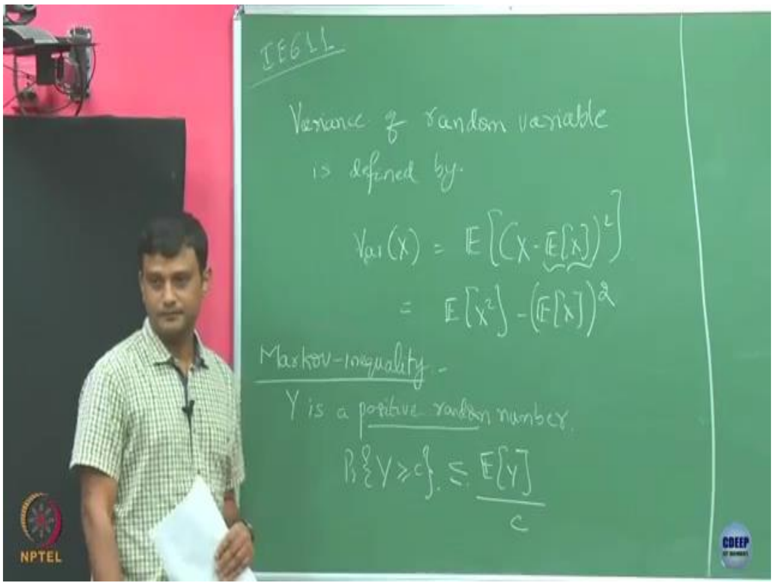
Introduction to Stochastic Processes
Prof. Manjesh Hanawal
Industrial Engineering and Operations Research
Indian Institute of Technology Bombay
Lecture 09
Variance and some inequalities of random variables

So, in the last class I talked to you what is a Discrete Random Variable? What is a Continuous Random Variable? And I introduced the notion of a Probability Density function for continuous random variable. Then we talked about what? We talked about Expectation of a Random Variable.

So, moving on, expectation is one characterization of your random variable like that gives in a sense like on an average, what is the outcome I am going to see, when I am going to perform this experiment. Other question you could ask is fine like this is the average value but when I do the experiment it is not like I am going to see some expected value like I am going to see different, different realization taken by the random variable, how the samples are going to be other from this mean value? So, mean value give some average characterization. This is what, but the actual outputs I got, how far they could be, is it that they will be very close to this mean value or they will be very far, how that? So, to characterize will have another notion called Variance.

(Refer Slide Time: 1:39)



So, see here what I am doing is, I am defining the variance to be expectation of where variable X but kind of centralizing it like I am looking, removing the mean value from it or that means what I am looking at the variation of its random variable X around its mean. But I am not worried about whether this variation is happening on the right side or on the left side. I do not want to care about it. So that is why I take the square and then I look at the expected value of this so, this is called the variance of the random variable.

Now, if you are just going to look, expand this value, you are going to see that and if you just expand it, so see here, this is a constant, right already, what you are removing from this X is a constant value that is the mean value. So this is another random variable, what we call center it random variable. And then we have squared it and took the expectation. If you are just going to expand this, this is going to be going to be this quantity where now, we have the variances, the expected value of the square of the random variable which we call it as? We call this as a second moment here, because the first moment is, my simple expectation. My second moment is expectation of X square, my mth moment is expectation of X to the power m, like that will come to that. So, this is the second moment minus the square after

Student: (())(4:14).

Professor: The first mean. So if I give a random variable and ask you to find the mean, that mean can be positive and negative, it can be both, what about variance?

Student: (())(4:37).

Professor: It is always going to be positive quantity. So, now let us see, we have seen mean variance and one of you asked say like, the mean looks like some average quantity but that is not what I am going to get, when I perform the experiment, so what you are... but mean is still some quantity of interest, like it kinds of give you the globally, what is happening, like, what kind of values I am going to observe on an average.

But what you would be interested is if I perform my experiment, what is the probability that my outcome will be larger or smaller than my mean value? So suppose, let us say the height, if you are going to take the example of height of the population, let us say mean value is, let us say 5 feet but if I am going to pick an arbitrary person, it is not necessarily that his height is going to be 5 feet But his height may be more than 5 or less than 5. Now, you may be interested in asking the question, what is the probability that, he sample I picked has a value which is larger than this 5? So how you are going to characterize such quantities?

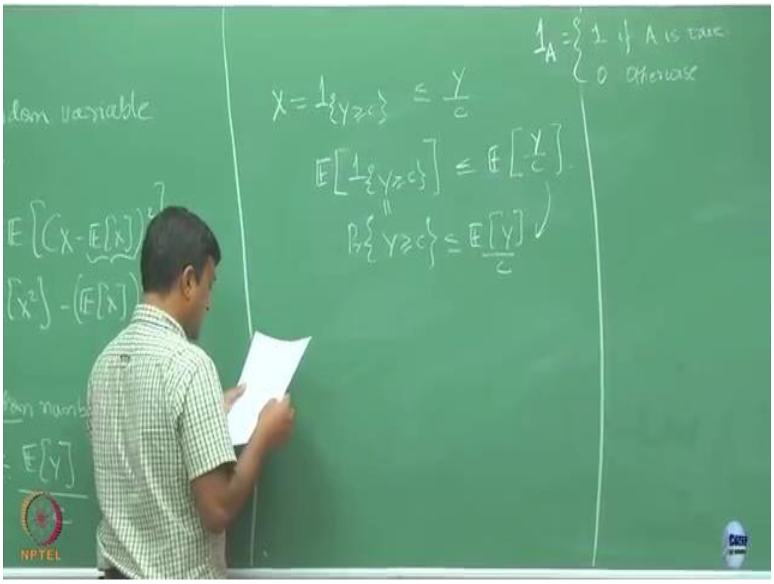
So, that comes from something called Markov’s inequality. Markov’s inequality gives a… I mean, it slightly answers a different question that I posed right now, what it tells is, if I had a random variable, what is the probability it takes a value larger than a certain number? So suppose let us say, Y is a positive random number. Then you may be interested in asking this question, what is the probability that Y is going to be greater than or equals to c. You perform an experiment and you want to know like, the outcome of your experiment is going to be larger than or equal to c. For example, let us say you are going to put your money in casino or let us say stock or something, you want to ask at any point, whether my returns are going to be larger than this quantity?

Let us say I am going to make, today if I invest in the market, the outcome you are going to get by our investment is a random quantity. But you may be interested in asking the question, Whether my outcome, my returns are going to be at least like let us say 10,000 rupees, how you are going. So, you have this, this is nothing but what? This is like compliment of your CDF? CDF is Y less than or equal to c. But let us not worry about the equality here. So, suppose you want, now, Markov, it says that this quantity is upper bounded by that you are going to be larger than this quantity c will be upper bounded by this ratio.

So now let us plug into some value. Suppose let us say c is some real number let us take c to be expected value of a random variable Y itself. So what I am asking, if I do that, well, the question I am asking is, what is the probability that Y is going to be larger than or equal to its mean value. This is going to give a value of one, but that has no meaning to me. Because I know that probably it is always going to be less than or equal to one. But if you are going to ask the question, if you are going to choose the c to be much larger than the mean value, let us say in your experiment, Y is such that Y has the mean of let us assume 0.5.

Now, if you are going to ask the question, what is the probability that Y is going to be greater than or equal to 0.5? Through this inequality, you get a trivial answer which has no meaning to me, and which has no extra information to me. But suppose if you are going to set this c to be 0.8, you are going to ask the question, ‘What is the probability that my Y is going to be large than or equal to 0.8, then this has some value, this is going to be 0.5 divided by 0.8 and in that way it characterizes, you being other from the mean value, how fast it is going to be shrink. Okay this is a simple relation, but this is one of the basic results that is useful in many, many scenarios. So, let us try to understand why this is true. And the note is that I have written this inequality only for positive random variable. This comes straightforwardly.

(Refer Slide Time: 10:27)



So if I write an indicator function like this, so, Y is a positive random variable, all of you understand this notation? This is an indicator notation what does this say?

Student: (())(10:52).

Professor: Yes, so, if this is I have defined a random variable like this, it says, you perform an experiment, if the outcome of Y is going to be greater than or equal to c that means this condition is true, right, then it is going to take the value 1. If this Y happens you take a value less than c, this condition is not true and this quantity becomes 0. So it is simply saying that so, what I am basically saying is, so A is an event. So, this is going to take 1 if A is true, 0 otherwise, such a function is called Indicator Function.

Student: If 1 or 0 is the value of A?

Professor: This 1 and 0?

Student: Is which variable's value?

Professor: We are going to define something, I will give you A and I am going to say this is the indicator function. So, here you are going to define this function, let us call this extra random variable X. What you are doing, so, X depends on Y in what way? Whenever Y takes value greater than or equals to c, X takes value 1, otherwise it takes 0. So, here, Y could be a continuous random variable, it could be taking any value. But now through this indicator function, you have defined another X which takes only 1 or 0. If you have another binary random variable, which is a function of this Y. Now, if I write like this, is this relation true for a positive random variable? now let us see, suppose Y is greater than or equal to c, the left… this quantity is 1 what about the right quantity?

Student: Greater than 1.

Professor: Greater than 1, so it hold, suppose Y is less than c, this is 0, but what is this?

Student: (())(13:12)

Professor: Some positive number, non zero, I mean positive number. So, this relation again holds. So whatever be your choice of c, this relation holds. So, now if I take expectation on both sides so, this is true for all ci. so now I am going to take expectation on both sides. If I take expectation of both sides, my claim is the expectations also return the… the inequality direction remains the same if I take the expectation. Why is this true?

Student: The random one will always be less than that.

Professor: So, we had stated this property, when did we state this property?

Student: If the property of X is greater than Y (())(14:21).

Professor: Right, so, we said that if there are two random variables X and Y such that X always dominates Y that is X is greater than or equals to Y with probability 1 then we said that expectation of X is greater than or equal to the expectation of Y, we had a name for the property.

Student: Preservation of the order.

Professor: Preservation of the order, did I apply it correctly here? So, because this is true probability 1, that means, this relation always holds irrespective of what value Y takes. And for any given c so, this is true and I could apply this, now what is this further? What is the expected value of this indicator function? So if I say this is nothing but probability Y greater than or equal to c, is this correct?

Student: Yes.

Professor: Fine, so (())(15:23) then we done right? So here, Y and c. So, from here to here, what property did I use?

Student: Scaling property.

Professor: We have just use the scaling property, right because c is a constant. So, it is why we get this.

Student: Does Markov inequality holds for all values of c?

Professor: Yes, if I am going to assume, if I do not assume c is going to be greater than 0, if this become negative, yes then it may not hold, right because I already assuming Y to be positive random variable, it does not make sense to take c negative because I know that Y is.

Student: Is it always going to be greater than zero?

Professor: It is always going to be greater than 0 at least.

Student: That quantity will be 1 and this quantity may be negative, (())(16:15).

Professor: Which one, this part? Yes, if I am going to take, this is going to be.

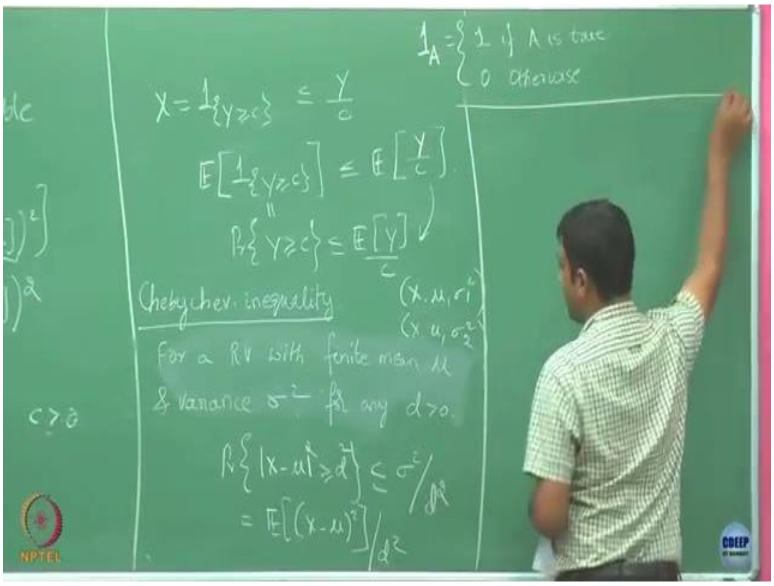
Student: c will be negative.

Professor: Yes.

Student: So, then right hand side will be (())(16:26)

Professor: Right, that is fine, so, this is going to be 1 in that case, and this is going to be negative. So, does not make, this inequality does not hold. But, fine we have to explicitly put, if you want to make it correct, so let us say c greater, strictly greater than 0 because otherwise not true, but I mean, this for a positive random variable having c to be negative, taking negative does not make any sense because I already know this probability is 1, why you need a bond, right and then the next inequality is called Chebyshev inequality.

(Refer Slide Time: 17:23)



So, the Chebyshev inequality exactly tells how much, what is the probability that my random variable X will be taking value away from the mean by certain amount. So, let me, so suppose if you are interested in asking this question. Another thing I am often going to use sigma square to denote variance of a random variable. So, when I want to make specific that this variance is associated with the random variable, I will put X like this and if I am not specifying, I mean, always I do not need to specify which random variable it is associated with in this, I will simply write it as sigma square.

So, any d greater than 0, you want to make it then we can write probability that X minus mu greater than d is upper bonded by sigma square by d square. So, what is this basically saying? Mu is what? Mean of, this X and sigma square is variance of this X. Now if I ask the question, what is probability that the value taken by random variable A, when I compare it with its mean value, the probability that it will be away I means the difference of the X and the mean is at least d. What is probability? This statement says, that this probability is upper bonded by the ratio of variance divided by d square.

So, do you see in anywhere like this quantity over here can be derived from this? In what way? How?

Student: d square (())(20:50).

Professor: So, this is I have deliberately put absolute value, If it is absolute value, if I am going to take square on both sides, the other is preserved and the probability should not change. Then?

Student: Then we apply Markov's inequality on that particular (())(21:08).

Professor: So, now if you apply Markov inequality, how we are going to apply? You are going to treat this quantity here as let us say Y and this quantity is a positive random variable because this is the squared value. Now, what is the expected value of that? So, when I am taking square, I really do not need to worry about this absolute value here, right. And what is c? So, c here should be what? d or d square?

Student: d square.

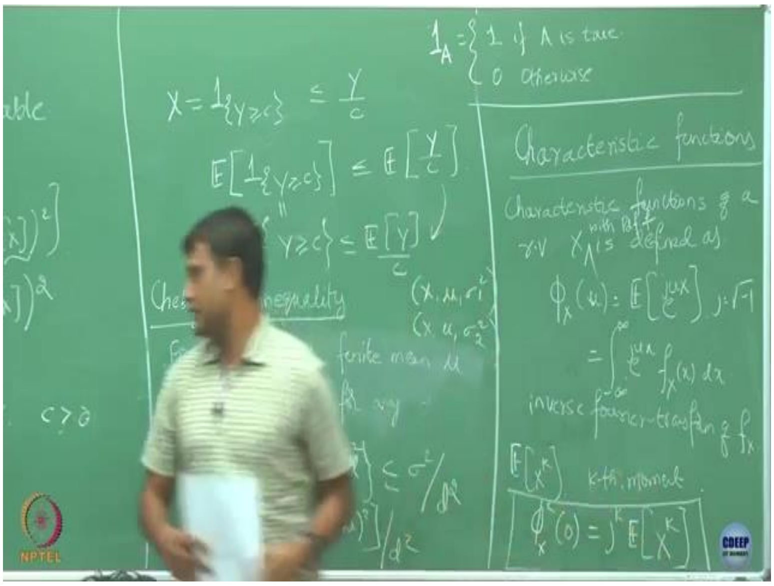
Professor: d square, now, what is this expectation of X minus mu whole square?

Student: (())(21:53)

Professor: That is variance and that is exactly this So, you see that like, the Chebyshev inequality here, whatever I have is implied by Markov's inequality. Now, what it is saying, suppose let us take two scenarios. Let us take two scenarios, let us say I have random variable X with a mu and let us say, so let us say I have two random variables X1 with mean mu and a variance sigma 1 square. And I will take another case where I will change this variance to sigma 2 squared, while keeping mean safe. So, by our understanding of this variance, it should be such that in this case my X, the spread of my X around mean is supposed sigma 2 is larger than sigma 1.

So, by our understanding vaguely at his definition, that variation is more. So, if sigma 2 is more than sigma 1, in this case, this X should be about, its mean value, it should be like spread. So, in that way, if I am going to increase the sigma square here, this probability is increasing, so, that means that me looking at this X being spread around mu, then that probability is also larger, so in that way, this Chebyshev inequality is kind of capturing how your random variable X spread about the mean value. So we will see more about this when we look into specific distribution.

(Refer Slide Time: 24:24)



So, next I am going to look at something called Characteristic Functions. Which is my standard (())(25:10) number and this is nothing but, so, let us take a random variable which has a PDF. That means X is here continuous random variable. I am going to define its characteristic function as for any u that is given, it is going to be defined like this, expectation of e to the power ju and here X here. So, this is the random quantity here, u is fixed j is fixed, j is your simple Euler number, complex number and by definition our expectation this is the value of characteristic function. So, now if you are going to just look at this f of X as some function given to u, what is this quantity indicates? You should look into like Fourier terms, in terms of Fourier terms. So, how you guys recall what is a Fourier transform? So, Fourier transform is defined for a function, give me a function I will define the Fourier transform of that, so that I can go from frequency domain to time domain and vice versa. So, if I am going to treat f of X, as my pdf as a function here what is this quantity here?

Student: f of (())(27:03).

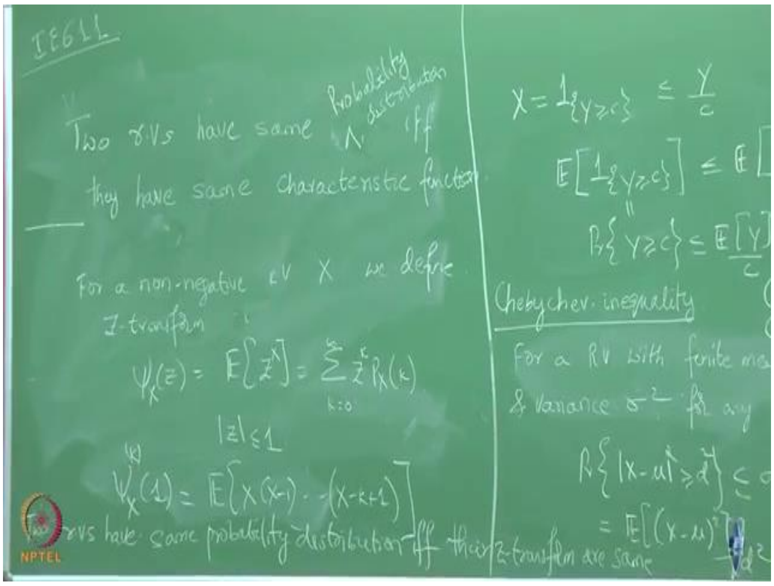
Professor: So, this is nothing but the inverse Fourier transform of your function f of X and also there is (())(27:24) pi factor here which we have not considered here. So, this is just a definition okay what is the consequence of this definition? Why this is useful? So, earlier I told that we are going to say that this is X to the power K, I do not know if this exists or not but this I am going to call it as Kth moment, because K is equal to 1 correspond to expectation K is equal to, to correspond to the second moment like that. So, by the way, if I said my variance is finite, suppose, let us say is my second moment finite? So, if I say this quantity is finite, variance is finite, it is given as the difference of these two quantities. It must be the case that this should be finite, this must be finite otherwise that quantity is not going to finite. So, its variance is finite, my second moment is already finite, but I do not know like the higher moments will be finite or not, but at least if I say as long as variance is finite, a second moment is finite. You will also see that if your second moment is finite, also implies that your first moment is going to be finite.

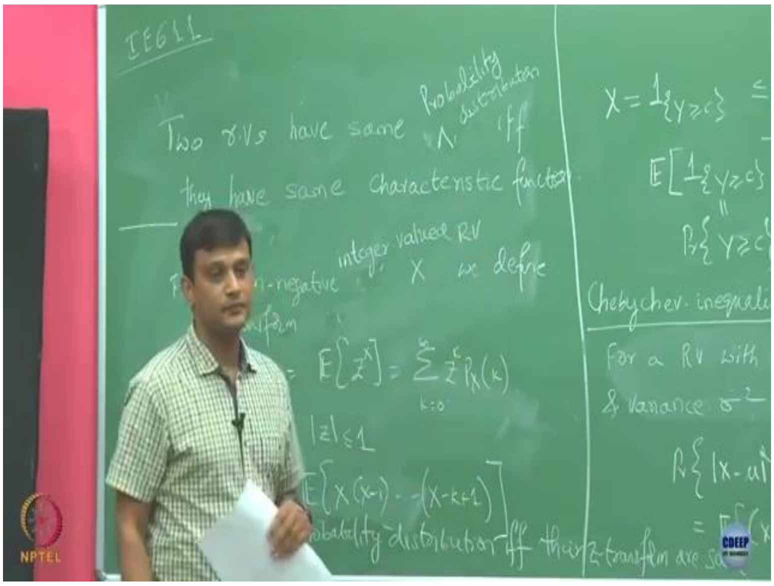
Now, if I want to find the Kth moment of random variables, then this character function comes handy. How is that? So, suppose let us say we have, I know this characteristic function of a random variable X, all I need to do is take the Kth derivative of this function and compute its valued at u equals to 0, then that is going to give me the Kth moment. So, I am going to just write it. So, this relation holds. So, what is this when I write like this v of K with the superscript K, that means, this is Kth derivative and I compute it at 0 this is simply going to be j to the power K and expectation value, expected value of X to the power K.

So, that means this quantity is simply going to be related to the Kth moment directly. So, you can directly see that, right like if you have like this, if you are just going to differentiate this with respect to u every time you differentiate this with respect to u, our X factor comes and every time second derivative another X factor comes after K derivative, X to the power K factor will come. And when you plugged in u equals to 0, this term vanishes, and you will end up with this, just this is computation just look into that.

So, one good thing about this is why this is a form of interest. The character function is, every pdf has a unique characteristic function. So, if you know pdf function, you already know how its character function is going to look like. So, if you know already characteristic function, you also, you can immediately say this character function belongs to this pdf function.

(Refer Slide Time: 31:41)





So I am going to say that two random variables have same pdf if and only if they have same or instead of probability distribution have same characteristic functions, I think this is slightly stronger than, not just pdf it is going to be probability distribution. So, if you have a probability distribution, and you are going to complete its character function, that is going to have a unique one. And if you have a characteristic function you can associate it with a unique distribution that way. So, this is we are going to use later, when we are going to do some theorems like, I think we are going to use it to when you prove Central Limit Theorem and all.

So, this is for a case where I have a pdf. And we can also define similarly for discrete case. So, for a special case when my random variable is non-negative, for a non-negative random variable X we are going to define Z transform which is defined as, v of X of Z is equals to expectation of and where I want this region of convergence range to be less than 1 and again, we can show that for this is for a discrete random variable, which is taking non-negative value, again we can show that v of X1 is nothing but expectation of X, X minus one X all the way to X minus K plus 1.

So, here it was straightforward that in the characteristic function, if you take the Kth moment and put compute at 0, it was directly related to the Kth moment, but here in Z form, it is not so straightforward related. So, I am going to, if you are going to take the Kth derivative of this function in Z, and then compute it at the value Z equals to 1, this is what you are going to get. And you see that this expectation is now involves all the moments, if you want to just expand the X, X minus 1 all the way to X minus K minus 1, it will going to have expectation of X to the power K, expectation of X to the power K minus 1 and it will also have expectation of just X. If you just expand this and again two random variables have same probability distribution if and only if the Z transform are same, now here in this case X is discrete random variable, non-negative, we are just going to say, sorry, non-negative, I have to say integer random variable.