Dealing with Materials Data: Collection, Analysis and Interpretation

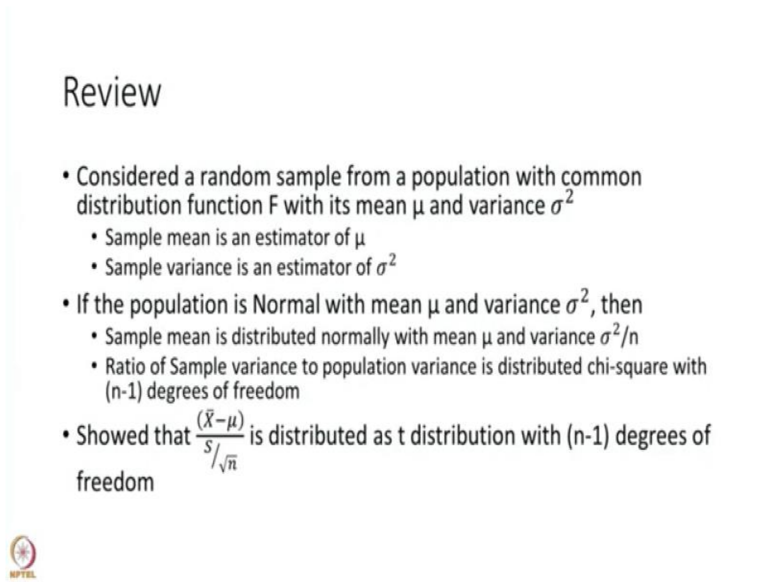
Professor. Hina A. Gokhale,
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay

Lecture 49

Sampling Distribution 3

Hello and welcome to the course on Dealing with Materials Data. Currently we are going through the sessions on describing Sampling Distributions.

(Refer Slide Time: 00:33)

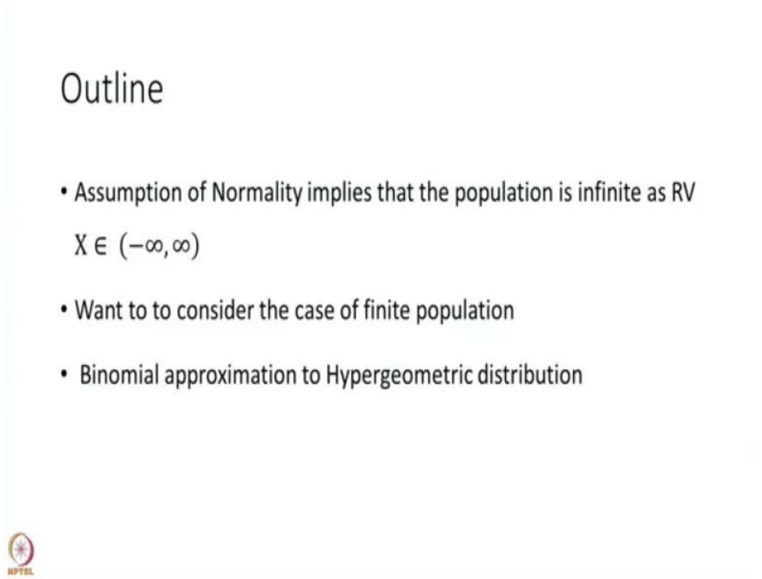


We have considered a random sample from a population with a common distribution function F with a mean value mu and variance sigma square. Then we found that sample mean is an estimator of population mean mu and the sample variance is an estimator of population variance sigma square.

Then we made an assumption that the population itself is a normal with a mean value mu and variance sigma square and then, in the last session we saw that the sample mean itself in that case is distributed as a normal with mean mu and a variance sigma square by n, where n is the size of the sample and the ratio of sample variance to the population variance is distributed as a chi-square with n minus 1 degrees of freedom.

Then we went ahead and showed that if you take the ratio of difference of sample mean from the population mean and divided it by the estimated population variance by sample standard deviation, then it is distributed as a t distribution with n minus 1 degrees of freedom.

(Refer Slide Time: 01:57)

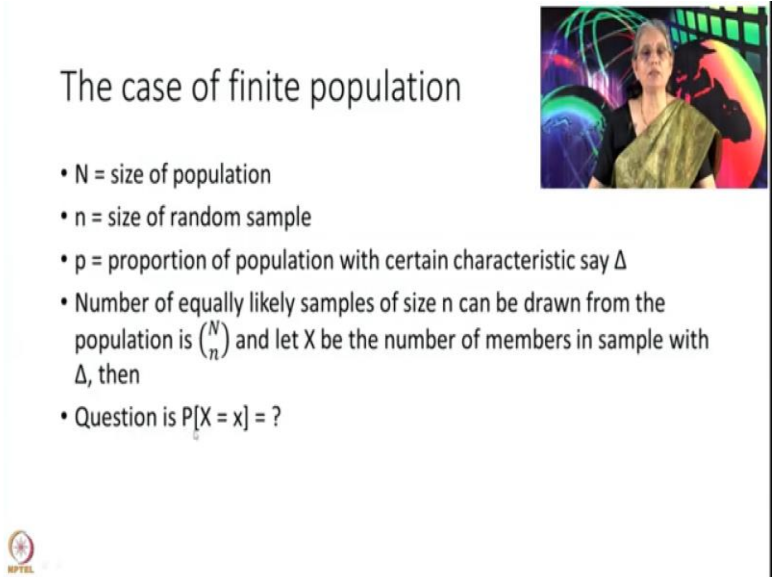


Now you note that in all this case we have assumed normality in the previous case and derived certain distributions. Now normality again assumes that the random variable, the population size is infinite. The random variable varies from minus infinity to infinity and it can take any value. So, your population itself is a kind of infinite population.

What if the population is finite and this is the case we would like to study here through an example. We are going to consider the case of a hyper geometric distribution for a finite population size and we would also like to show that if this finite population size is very large then it can be approximated by a binomial distribution.

Remember here, the previous approximations that we have talked about is binomial distribution approximated as exponential or binomial distribution approximated as a normal distribution in which the number of Bernoulli trials were considered large. But now we are going to consider the population itself as large. Remember there we were considering the sample size tending to infinity, here we are going to consider the population size itself tending to infinity.

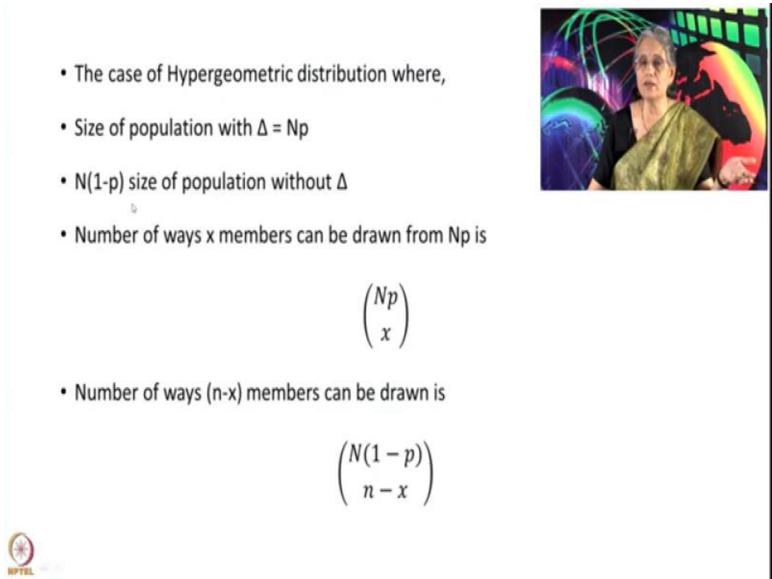
(Refer Slide Time: 03:42)



Let us start, capital N let us say is a size of population, small n represents a size of a random sample as always, p is the proportion of population with a certain characteristic say delta. For example, its number of defectives, its number of successes. So, the characteristic could be defective, characteristic could be success, so there is some characteristic delta and the, we know that the population in the population p proportion of population has a characteristic delta.

The number of equally likely samples of sizes n that can be drawn from the population is obviously capital N choose small n, it is capital N choose small n and let X be the number of members in the sample with a characteristic delta. The question we are trying to answer is what is the probability that the capital X takes a value small x, what is the probability that X is equal to small x?

(Refer Slide Time: 04:56)



So, let us work it out it is a case of a hyper geometric distribution if you recall the introduction of discrete distribution functions we talked about hyper geometric distribution, so the size of population with characteristic delta would be capital N multiplied by p, because p is the proportion of population which has a characteristic delta.

So, the size of population with a characteristic delta will be Np and therefore obviously N times 1 minus p is the size of population without delta. Now x is chosen x members are chosen from the population, because x members have the characteristic delta.

Number of ways x members can be drawn from Np is

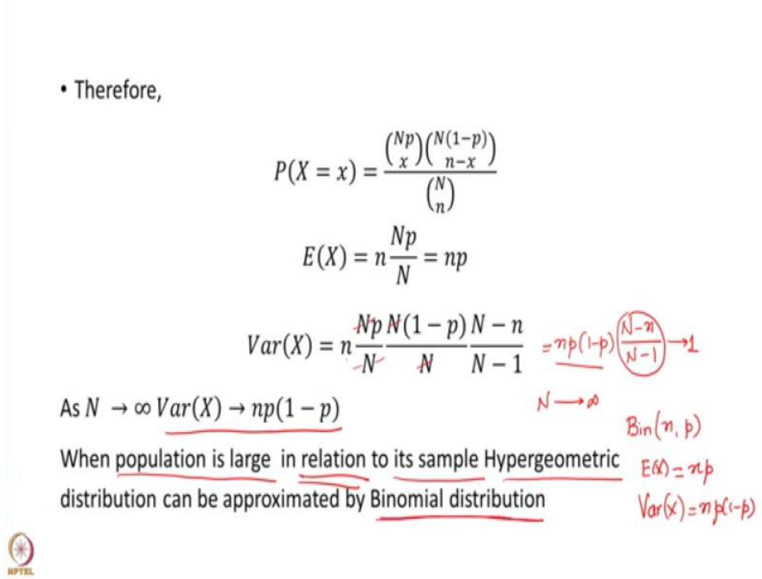
$$\binom{Np}{x}$$

So, this many ways you can select the x members of the population having a characteristic delta.

The number of ways rest of the members have being chosen or have being drawn, they must have come from n times 1 minus p population, because they do not have the characteristic delta and therefore that has to be

$$\binom{N(1-p)}{n-x}$$

(Refer Slide Time: 06:36)



And therefore probability of x is equal to x becomes

$$P(X = x) = \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}}$$

So, if you apply the hyper geometric distribution expected value of x

$$E(X) = n\frac{Np}{N} = np$$

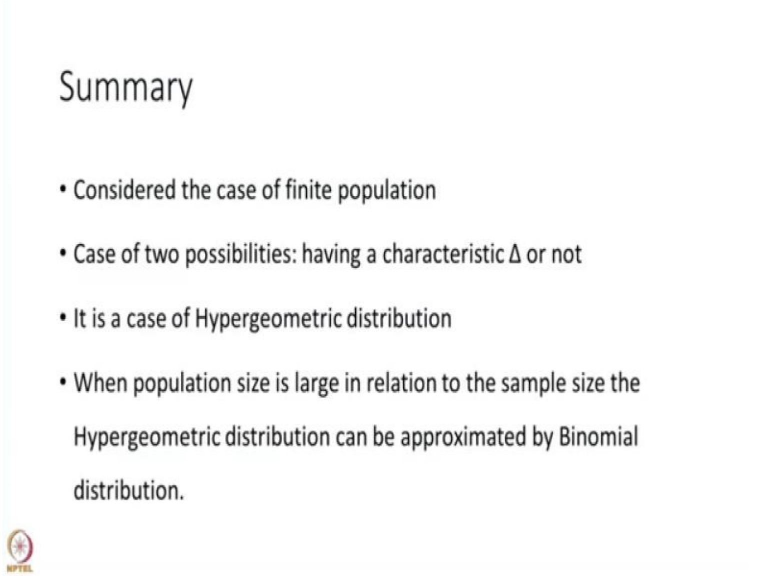
$$Var(X) = n\frac{Np}{N}\frac{N(1-p)}{N}\frac{N-n}{N-1}$$

As $N \to \infty$ $Var(X) \to np(1-p)$

This is very easy to show, I will leave it to you to show it, so this tends to 1 and therefore it is said that when population is very large in relation to its sample, when the population is very large in relation to its sample, so the sample size is not that large. Then hyper geometric distribution can be approximated by binomial because remember in binomial distribution, in binomial distribution with n, p, the expected value of x is np and variance of X is np 1 minus p. So, this shows that when population is large in relation to its sample because it is necessary here.

Distribution can be approximated by a binomial distribution and then you can further say that when N becomes very large that is you also consider it sample to be very large you can further approximate. In the R sessions you will come across some of the problem that you will solve in which you will need this approximation.

(Refer Slide Time: 10:06)



So now we summarize what we considered today, we considered the case of finite population case of two possibilities of having a characteristic delta or not having the characteristic delta. We showed that it is a case of a hyper geometric distribution and then we showed that when the population size is large in relation to the sample size the hyper geometric distribution can be approximated by a binomial distribution, thank you.