Dealing with Materials Data:Collection, Analysis and Interpretation

Professor. M P Gururajan
Professor. Hina A Gokhale

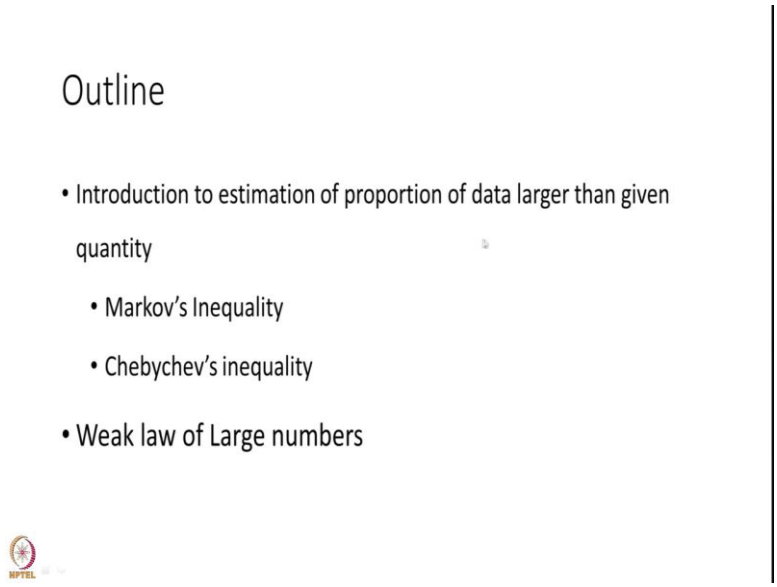
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture-07

Random Variable and Expectation 4

Hello and welcome to the course on Dealing with Materials Data. In the present sessions, we are working with random variable and its expectation and today what we are going to work on is little different form that. Anybody who has a data or is dealing with a situation or an event, always has a question as to what proportion of data would be larger than certain value, or what proportion of data would lies between this two bounds, or what proportion of value data will be more then or less then certain value.

So, these kind of questions are being answered in statistics by several inequalities and this, in today's session we are going to talk about two such inequality and one such law which is called which is called a Weak law of large numbers.

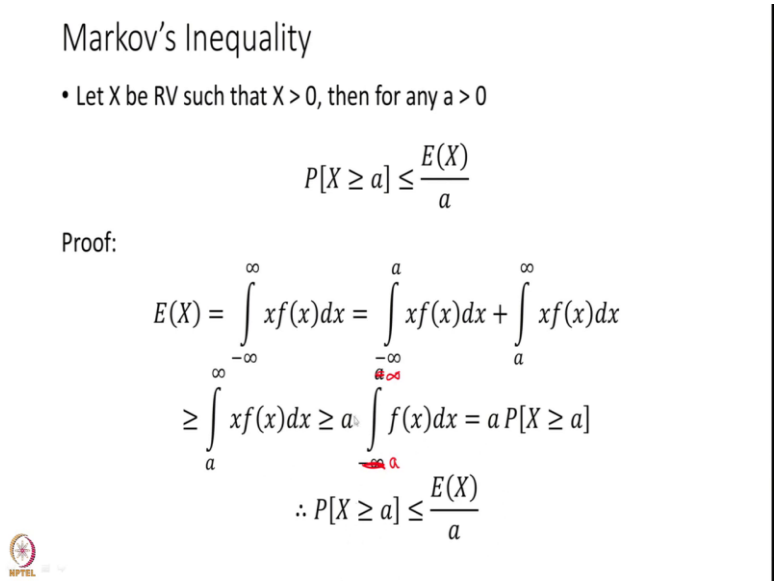
(Refer Slide Time: 01:17)



So, in outline let us see, we would like to introduce the estimation of proportion of data larger than the given quantity. We would like to do this through Markov's inequality, secondly

Chebychev's inequality, these are all improvement on one or the other and finally, we will introduce what is known as Weak law of large numbers and let us start.

(Refer Slide Time: 01:49)



So, Markov's inequality is define as let X, if X is an random variable such that it is a positive random variable and there is any number a which is positive,

$$P[X \geq a] \leq \frac{E(X)}{a}$$

The proof is very simple expected value of X

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{a} xf(x)dx + \int_{a}^{\infty} xf(x)dx$$

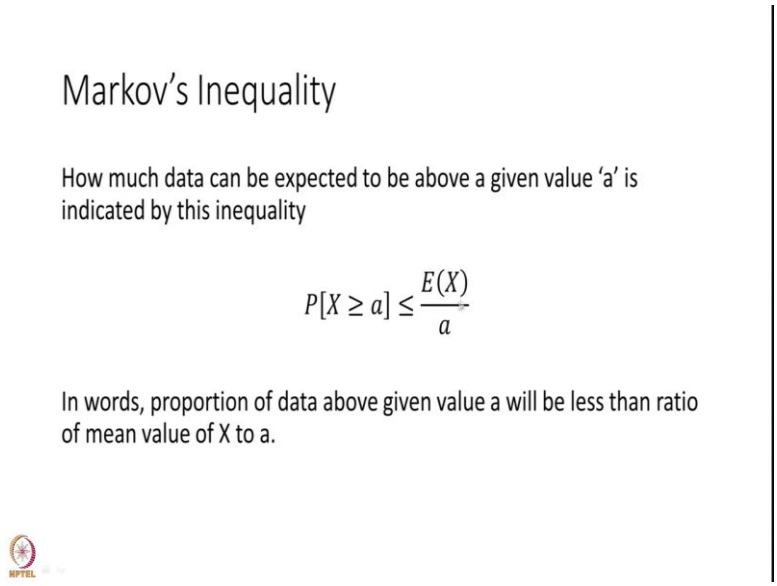
Now, because all the quantities are positive here each of this summation is positive, therefore, if you drop one of the quantities, it will be smaller than the given quantity expected value of X, so you get

$$E(x) \geq \int_{a}^{\infty} xf(x)dx \geq a \int_{a}^{\infty} f(x)dx = a\,P[X \geq a]$$

$$\therefore P[X \geq a] \leq \frac{E(X)}{a}$$

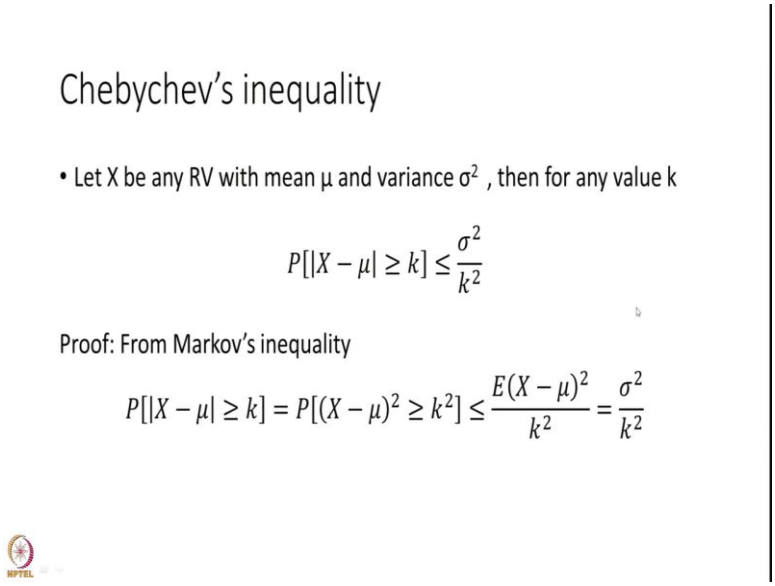
I think I have made a mistake here, this has to be a to infinity, so please make the correction, I will make it right here, let us do that.This should have been a to infinity just as this case and therefore, we have expected value of a is equal to probability that or is equal to a multiplied by probability of X greater then equal to a. I hope I said expected value of X and not expected value of a, so it is expected value of X which is greater than a times probability of X greater than or equal to a and which proof the inequality.

(Refer Slide Time: 04:02)



So, what it really says is that how much of data can be expected above a given value a, this bound is given by this, it cannot be larger than expected value of X or the it cannot be larger than the ratio of expected value, mean value of X divided by a quantity a. This is called Markov's inequality.

(Refer Slide Time: 04:28)



We move to another inequality, which might be known to you in other cases also, which is called Chebychev's inequality. Again X is the random variable, but now we are not assuming that it is a positive random variable, so we are generalizing the case. It has a mean and the variance sigma square, then for any value k, we are again not putting a restriction of k being greater than 0 as in the case of Markov's inequality, this is for any value k,

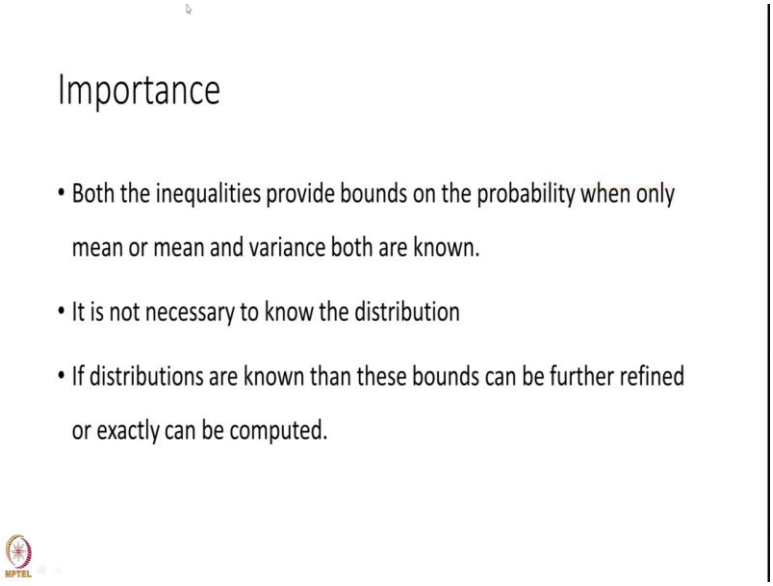
$$P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

This can easily be derived as shown here from Markov's inequality,

$$P[|X-\mu| \geq k] = P[(X-\mu)^2 \geq k^2] \leq \frac{E(X-\mu)^2}{k^2} = \frac{\sigma^2}{k^2}$$

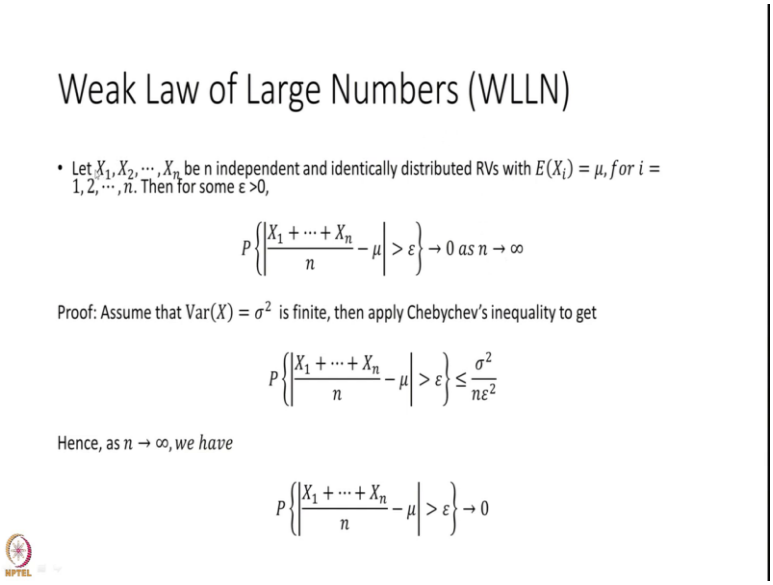
I think here instead of saying any value of k, k is implicitly coming out to be positive, so that may be noted.

(Refer Slide Time: 06:01)



So, both this inequalities provide bound on the probability when only mean or mean and variance are known to you for any random variable. In Markov's case it is only a positive random variable. In the case of Chebychev's inequality we do not put such a restriction. Please note that no distributional assumptions have been made, it is a normal distribution of course this distribution have to be introduce later, but no distributional assumptions are made, but we should know that, if you know the distribution this bound can be further refined and can be exactly computed.

(Refer Slide Time: 06:50)



From this we move, because these inequalities will be useful for us to proof a Weak law of large number. What does a Weak law of large number says? Its say that if X_1, X_2, X_n are independently an identically distributed random variables. We call it IID, so they are independently identically distributed random variable with expected value of E($\mathbf{X_i}$) is mu, for the all the i's and for some epsilon which is greater than 0 .

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right\} \to 0 \; as \; n \to \infty$$

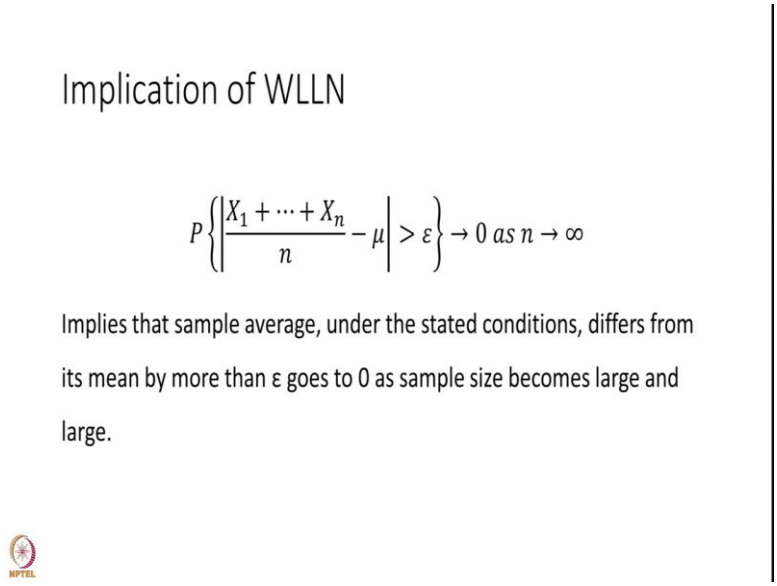
It means that, the difference between the average of these random variables X_1, X_2, X_3, X_n and its common mean value, almost vanishes as n becomes large an large, as your sample, in future you are going to called X_1, X_2, X_3, X_n a sample of size n.

So, from a random variable X, you take a sample of size Xn which is your data, so as you are data becomes larger and larger the arithmetic average, the average of these sample is going to be very close. The difference between this average and the mean common mean value is going to diminish as n becomes large an large. The proof is rather simple. We have to make one assumption that there is this random variables have a variances, sigma square which is finite.

See there are cases in which you have variances also, so here we are defining that assume that variance of all X is sigma square which is finite. Then we apply the Chebychev's inequality to this and then if you apply the Chebychev's inequality, it is obvious that as n tend to infinity this quantity tends to 0.

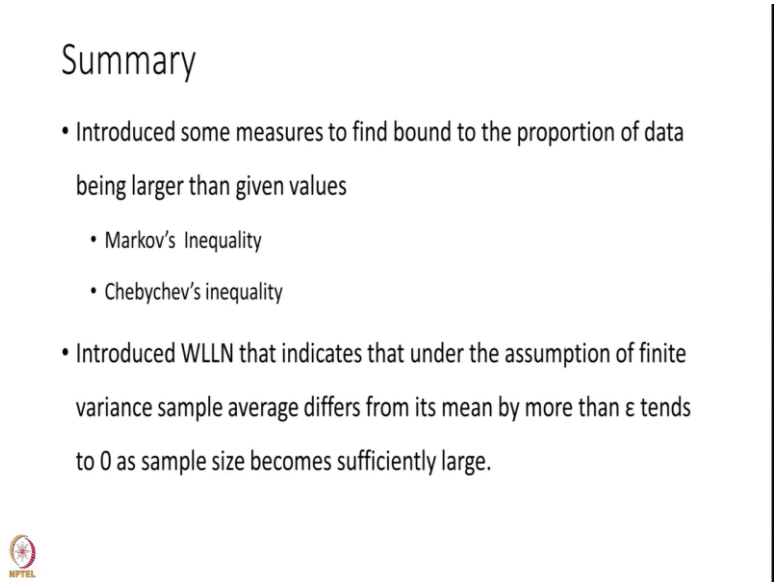
So, this once again, this is a Weak law of large number. Let me tell you why it is called a Weak law, because it is the probability which tends to 0 and not actually the difference which tends to 0. Therefore, you call it a Weak law of large number. We are going to have a central limit theorem, which is a stronger law of large number, we will come across when we move further in this course.

(Refer Slide Time: 10:20)



So, what is the implication? What it says is that as I said before, the sample average under the stated condition. By stated condition I mean that the common variances of X_1, X_2, X_3, X_n is exist and it is finite, then it defers from its mean value by more than epsilon or rather the difference between the sample average and the mean value as n tends to infinity, it eventually diminishes. So, if it is more than epsilon that probability goes to 0 as the sample become large and large.

(Refer Slide Time: 10:53)



So, let us summaries what we have learnt today, we have introduce some measures to find bound on the proportion of data being larger than given quantities. The first inequality we introduce was Markov's inequality, in which we assume that the random variables are positives. In the Chebychev's inequality, we removed the, restriction of random variable being positive, we said that it could be any random variable with having a mean value mu and a common variance sigma square and then we gave an inequality in which we showed that the difference between the mean and the arithmetic average actually the random variable X actually is smaller than the ratio of variances and the quantity k above which you are looking for it, looking for the value to be large or let me clarify very specifically, it means, it tries to say that, the random variable X minus its mean value, the absolute difference is large than a value k. This proportion is smaller than the variance of X divide by k square.

This is the bound it gives and finally we introduce a Weak law of large number, which indicates that under the assumptions of finite variances, the sample average defers from its mean value more than any small value epsilon, it tends to 0, as sample size becomes sufficiently large. So, as sample size is large, the difference between the sample average and the mean value diminishes. With this we cover up the complete session on the random variable and its expectations and next we will move on to the special distributions.

Thank you.