g is integrable. What do we do about this proof? So, we want to prove this by some appropriate inequalities. Let us go back to this certain inequalities of this form we want to claim that $|f + g|^p$ will be dominated by $|f|^p + |g|^p$ multiplied by a suitable constant.

So, you have to figure out this suitable constant but if you can figure out this suitable constant then using the integrability of the terms that appear on this right-hand side just by taking integration in this inequality you will get the required result that $|f + g|^p$ is integrable. There are two methods we discuss here both give some ideas in how to do this kind of estimates.

First start with $|f + g|^p$, observe that $|f + g| \leq |f| + |g|$, therefore, raise to the power p you will also get this inequality. Then look at the maximum of the functions between $|f|$ and $|g|$ then $|f|$ is dominated by $\max\{|f|, |g|\}$ and similarly $|g|$ is also dominated by $\max\{|f|, |g|\}$.

Therefore, you get 2 max $\{|f|, |g|\}$. That is all that calculation is happening within the bracket and you are just raising these inequalities to the power p. Now you observe that you can now push the power p inside and say that this is $\leq 2^p \max\{|f|^p, |g|^p\}$. Then observe that maximum of these two quantities is less or equal to the sum of these two quantities, these are non-negative quantities therefore this relation is true.

You get this constant 2^p that we were after therefore, what we have is that $|f + g|^p$ is less or equal to some constant which is 2^p here multiplied by $|f|^p + |g|^p$. Another method will improve this constant slightly, so let us try to say this.

(Refer Slide Time: 26:12)

$$\leq 2^p \max\{|f|^p, |g|^p\} \leq 2^p(|f|^p + |g|^p).$$

(Method 2):

Since $p \geq 1$, the function $(0,\infty) \to (0,\infty)$, $x \mapsto x^p$ is convex. By Lemma 1,

$$|f+g|^p \leq 2^p\left(\frac{|f|+|g|}{2}\right)^p \leq 2^p \cdot \frac{|f|^p+|g|^p}{2}$$

$$= 2^{p-1}(|f|^p+|g|^p)$$

This completes the proof.

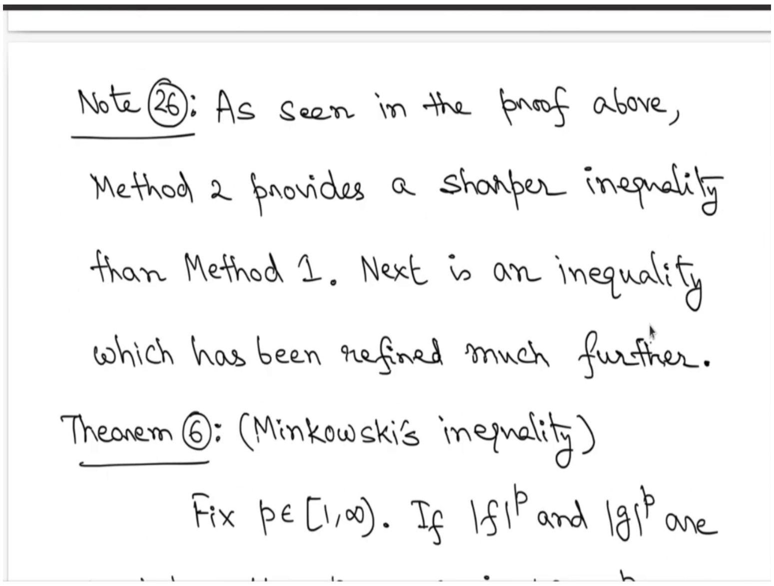
$$|\text{Covariance}(X,Y)| \leq \sqrt{\text{Variance}(X)}\sqrt{\text{Variance}(Y)}$$

where $\text{Covariance}(X,Y) = \mathbb{E}\big((X-\mathbb{E}X)(Y-\mathbb{E}Y)\big)$.

Proposition 1: Fix $p \in [1,\infty)$. If $|f|^p$ and $|g|^p$ are μ-integrable, then so is $|f+g|^p$.

Proof: It is enough to establish an inequality of the form $|f+g|^p \leq c \cdot (|f|^p + |g|^p)$

for some constant $c > 0$ and use $\int |f|^p d\mu < \infty$



Here we are choosing $p \geq 1$ this is as per the statement and then look at this function $x \to x^p$ on $(0, \infty)$. Now you can check that this function is convex, if you use this fact then by this lemma 1 which is the finite form of the Jensen's inequality you get $|f + g|^p \leq 2^p \left(\frac{|f|+|g|}{2}\right)^p$.

So, that is all we are doing here, that is the first step. If you push the power p inside by applying this convex linear combination here, so that is all we are doing you are applying the finite form of the Jensen's inequality. Here you get the power p average of that, that is all. But then what happens is that you get this inequality with 2^{p-1} appearing as that constant that we have talked about.

We are getting this fact that $|f + g|^p$ is less or equal to some appropriate constant multiplied by $|f|^p + |g|^p$. Either of these bounds will imply that $|f + g|^p$ is μ integrable or equivalently $|(f + g)^p|$ is μ-integrable.

As seen in this proof above we have applied two methods and the second method gave a sharper inequality. You were able to improve that constant in the upper bound. Now you would like to understand if there is a better result in this direction. We see this inequality which has been refined much further and this is called the Minkowski’s inequality.

(Refer Slide Time: 28:09)

Theorem ⑥: (Minkowski's inequality)

Fix $p \in [1,\infty)$. If $|f|^p$ and $|g|^p$ are μ-integrable, then so is $|f+g|^p$ and

$$\left(\int |f+g|^p \, d\mu\right)^{1/p} \le \left(\int |f|^p \, d\mu\right)^{1/p} + \left(\int |g|^p \, d\mu\right)^{1/p}.$$

Proof: Case $p=1$: observe that $|f+g| \le |f| + |g|$ and hence

$$\int |f+g| \, d\mu \le \int |f| \, d\mu + \int |g| \, d\mu$$

Case $p>1$: Choose $q \in (1,\infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$\int |f+g|^p \, d\mu = \int |f+g| \cdot |f+g|^{p-1} \, d\mu$$

$$\le \int |f| \, |f+g|^{p-1} \, d\mu + \int |g| \, |f+g|^{p-1} \, d\mu. \quad \cdots\cdots (*)$$

Check that $\int \left(|f+g|^{p-1}\right)^q d\mu < \infty$ (Exercise)

Check that $\int \left(|f+g|^{p-1}\right)^q d\mu < \infty$ (Exercise)

Now, we apply the Hölder's inequality on the right hand side of $(*)$. We obtain,

$$\int |f+g|^p \, d\mu \le \left(\int |f|^p \, d\mu\right)^{1/p} \left(\int |f+g|^{(p-1)q} \, d\mu\right)^{1/q}$$

$$+ \left(\int |g|^p \, d\mu\right)^{1/p} \left(\int |f+g|^{(p-1)q} \, d\mu\right)^{1/q}$$

We are choosing this number $p \in [1, +\infty)$, so ∞ excluded but 1 allowed. If it so happens that $|f|^p$ and $|g|^p$ are μ-integrable then we want to again claim that this is also integrable so that is part of the statement. Then here we are going to say an important inequality that is going to give you the relation between these summations.

Let us try to see this. In the case $p = 1$ observe that $|f + g| \leq |f| + |g|$ and therefore, if you integrate these functions on both sides by this measure μ then you get the required relation. The required relation is very easy to check when $p = 1$ but now you want to show this when $p > 1$. In this case what do you do?

You choose $q \in [1, +\infty)$ such that this convex linear combination is formed, so p is given to you and you are going to choose this q appropriately. So that $\frac{1}{p} + \frac{1}{q} = 1$, In this case what happens? Then integration of $|f + g|^p$ you can write it as a product of $|f + g|\ |f + g|^{p-1}$.

You are writing in terms of a product in preparation for applying the Holder's inequality and that is why you are also choosing this q to form this convex linear combination. For this $|f + g|$ part you apply this inequality that was observed earlier that $|f + g| \leq |f| + |g|$. Therefore, you get these two separate integrations this is simply by additivity of the integrations you get two separate integrations.

Here integrability is not an issue because we are only dealing with functions which are taking non-negative values, but here we want to focus on this part here that is $|f + g|^{p-1}$. Again, $p > 1$ so there is no problem in this definition. We claim that this function here raise to the power q, if you want to integrate it with respect to μ you will get a finite quantity. This is left as an exercise please check this.

What do we do with this observation at hand we are going to apply the Holder's inequality on the right-hand side of this star relation, so that was in preparation for applying the Holder's inequality that in both these integrations the integrands are factored and we are going to apply the Holder's inequality here.

(Refer Slide Time: 31:09)

Now, we apply the Hölder's inequality on the right hand side of (*). We obtain,

$$\int |f+g|^p \, d\mu \le \left(\int |f|^p \, d\mu\right)^{1/p} \left(\int |f+g|^{(p-1)q} \, d\mu\right)^{1/q}$$

$$+ \left(\int |g|^p \, d\mu\right)^{1/p} \left(\int |f+g|^{(p-1)q} \, d\mu\right)^{1/q}$$

$$= \left[\left(\int |f|^p \, d\mu\right)^{1/p} + \left(\int |g|^p \, d\mu\right)^{1/p}\right]$$

$$\int |f+g|^p \, d\mu = \int |f+g| \cdot |f+g|^{p-1} \, d\mu$$

$$\le \int |f| \, |f+g|^{p-1} \, d\mu + \int |g| \, |f+g|^{p-1} \, d\mu. \quad \cdots\cdots (*)$$

Check that $\int \left(|f+g|^{p-1}\right)^q d\mu < \infty$ (Exercise)

Now, we apply the Hölder's inequality on the right hand side of (*). We obtain,

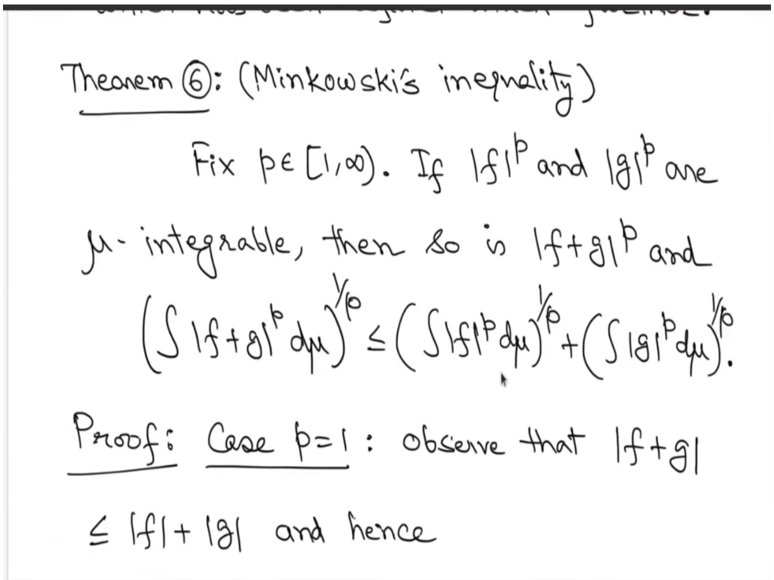
$$\int |f+g|^p \, d\mu \le \left(\int |f|^p \, d\mu\right)^{1/p} \left(\int |f+g|^{(p-1)q} \, d\mu\right)^{1/q}$$

$$+ \left(\int |g|^p \, d\mu\right)^{1/p} \left(\int |f+g|^{(p-1)q} \, d\mu\right)^{1/q}$$

$$= \left[\left(\int |f|^p \, d\mu\right)^{1/p} + \left(\int |g|^p \, d\mu\right)^{1/p}\right]$$

$$\times \left(\int |f+g|^p \, d\mu\right)^{\frac{p-1}{p}}.$$

The inequality follows.



Theorem ⑥: (Minkowski's inequality)

Fix $p \in [1,\infty)$. If $|f|^p$ and $|g|^p$ are μ-integrable, then so is $|f+g|^p$ and

$$\left(\int |f+g|^p d\mu\right)^{1/p} \leq \left(\int |f|^p d\mu\right)^{1/p} + \left(\int |g|^p d\mu\right)^{1/p}.$$

Proof: Case $p=1$: observe that $|f+g| \leq |f| + |g|$ and hence

If you apply this then you immediately obtain that $|f + g|^p$ integration with respect to the measure μ that is on the left-hand side is less or equal to these two terms. So, the first integral that was there on the right-hand side will contribute this term that the p-th root of the integration of $|f|^p$, so that is appearing here and similarly we will get this term $(|g|^p)^{1/p}$ integration of that, that is all you are getting.

Now, this common term is appearing on both the terms, so let us just go back to this integration in start. You are just factoring it, you are raising this part to the power p, this part to the power q, similarly, you are raising this part to the power p, this part to the power q, so that is all you are applying and you are getting this as a consequence of the Holder's inequality.

Now, this term is common, you take that out and if you do this calculation a bit carefully you can show that $|f + g|^{(p-1)q}$ that quantity is nothing but this quantity, so please try to check this. What is happening? So, on the left-hand side I have $|f + g|^p$ integration of that and we have some term within the bracket here multiplied by the left-hand side raised to the power $\frac{p}{1-p}$.

Let me repeat, so I have the left-hand side which is less or equal to this term within the brackets multiplied by the expression on the left-hand side raised to the power $\frac{p-1}{p}$. So, you cancel things off and you are going to get the required inequality that the p-th root of the left hand side will be less equal to the sum of these p-th roots.

That is the statement that was given in this inequality here in Minkowski's inequality. We get this result and as a consequence you also get the integrability of $|f + g|^p$.

(Refer Slide Time: 33:19)

The inequality follows.

Note (27): The inequality in Theorem (6) is true if any one of the integrals on the right hand side is infinite. We can make comments similar to Note (24)(i).

Proposition (2): (Markov-Chebyshev inequalities)

Fix $p \in [1, \infty)$. If $|f|^p$ and $|g|^p$ are μ-integrable, then so is $|f+g|^p$ and

$$\left(\int |f+g|^p d\mu\right)^{1/p} \leq \left(\int |f|^p d\mu\right)^{1/p} + \left(\int |g|^p d\mu\right)^{1/p}.$$

Proof: Case $p=1$: observe that $|f+g| \leq |f| + |g|$ and hence

$$\int |f+g| d\mu \leq \int |f| d\mu + \int |g| d\mu.$$

Proposition ②: (Markov-Chebyshev inequalities)

(i) let f be any measurable function and fix $\varepsilon > 0$. Then

$$\mu(|f| \geq \varepsilon) \leq \frac{1}{\varepsilon} \int |f| \, d\mu.$$

More generally, $\mu(|f| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \int |f|^p \, d\mu$ for any $p \in (0, \infty)$.

More generally, $\mu(|f| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \int |f|^p \, d\mu$ for any $p \in (0, \infty)$.

(ii) let X be an RV and fix $\varepsilon > 0$. Then

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}|X|.$$

More generally, $\mathbb{P}(|X| \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \mathbb{E}|X|^p$ for any $p \in (0, \infty)$.

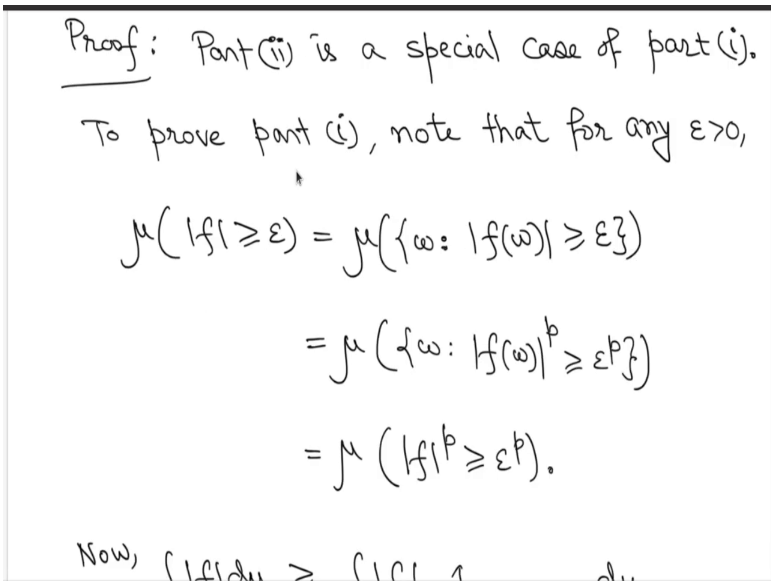
We see some comments involving this Minkowski's inequalities, so the inequality in this theorem is true even if one of these integrations on the right-hand side is infinite. On the right-hand side, you get a summation of two integrations. If anyone of these terms is infinite even then this inequality is still true but we are really interested in the case when these terms are finite that is when you get some useful information bounding this left-hand side.

That is why in the hypothesis we had assumed these integrability conditions so that this right-hand side becomes finite. Even if that is not the case you can still get this inequality but you do not really get any useful information for the bound for the left-hand side. We now have some good ideas about the estimates involving these moments, but now we want to look at certain more interesting inequalities and this is a very very useful inequality and referred to as Markov-Chebyshev inequality.

Here what do we do? So, choose any measurable function and fix epsilon to be positive then look at the size of this set where $|f| > \epsilon$ and we say that we can dominate it, we can estimate this by this quantity the integration of $\frac{|f|}{\epsilon}$. More generally there is this upper bound given to us which is $\frac{1}{\epsilon^p} \int |f|^p \, d\mu$. Here $p \in (0, \infty)$ and this relation holds for any fixed $\epsilon > 0$. Now for the case of random variables you get this special statement which can be represented in terms of probability of certain events. We are just rewriting this inequality that was stated here in terms of the random variable. What is this? We are saying that $\mathbb{P}(|X| \geq \epsilon) \leq \frac{1}{\epsilon} E(|X|)$.

Here again the inequality will be meaningful if X is integrable that is $E(|X|)$ is finite. More generally in this case when you are involving this scalar $0 < p < \infty$ you can still get these estimates of these probabilities in terms of this absolute moment. This is anyway a special case of part 1.

(Refer Slide Time: 35:50)



$$= \mu\left(|f|^p \geq \varepsilon^p\right).$$

Now, $\int |f| d\mu \geq \int |f| \, \mathbb{1}_{(|f| \geq \varepsilon)} d\mu$

$$\geq \varepsilon \int \mathbb{1}_{(|f| \geq \varepsilon)} d\mu$$

$$= \varepsilon . \mu(|f| \geq \varepsilon).$$

Similarly, $\int |f|^p d\mu \geq \varepsilon^p \mu(|f|^p \geq \varepsilon^p)$

$$\geq \varepsilon \int \mathbb{1}_{(|f| \geq \varepsilon)} d\mu$$

$$= \varepsilon . \mu(|f| \geq \varepsilon).$$

Similarly, $\int |f|^p d\mu \geq \varepsilon^p \mu(|f|^p \geq \varepsilon^p)$

$$= \varepsilon^p \mu(|f| \geq \varepsilon).$$

This completes the proof.

Note 28: If $\mathbb{E}X^2 < \infty$, then observe the

Let us see how we can prove this, so here for any epsilon positive look at this size of this set which we are interested in. Here what is happening first observe that this is the set of points in the domain such that $|f(\omega)| \geq \epsilon$, but if you raise it to the power p when $p \in (0, +\infty)$ you still get this inequality.

These two sets are identical therefore, you get the same measure great but then you can rewrite this inequality in our standard form that removing this sample point ω you can get that. These two measures of the sets are the same because the sets are the same. Look at the quantity that we are interested in which is $\int |f|$, but then observe that this function $|f|$ is an upper bound for this function which is $|f|$ multiplied by the indicator of the set we are interested in.

You just multiply by this indicator therefore, you get this inequality relation because you are saying that this function dominates this function, but then on this set the value of $|f|$ is at least ϵ and use that this quantity will dominate ϵ times the integration of this indicator function. Integration of indicator function is nothing but the size of that set and therefore, you get the inequality and you can rewrite it as measure of this set is less or equal to $\frac{1}{\epsilon}\int |f|$ which you wanted.

How do you get the version involving that scalar p? You start with this integration of $|f|^p$, if you apply this and for this scalar ϵ^p you get this relation simply by the previous argument. Instead of ϵ you are now putting ϵ^p and for the function $|f|$ you are putting $|f|^p$. You get this relation but then you have immediately observed that measure of this set is exactly equal to measure of this set.

Therefore, the inequality will follow. So, this is a very simple inequality you are just using the fact that $|f|$ dominates this function as soon as you multiply $|f|$ by this indicator. This is a very very simple proof but this has a very wide range of applications.

(Refer Slide Time: 38:15)

Note 28: If $\mathbb{E}X^2 < \infty$, then observe the following.

(i) Variance(X) is finite.

(ii) $\mathbb{E}|X| \leq (\mathbb{E}|X|^2)^{\frac{1}{2}} < \infty$, by Note 25(i) and hence X is integrable. In particular, $\mathbb{E}X \in \mathbb{R}$.

(iii) Taking $Y = |X - \mathbb{E}X|$, by the right hand side is infinite. We can make comments similar to Note 24 (i).

<u>Proposition 2</u>: (Markov-Chebyshev inequalities)

(i) let f be any measurable function and fix $\varepsilon > 0$. Then

$$\mu(|f| \geq \varepsilon) \leq \frac{1}{\varepsilon} \int |f| \, d\mu.$$

$\mathbb{E}X \in \mathbb{R}$.

(iii) Taking $Y = |X - \mathbb{E}X|$, by Proposition 2, we have,

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) = \mathbb{P}(|Y| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathbb{E}|Y|^2$$

$$= \frac{1}{\varepsilon^2} \operatorname{Var}(X),$$

for all $\varepsilon > 0$.

Let us look at it from the perspective of moments of random variables. Consider the fact when the second moment of X exists and is finite then variance is finite so that something that you have already known from your basic probability theory. Observe that we have already mentioned that if the second moment is there then it will also tell you that $E(|X|)$ is also finite and it implies that X is integrable and in particular expected value of X is a real number.

You can consider this random variable Y which is $|X - E(X)|$ and therefore if you apply this proposition to, so what was proposition 2 once more? So proposition 2 is nothing but this Markov-Chebyshev inequality. If you apply it to the random variable Y instead of X what do you get you get the fact that $\mathbb{P}(|X - E(X)| \geq \epsilon) = \mathbb{P}(|Y| \geq \epsilon)$ and you can rewrite this and you will get this expression involving $var(X)$, because the second moment of Y is nothing but the $var\ (X)$ and this is true for all $\epsilon > 0$.

Therefore, you get some estimates of the deviation of X from the mean by this ϵ quantity in terms of the variance and that is scaled by this appropriate factor involving ϵ this is true for any ϵ.

(Refer Slide Time: 39:53)

We state the next result without proof.

Proposition 3: (Jensen's Inequality)

Let X be an integrable RV. Then for any convex function $\phi: \mathbb{R} \to \mathbb{R}$ such that $\phi(X)$ is also integrable, we have

$$\phi(\mathbb{E}X) \leq \mathbb{E}\,\phi(X).$$

Note 29: We have seen special cases of

Note 29: We have seen special cases of Jensen's inequality earlier.

(i) Taking $\phi(x) := |x|, \forall x \in \mathbb{R}$, we get $|\mathbb{E}X| \leq \mathbb{E}|X|$.

(ii) Taking $\phi(x) := x^2, \forall x \in \mathbb{R}$, we get $\mathbb{E}|X| \leq (\mathbb{E}X^2)^{\frac{1}{2}}$.

get $\mathbb{E}|X| \leq (\mathbb{E}X^2)^{\frac{1}{2}}$.

(iii) For $x_1, \ldots, x_n \in \mathbb{R}$ and scalars $a_1, \ldots, a_n > 0$, Consider the discrete RV $X \sim \sum_{j=1}^{n} \frac{a_j}{a_1 + \cdots + a_n} \delta_{x_j}$. Then, we have

If we state this result without proof so this is the Jensen's inequality. Again, remember that we had mentioned this finite form of Jensen's inequality right at the beginning here we state the general form without proof. Here what we are doing? We are starting up with an integrable random variable that means $E(X)$ is a real number here we are saying that if you take any convex function defined on the real line just to remind you a convex function is continuous and hence measurable.

Therefore, this composition is well defined it will satisfy the measurability requirement. We are saying that if $\phi(X)$ is also integrable then we get this inequality. What are the hypotheses? X is integrable, $\phi(X)$ is integrable and phi is convex, provided these three things happen you get this inequality $\phi(E(X)) \leq E(\phi(X))$.

We are not going to prove this but we are going to discuss some special cases of this inequality what are this? So, if you choose this convex function $X \to |X|$ you get $|E(X)|$ is dominated by $E(|X|)$, so this we have seen earlier. Similarly, if you choose the function X going to X^2 then you get this familiar inequality once more. That the first absolute moment of X is dominated by the square root of the second moment.

You can choose this as a curiosity that you can choose this numbers $x_1, x_2, \ldots, x_n$ from the real numbers and scalars $a_1, a_2, \ldots, a_n$ positive. Here you get this discrete random variable X with this specific law that it is focused on these Dirac masses but with these weights

$\frac{a_j}{a_1+a_2+..+a_n}$. Therefore, this will give you a convex linear combination and this is a genuine probability measure.

(Refer Slide Time: 41:51)

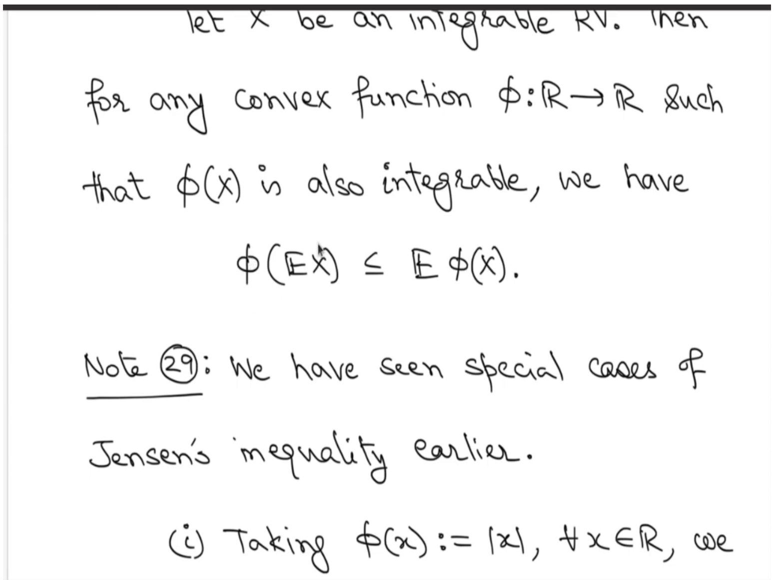
$$X \sim \sum_{j=1} \frac{a_j}{a_1+\cdots+a_n} \delta_{x_j}.$$ Then, we have

$$\phi(X) \sim \sum_{j=1}^{n} \frac{a_j}{a_1+\cdots+a_n} \delta_{\phi(x_j)}, \quad \mathbb{E}X = \frac{\sum_{j=1}^{n} a_j x_j}{\sum_{j=1}^{n} a_j}$$

and $$\mathbb{E}\phi(X) = \frac{\sum_{j=1}^{n} a_j \phi(x_j)}{\sum a_j}.$$ Lemma ① follows.

$$\phi(X) \sim \sum_{j=1}^{n} \frac{a_j}{a_1+\cdots+a_n} \delta_{\phi(x_j)}, \quad \mathbb{E}X = \frac{\sum_{j=1}^{n} a_j x_j}{\sum_{j=1}^{n} a_j}$$

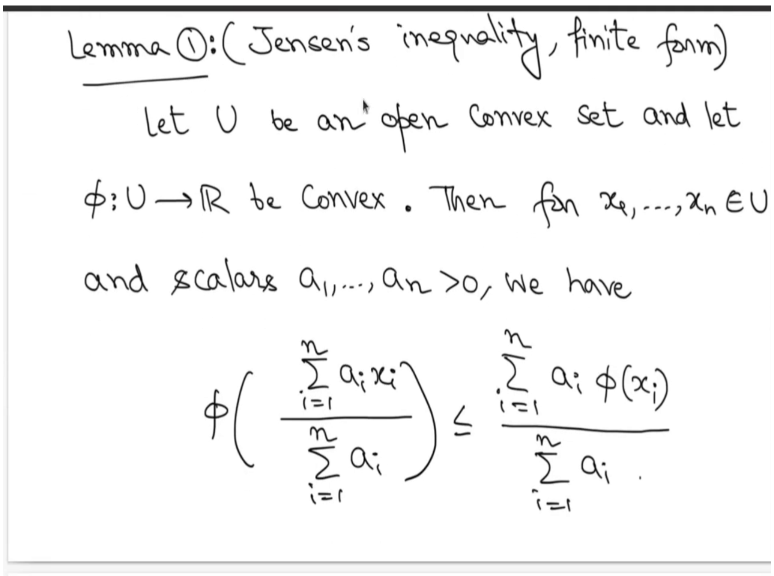
and $$\mathbb{E}\phi(X) = \frac{\sum_{j=1}^{n} a_j \phi(x_j)}{\sum_{j=1}^{n} a_j}.$$ Lemma ① follows.



Here what do you get you look at $\phi(X)$ that is nothing but these random variable once more but taking values at $\phi(x_j)$ therefore, the law of $\phi(X)$ is essentially these Dirac masses with the same convex linear combination. Look at the expected values, expected values are nothing but the weighted averages of x_j's for $E(X)$ and a weighted averages of $\phi(x_j)$'s for $E(\phi(X))$ and then immediately if you look at Jensen's inequality that will tell you that lemma 1 will follow why because you are just looking at the same inequality.

If you go back to this, so this is nothing but phi applied to this weighted averages of the quantities less or equal to $E(\phi(X))$ which we have computed at this quantity.

(Refer Slide Time: 42:47)



$a_1, \dots, a_n > 0$, Consider the discrete RV

$X \sim \sum_{j=1}^{n} \frac{a_j}{a_1 + \cdots + a_n} \delta_{x_j}$. Then, we have

$$\phi(X) \sim \sum_{j=1}^{n} \frac{a_j}{a_1 + \cdots + a_n} \delta_{\phi(x_j)}, \quad \mathbb{E}X = \frac{\sum_{j=1}^{n} a_j x_j}{\sum_{j=1}^{n} a_j}$$

$\sum^{n} a_j \phi(x_j)$

let X be an integrable RV. Then for any convex function $\phi: \mathbb{R} \to \mathbb{R}$ such that $\phi(X)$ is also integrable, we have

$$\phi(\mathbb{E}X) \leq \mathbb{E}\phi(X).$$

Note 29: We have seen special cases of Jensen's inequality earlier.

(i) Taking $\phi(x) := |x|, \forall x \in \mathbb{R}$, we

This is exactly what we had stated in lemma 1, so let us see this, so again this weighted average is evaluated at this point. If you look at that it is less equal to the weighted average of the function value, so that is exactly what was obtained as a consequence of the general Jensen's inequality.

Let us go back to that therefore, the lemma 1 follows as a special case of the Jensen's inequality, which is the general form of this, so we do not prove this. In this lecture we have discussed many of the interesting inequalities involving moments and more generally of expected values of functions of X.

We use this for estimating many of the important quantities involving the random variables and in particular by the Markov-Chebyshev inequalities we can also estimate certain important events and their probabilities. This is the final lecture in terms of the contents of this course, but in the next lecture we are going to consolidate whatever we have learned and make some concluding remarks. So let us stop here.