Dealing with Materials Data: Collection, Analysis and Interpretation

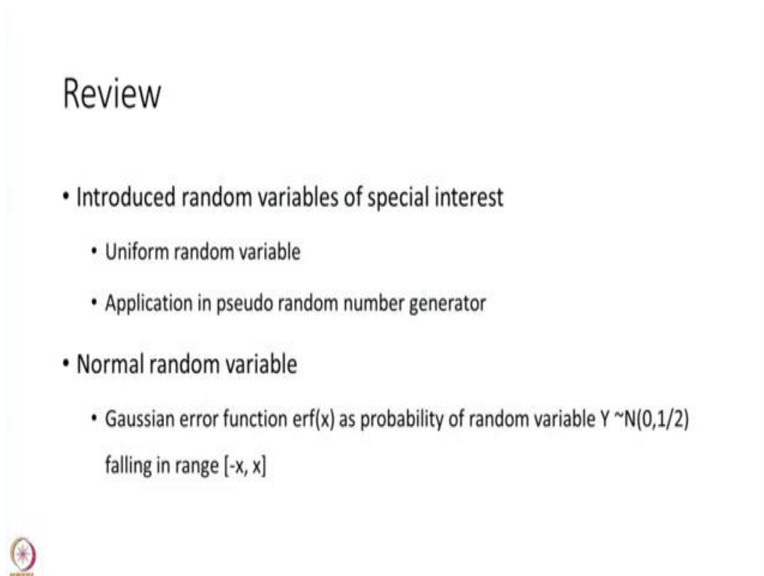
Professor. Hina A. Gokhale
Department of Metallurgical Engineering and Materials Science,
Indian Institute of Technology Bombay

Lecture 33

Special Random Variables 3

Hello and welcome to the course on dealing with Materials Data. In previous few sessions, we are working with a variety of distributions, model distributions that may arise through the experiments that we conducted, that is the results of the experiment may follow a certain kind of pattern of distribution and this is what we want to call them, we call them as Special Random Variables.

(Refer Slide Time: 00:53)



So, far we have introduced two kinds of special random variables, one is a uniform random variables, these are all the continuous in nature, if you recall back we introduced the Discrete Random Variables, special random variables such as discrete uniform distribution, binomial distribution, Bernoulli trial, negative binomial etcetera, etcetera.

Then we introduce a continuous type of special random variables. So, in the previous session we introduced uniform random variable, even showed that it is useful for pseudo random, is useful in random number generation, and it uses the pseudo random number generator to generate any random, pseudo random number that we wish to have from any distribution.

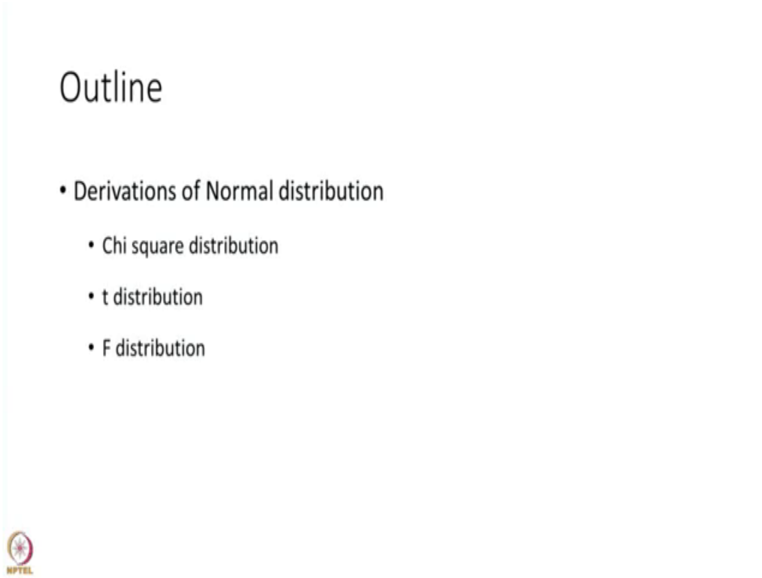
Then, we introduced a normal random variable, we said that this random variable or this distribution was noticed. Actually, it was noticed during the Galileo's time when he was very upset that under the same circumstances, his students are bringing out different observations of stars and constellations.

But he took around 200 years for Gauss to come and say that errors are normal to occur, they are natural to occur and they have a certain kind of distribution. And that distribution we called it a normal distribution or a Gaussian distribution. We also said that the Gaussian error function is directly related to the normal random variable.

That is Gaussian error function actually shows the probability of a random variable Y falling between minus x and plus x when Y is distributed as a normal with mean value 0 and a standard deviation or the variance as 1 over 2 or variance as a half.

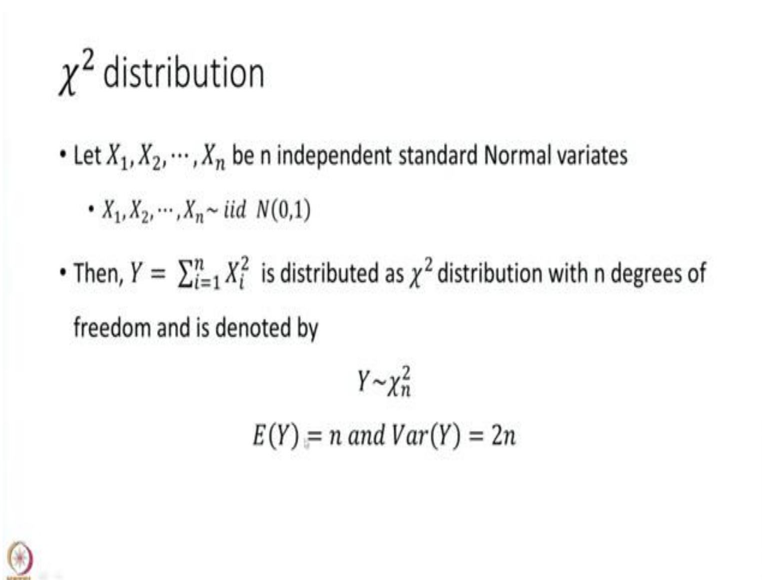
Now, what we want to do this time is we want to introduce some of the distributions which are going to arise in future, that is when we deal with assumption that certain data comes from a normal distribution and we do certain inference on the data, we do the estimation of the data. We do the confidence interval estimation of the data or we do the hypothesis testing, all this we are going to do in the future, when we do these things on the data, we need certain distribution in order to estimate the probabilities or the estimate the uncertainties.

(Refer Slide Time: 3:34)



So, this time we are going to, in this session we are going to introduce three such distributions, which are derived from the normal distribution. They are Chi square distribution, t distribution and F distribution. So, let us start, as I described already in the inference for the, to make the inference from the observations coming out of normal distribution we need certain kinds of distributions and these are the distribution we want to describe here.

(Refer Slide Time: 4:10)



First we go by Chi square distribution, let us say that we have n random variables, which are independent and they are standard normal variance, it means that they are all independently distributed and identically distributed as a normal distribution with mean 0 and variance sigma square.

$$Let\, X_1,\ X_2,\ \cdots, X_n\ be\ n\ independent\ standard\ Normal\ variates$$

$$X_1,\ X_2,\ \cdots, X_n \sim iid\ \ N(0,1)$$

$$Then, Y = \ \sum_{i=1}^{n} X_i^2$$

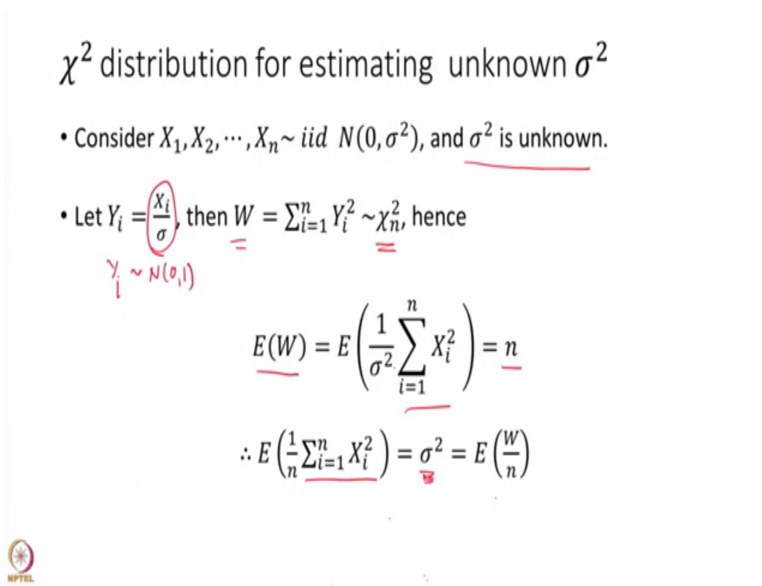
$$is\ distributed\ as\ \chi^2\ distribution\ with\ n\ degrees\ of\ freedom\ and\ is\ denoted\ by$$

In that case.

$$E(Y) = n\ and\ Var(Y) = 2n$$

the expected value of Y is n and variance of Y is 2n and this is very interesting, variances exactly double the time, double the expectation value of Y.

(Refer Slide Time: 05:21)



This distribution is useful to estimate the unknown variance sigma square of a normal distribution. So, let us consider that you have X1, X2, X3…Xn coming from a normal distribution with mean 0, but variance sigma square, remember when we defined the Chi square we said sigma square is 1. Here we are saying that there is a sigma square and we do not know, the sigma square is unknown.

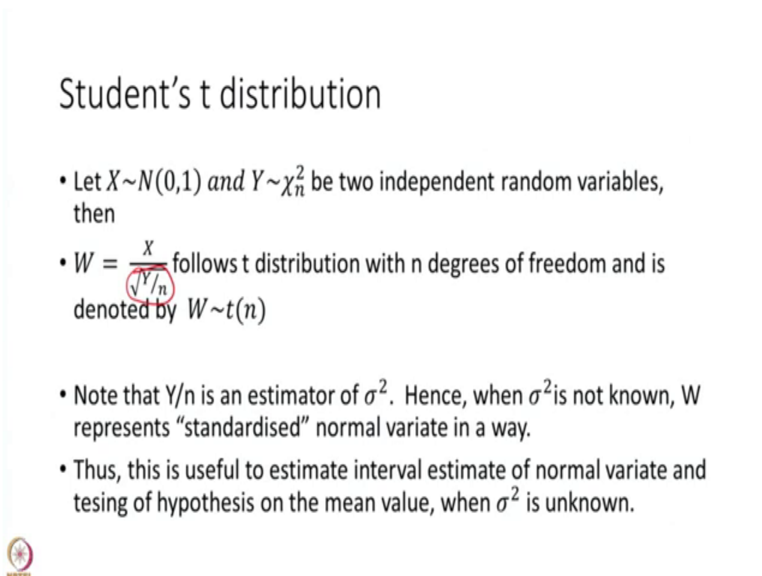
$$Let\, Y_i = \frac{X_i}{\sigma}, \; then\, W = \sum_{i=1}^{n} Y_i^2 \sim \chi_n^2, \; hence$$

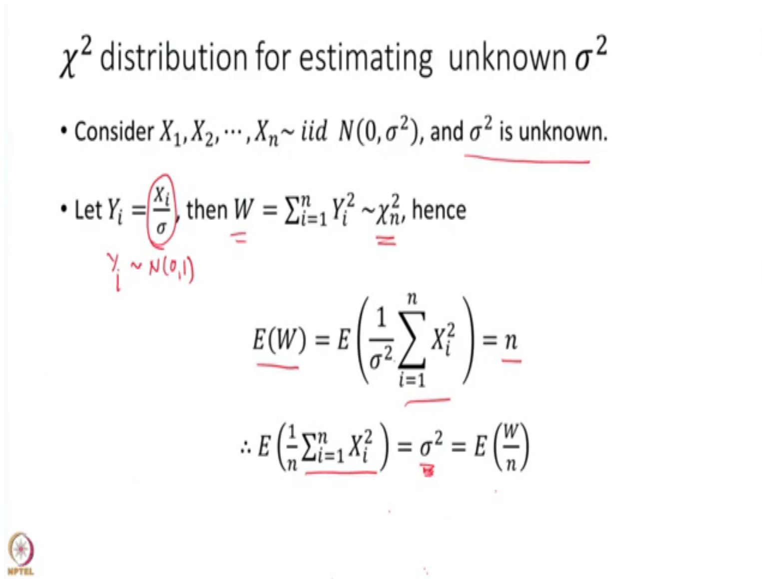
$$E(W) = E\left(\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2\right) = n$$

$$\therefore E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \sigma^2 = E\left(\frac{W}{n}\right)$$

And this is the way, if you have n random variable coming from a normal with 0 mean and sigma square and sigma square is unknown, Chi square distribution is useful to estimate the value of sigma square, it gives you expected value of 1 over n summation Xi square as a sigma square.

(Refer Slide Time: 7:45)





Next we move on to Student's t distribution. This distribution is also known as a student's t distribution, because the person who, means the t distribution in he took, he took the ten name of student and therefore it is called a Student's t distribution. Let X be distributed as ~N(0,1) and Y be distributed as a Chi square distribution with n minus 1 degrees of freedom and these X and Y are two independent random variables.

Then you defined a new random variable W as a ratio of X divided by square root of Y divided by n. Then this follows a t distribution with the same n degrees of freedom and it is denoted as W distributed as small t, n degrees of freedom.

Sometimes n is also written as a subscript, it is just a case of notation. Kindly note one thing, that there is a square root in the bottom, there is a square root, while taking the ratio we have taken a square root. You can imagine that if X is a result of a experiment it has a unit connected to it and Y if you look at it as a Chi square distribution, the unit of the random variables coming up for Chi square distribution will have exactly the square of the unit of X.

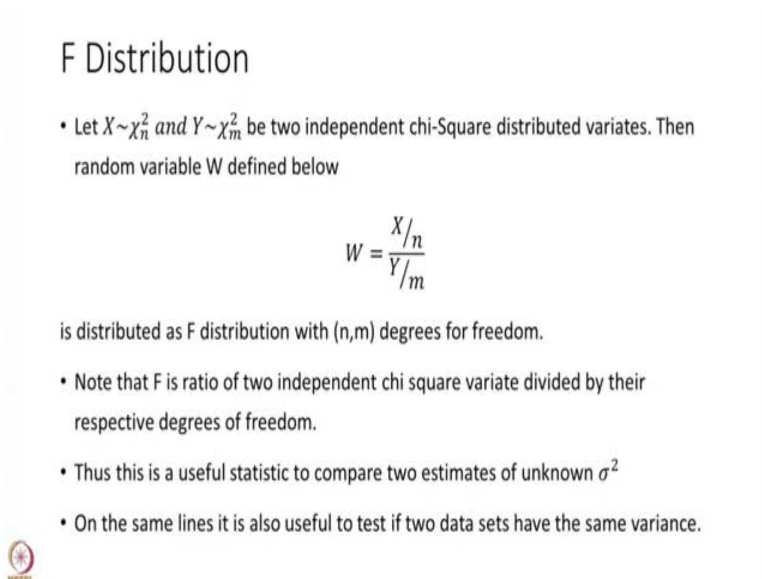
Please recall, let us go back, you have X random variables with some unit say l in that case, the W which gives you Chi square distribution has a l square unit and therefore, in the case of student t distribution, remember that it is unit less because you are dividing it by the square root and that is the reason you are dividing it by square root and therefore, W is equal to X divided by square root of Y by n.

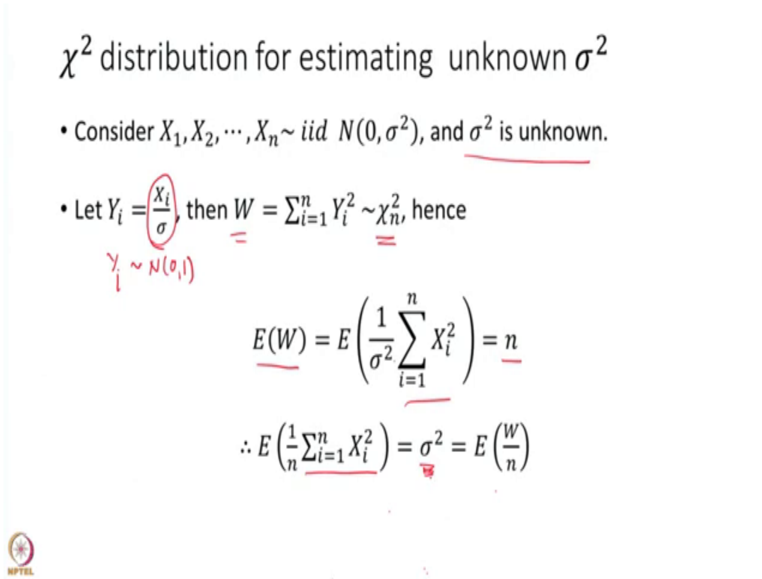
Easy to remember if you go by the units you know that t is not supposed to have any units and therefore, t distribution it has to be divided by the square root. So, such a W follows t distribution with n degrees of freedom, if Y by n as we discussed previously is an estimator of sigma square, then sigma square is n, if sigma square is not known, W either way represents the standardized to normal variate in a way. This is what I described in a different way.

If Y by n, suppose, you assume that X is not the, it is not a standard normal variate, it has a variance sigma square which is unknown. Then what we are trying to say is that if Y by n is an estimator of sigma, then by taking X divided by square root of Y by n you are kind of standardizing the variable, you are standardizing the normal variate in order to get or distribution and that distribution is t.

So, this distribution is useful to estimate an interval estimate of a normal variate or a normal mean value when the sigma square is not known. We will go into detail in future when we come to the estimation, parametric estimation.

(Refer Slide Time: 11:40)





The next distribution we would like to introduce is an F distribution. Let X a random variable be distributed as a Chi square with the n degrees of freedom and Y be another random variable distributed also as Chi square with m degrees of freedom. These are two independent Chi square distributed variates, then the ratio W of X by its degrees of freedom divided by Y divided by its degrees of freedom, is said to be distributed as a F distribution with two degrees of freedom, numerator degrees of freedom n and denominator degrees of freedom m.

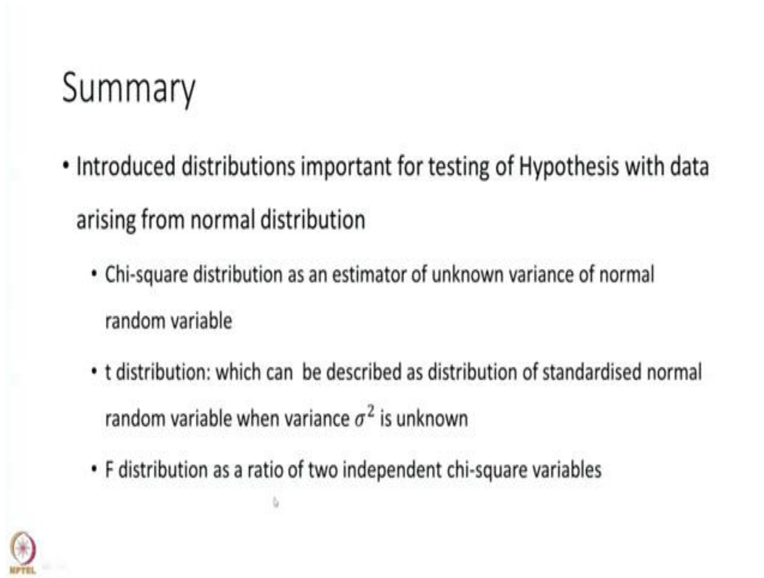
If we look at X by n and suppose, X by n is also an estimator of a variance, unknown variance sigma square of a normal variate, as we derived in the past, if you recall, we said that W by n, when W is a Chi square distribution with n degrees of freedom in certain circumstances, W by n is an estimator of a variance sigma square, unknown variance sigma square.

So, here what it says is that F is a ratio of two independent Chi square variates. It is two independent Chi square variates and therefore, if both the Chi square variates are estimating the same unknown sigma square, then this becomes a test to see if the two variates are same or they are different. This is what is going to come up when we learn analysis of variance, this distribution is going to play a very important role, it also plays important role to test the hypothesis that the two data sets have the same variance.

Suppose, you conduct an experiment, two independent experiment and you feel that the variation, variance in the two experiments are very similar. Then to test such a hypothesis, you need an F distribution. So, with this we introduce the three distributions derived from normal distribution, please remember, for normal distribution, we gave that it arises naturally in experimentation.

Why this Chi square t and F distribution do not necessarily arise naturally from performing any experiments, but they arise, they are useful for the future inference, when we may on different kinds of normal, data arising from a normal distribution and therefore, we introduce the theory distributions here.

(Refer Slide Time: 14:49)



Chi square distribution as an estimator of unknown variance of a normal random variable, t distribution, which can be described a distribution of standardized normal variable when variance sigma square is unknown, and F distribution as a ratio of two independent Chi square variables,

Thank you.