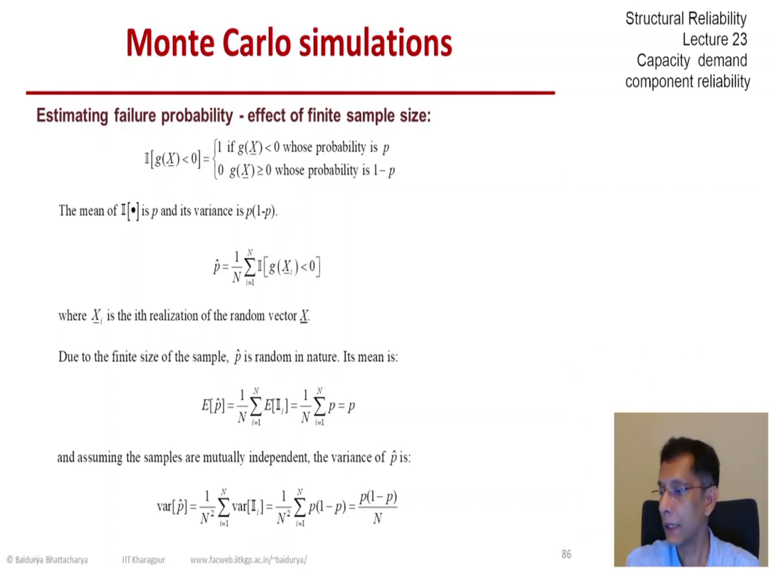
Structural Reliability
Prof. Baidurya Bhattacharya
Department of Civil Engineering
Indian Institute of Technology, Kharagpur

Lecture –184
Capacity Demand Component Reliability (Part 32)

(Refer Slide Time: 00:27)



So it is very good to know that the estimates of p will converge in the limit to the true value of p the failure probability backed by the strong law of large numbers as long as our samples from the join distribution of the basic variables are IID. But obviously we cannot sample forever. So, we have a finite sample size. So, what happens if we stop after a finite number of samples how uncertain is the estimate can we get some sort of confidence interval.

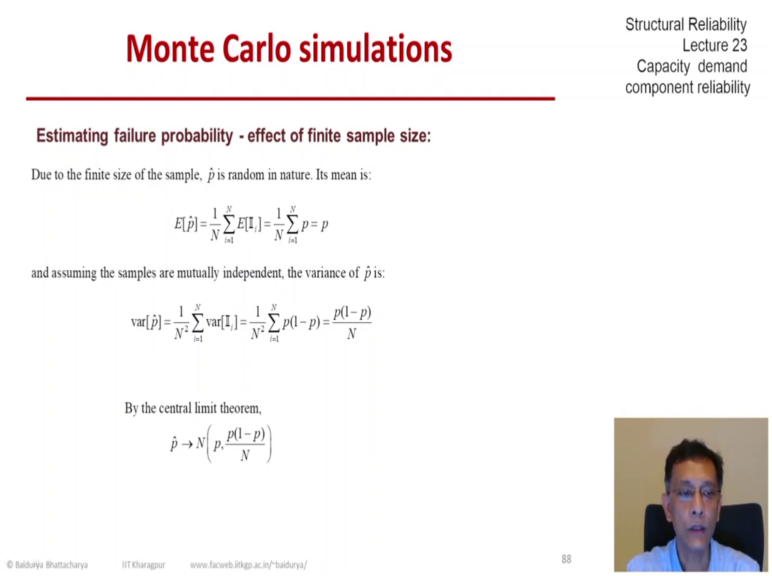
And how many samples are enough to achieve certain accuracy certain desired accuracy so for that we will invoke another very fundamental law but we will come to that in a second. So, to recap we have this indicator function i which evaluates whether failure has occurred or not it is one a failure has occurred whose probability is p and it is 0 otherwise. So, the mean of i is p and its variance is p times q or p times 1 minus p because i is a binary a binary random variable we have set up the scheme in which we estimate p, p hat as an expectation.

So, we generate x from the joint distribution many, many times or n times and then evaluate g every time see if i is one or zero and then add those i's and divide by n. So, we are computing the expectation now because p hat is estimated from a finite sample size of n then p hat is a random variable. So, if p hat is a random variable we would like to know what its mean is what its variance is and. So, on and we have seen such derivations before in this course.

So, the expectation of p hat expectation being a linear operator is what you see on your screen 1 over n the sum of the individual eyes and because the individual eyes are IID they have the same parameter small p. So, the expectation of p hat ends up being p itself. So, that is very good and what would be the variance of p hat if the samples are mutually independent which is what we have been stressing all this while then the variance of the sum is the sum of the variances multiplied by the square of the coefficient. So, that gives us p times q over n.

So, the variance goes down with the sample size. So, that is that is very good now it is time that we introduced the other great theorem are the central limit theorem.

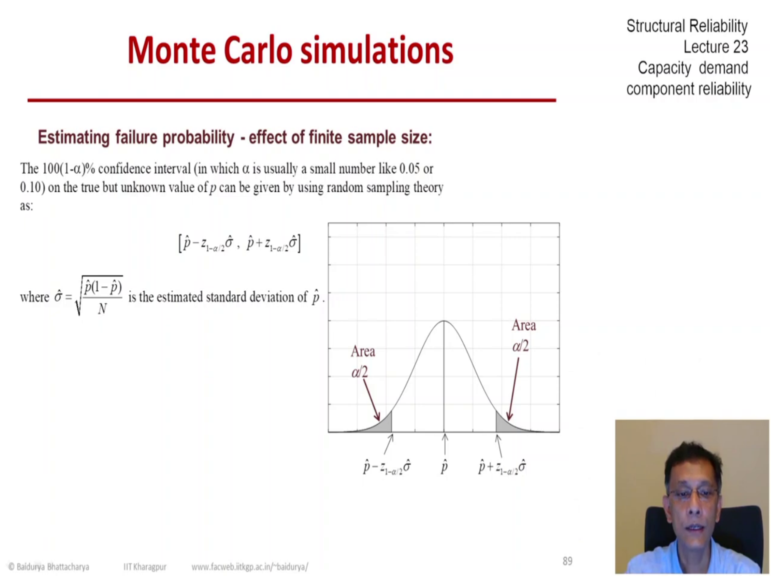
(Refer Slide Time: 04:01)



So, we also saw this earlier in this lecture in part A to recap basically if you add a good number of random variables that are not too dependent and that do not have a single very dominant component among them then that sum would approach the normal distribution regardless of the individual nature of the distributions of the individual random variables and we had the entire description of the central limit theorem we are not going to go into that now.

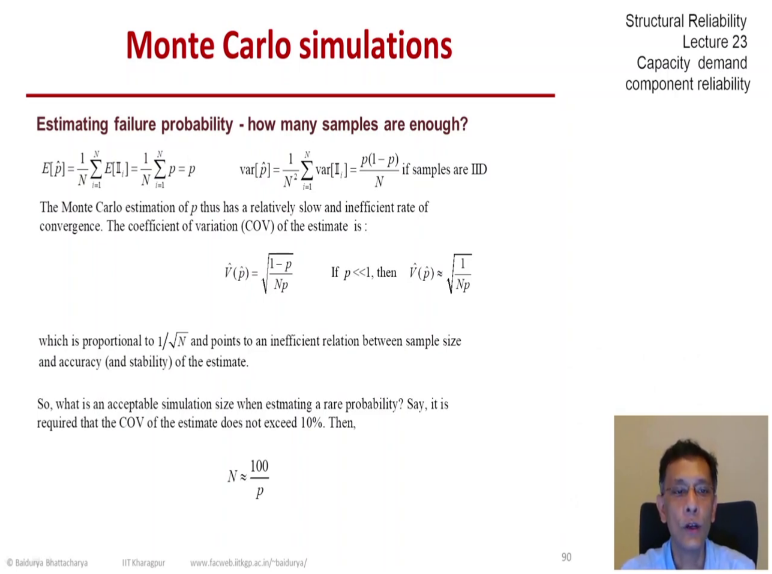
But basically this is this is what happens if you if the conditions are met the normalized normalized and centralized sum would approach the standard normal. So, that is what we are going to take advantage of because our samples in estimating failure probability a sample of the entire indicator function their ideas we have been stressing. So, with the backing of central limit theorem. Now we can say that the estimated p, p hat is going to approach a normal distribution with the mean of p and the variance of p times y minus p divided by n. So, that tells us that we are now ready to give bounds and confidence intervals on this estimated p.

(Refer Slide Time: 06:01)



And for that we can simply refer back to the standard normal distribution and say we want to have the one minus alpha percent confidence interval on p. So, if we have an estimate p hat then we can find out say the 90% and confidence interval 95% conference interval simply by looking at the standard normal deviates what you see on the screen is such an interval the area under the normal distribution curve alpha over 2 on either side those shaded regions and if we have p hat then plus and minus z of 1 - alpha by 2 times the estimated sigma which in turn is a function of n would give us the confidence interval on the estimated failure probability.

(Refer Slide Time: 07:08)



So, continuing with this thought how many samples would be enough then for a desired accuracy. So, we know that the mean of p hat is p and the variance of p hat is p q over n. So, provided the samples are IID of course that two things we see here is that because variance goes down as 1 over n the standard deviation goes down as 1 over square root of n and that is actually an inefficient convergence which is true for all such sampling techniques.

But if we want to find the estimated cov the coefficient of variation sigma over mean then it turns out to be the square root of 1 - p over NP or q over NP. Now we can invoke the general property that we are estimating failure probabilities and these failure probabilities are typically very small numbers. So, we actually can safely say that the coefficient of variation of the estimate p hat would be almost equal to square root of 1 over NP N being the sample size and P is the true unknown probability that we are trying to estimate.

So, then; this actually gives us an idea of how many samples to drive how many samples to generate. So, suppose we want the cov to be 10. So, that tells us that in that case our sample size has to be of the order of 100 over p. So, this gives us a kind of initial guess of how many samples to go for if we have an idea about what the failure probability is like or conversely if we have estimated a failure probability from a certain sample size and the cov turns out to be too large then obviously we have to keep sampling until we have reached our own comfort level in terms of the uncertainty involved.