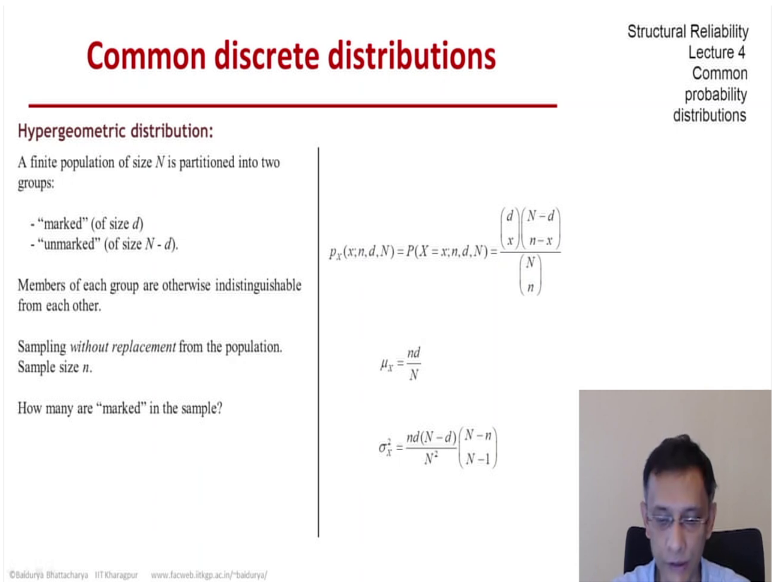
Structural Reliability
Prof. Baidurya Bhattacharya
Department of Civil Engineering
Indian Institute of Technology, Kharagpur

Lecture –39
Common Probability Distributions (Part - 10)

(Refer Slide Time: 00:28)

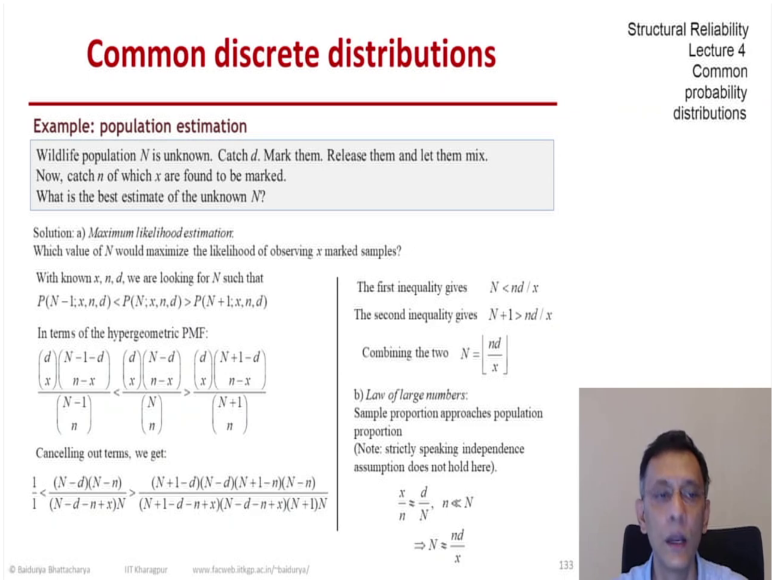


Let us continue our discussion on common discrete distributions. The next one is hypergeometric and we will end this discussion with the Poisson distribution. So, the hypergeometric distribution comes when you are selecting without replacement from a finite population which has two kinds of members marked and unmarked but otherwise these are indistinguishable. So, the question is if we perform a sample without replacement from this population of size capital N.

Capital N is not a random number it is a fixed number and the sample size is small n. So, what is the probability of finding a given number of marked members in the sample? So, that would be by using our definition of classical probability is in the denominator we have N choose n and in the numerator we have we apply the counting principle. So, we choose x from d marked members and n - x from the n - d marked n minus the unmarked members.

So, that ratio is the hypergeometric PMF and you could apply the definition of the expectation perform the summation the mean is the ratio of nd and capital N and the variance is also you can it can be derived in terms of these parameters of the problem. Now there is one classical problem of population estimation that can be solved using the hyper-geometric distribution and we will go through them now.

(Refer Slide Time: 02:38)



So, suppose there is an unknown wildlife population of size capital N we want to estimate it and this is the process that is followed. So, we catch d of them mark them in some manner and then release them and give sufficient time so that this marked members mix uniformly throughout the population. Now I take a sample from this without replacement of size small n and I find out these x of them are marked.

So, I release d marked and x of them have come back. Now what is my best estimate of the unknown capital N there are two ways of doing this the first one would be taking advantage of the maximum likelihood estimation. So, the question simply put is that I have the hypergeometric PMF which is dependent on capital N, d and small n. But for what value of capital N is that PMF at its highest.

So, we are looking for that capital N by with knowing with known x, n and d which would give me this two inequalities. So, p of n would be greater than p of n + 1 with the same parameters x n and d and p of N will be greater than p of N - 1 with those same parameters x small n and d. Now by plugging in the hypergeometric PMF expressions and doing some algebra we come up with two inequalities.

The first one which you see on the right gives me one condition the second one gives me another condition. So, if you combine these two you get a nice solution that the unknown capital N is nothing but the integer the floor value of nd divided by x. There is also another intuitively appealing explanation of this solution and that is by invoking the law of large numbers it is by asserting that the sample proportion approaches the population proportion.

Although the underlying requirements for the law of large numbers to hold is not strictly holding here because the samples are not strictly independent but assuming that small n is small compared to capital N we can still equate the two proportions and I get the same answer for the unknown population size.