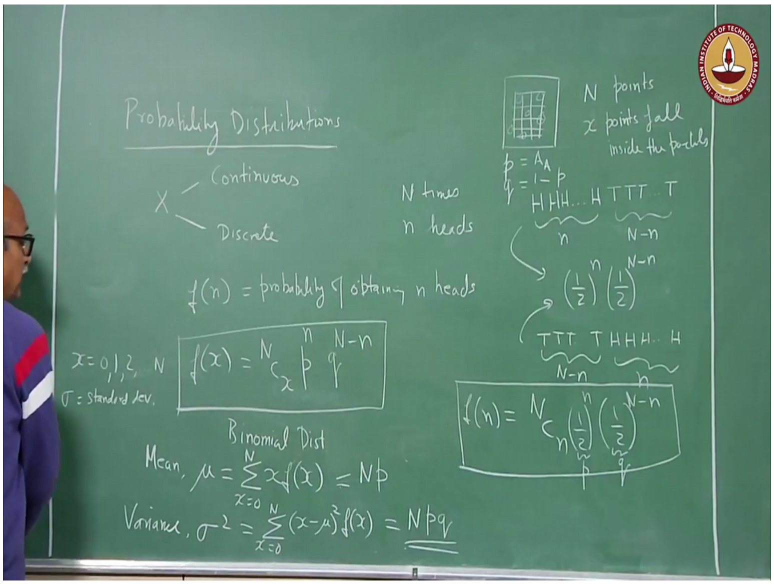
Elementary Stereology for Quantitative Microscopy

Prof. Sandeep Sangal
Department of materials Science and Engineering
Indian Institute of Technology, Kanpur
Prof. S. Sankaran
Department of Metallurgical and materials Engineering
Indian Institute of Technology, Madras

Lecture – 05
Geometric Probability

(Refer Slide Time: 00:17)



Now, before actually going into the stereology aspect I also want to introduce probability distributions. So, probability distributions are distributions of as you know random variables ok. So, there is some random variable X which will have some particular distribution. In fact so, as I said for that chord length I can get chord length distributions then the random variable in that case is the length of the chord.

If I look at the experiment that was done in the last lecture where you are counting the number of points falling inside those graphite nodules, then the random variable in that is number of points; which will vary every time you repeat the experiment it will somewhat vary it may cluster around some particular value which we call as a mean value, but it will have a distribution of points, a distribution of count.

In one case I have length of the chord which may be varying in another case I have these number of points which are varying. These are two different classes of random variables.

One of them is called continuous in the case of the length of the chord there is a uncountable number of values that the chord length can take and in the second case it is discrete because you can take countable number of values ok. In one case uncountable, another case countable both will have their own particular distributions for example, let me take up a very particular example of a discrete random variable that suppose I toss a coin N times and let us say in N tosses I get small n number of heads.

There would be some probability f n that is probability of obtaining n heads, what is the probability of obtaining n heads N capital n tosses ok. Let us try to work this out; one possibility can be I get head, head, head, head the first n small n number of tosses and then I get, what is the probability of getting the sequence as it is; that the first small n number of tosses I get heads and then the next N minus n tosses I get all tails. This should be half to power n times half to power N minus n.

I can have a sequence like this also I get first all tails and then I get small n number of heads, what is the probability of getting this sequence same that I am sure you can see now that there are going to be many more sequences I can get head tail, head tail, head tail, I can get header tail, tail, then head it can be scattered in many different ways. How many ways you can create these sequences; you can create these sequences in combination N c n ok. So, the probability of getting small n number of heads in capital N tosses would be N c n times half to power small n times half to power capital N minus n.

Now, let me just slightly modify this problem coming back to your experiment that we did. So, we are going to now bring in geometrical probability aspects here. I bring in that point counting experiment, where you had these graphite modules and you had put a grid of points, what is the probability of let us say my grid of total number of points are capital N points out of which let us say x number of points fall inside the graphite particles ok.

What is the probability of a point hitting a particle is the area fraction or let us say that is p; the probability of a point hitting a graphite nodule is p which is equal to the area fraction as we know. Probability of point not hitting the graphite nodule if I call that as q or basically 1 minus p or 1 minus area fraction then total number of points the distribution of those that every time I do an experiment I get a different number each one of you did you got different different numbers.

So, a distribution of x number of points falling inside graphite nodules they will follow a distribution similar to this only difference being that instead of half you will write p here and instead of half here you will write q p is the probability of getting a head here or probability of point hitting the graphite nodule q is probability of getting a tail which is probability of the point not hitting the graphite nodule.

So, hence this becomes this particular distribution is called as the binomial distribution if we had repeated that experiment very large number of times it can be very easily shown that the experimental data will confirm to a distribution like this ok.

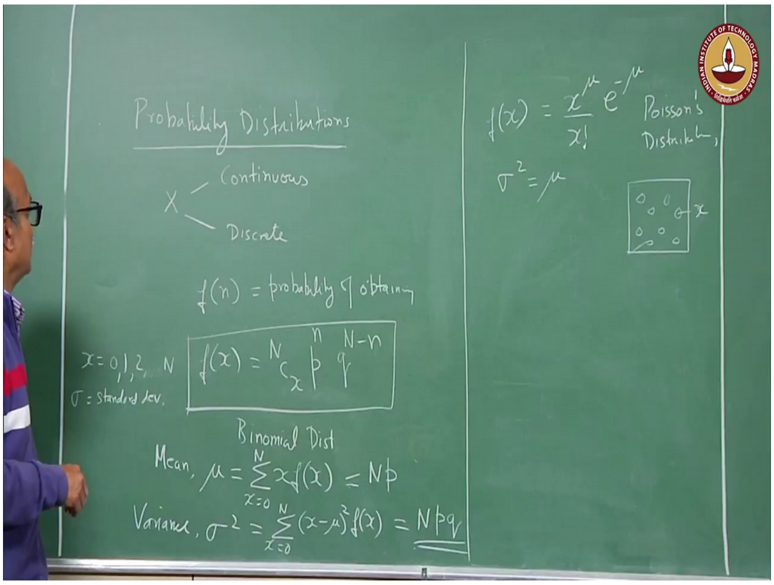
If this distribution has several properties we will not be able to cover all of those properties except for two, one is that the mean value of this distribution, how does one calculate the mean of a distribution for a and this is a distribution of a discrete random variable that we it can take specific values. In this case what can what values x can take, x can take 0 no point hits a graphite nodule only 1 point hits a graphite nodule, 2 points hit and maximum it can go is n all n points hit the graphite nodule so, will go from 0 to N.

In order to calculate the mean for this would be summation x times the probability function f x value of x multiplied by probability of getting this particular value where x goes from 0 to N. Now, if I do this summation I will simply get the result N times p ok, that sounds you know quite good that this is what I would expect also. Then there is something called as the variance or the standard deviation of a distribution right that gives you the spread whether the distribution is wide or narrow.

So, let me call that variance which is sigma square square root of that is sigma which is called the standard deviation ok. So, sigma is called the standard deviation and this sigma square can be written again as a summation, but now we have to measure this spread of the distribution about the mean value.

So, deviation from the mean square it multiplied by the probability of getting that specific value of x n x going from 0 to N this turns out to be if I do this summation of this series this turns out to be capital N times p times q. So, this is one kind of probability distribution dealing with the discrete random variable let me take just one more example of a probability distribution of a discrete random variable which is called as the Poisson distribution.

(Refer Slide Time: 13:14)

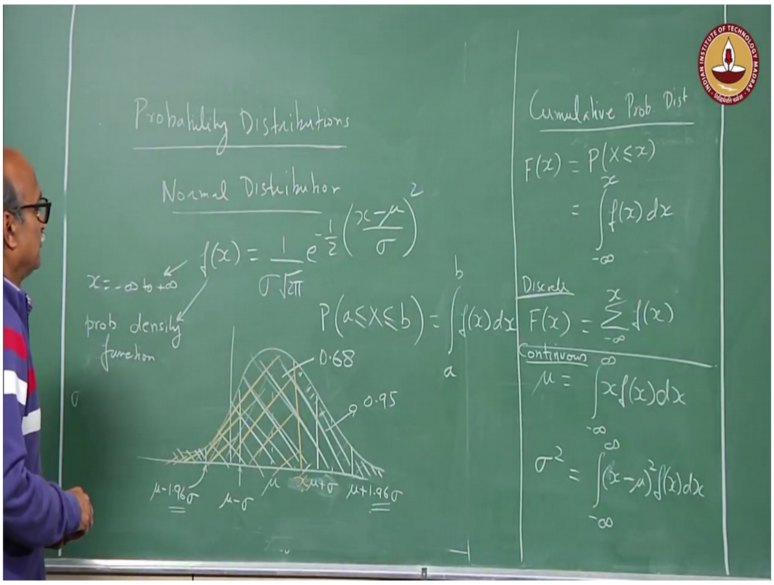


Poisson distribution is described like this it is x to power mu divided by x factorial multiplied by e to power minus mu where x is again a discrete random variable an example of this I can show that I have a microstructure of particles. Now, I examine this microstructure in different regions of my sample and I keep counting the number of number of particles in different regions and see how they are distributed.

How this number is distributed very often this number can be modeled using the Poisson's distribution and the property of these this particular distribution is mu is the mean x is the number that I have been talking about and the variance of this distribution is simply equal to the mean; I can calculate this variance using this relationship and I will get the variance as the mean itself ok. So, this is another example of a probability distribution of a discrete random variable.

Now, I have given you a couple of examples of this, let us take up now a couple of examples of a probability distribution of a continuous Random variable.

(Refer Slide Time: 15:00)



And the most well known distribution is what for the continuous random variable the heard of the bell shaped curve the Gaussian distribution of the normal distribution and this distribution function is this f x is equal to 1 upon sigma is the standard deviation square root 2 pi e to power minus half x minus mu upon sigma square where mu is the mean. So, the parameters mean and variance or standard deviations are parameters of this distribution as well.

Here x is a continuous random variable and x will vary from minus infinity to plus infinity if I plot this function this is how this distribution may look like where this is the mean what is the total area under this curve if I look at the total area under this curve should be 1, but if it has got to be a probability distribution.

So, this function is also called as probability density function, if I want to know suppose I have a process which is the generating values over this range and I know that those the that that process follows a normal distribution then if I want to know what is the probability that the value that I will get would be between a and b that would be the area under the curve ok.

So, the probability that x will lie between a and b is nothing, but a integral a to b f x d x because that represents a this integral represents the area under this curve. There are some properties for this distribution that are useful to know and that is let us say if I take this to be mu plus 1 standard deviation and if I take this as mu minus 1 standard deviation, then the area between these two is how much, this is 0.68 that is 6. If I have this data which is following a normal distribution, then a data point probability of a data point falling 1 standard deviation plus or minus 1 standard deviation about the mean that probability is 0.68 or 68 percent.

So, I it correct to say that if I repeat an experiment 100 times 68 of the results I should expect to get between plus or minus 1 standard deviation. If I extend this to about mu plus 1.96 times the standard deviation and on this side mu minus 1.96 times the standard deviation, then the area that falls between this almost to standard deviation that I am saying almost to standard deviations on both sides.

This area turns out to be 0.95 or 95 percent of the total area; that means, 95 times out of 100 on repeating the experiment I will get data point which will fall in this range and only 5 percent of time it will be outside this range in this in these 2 tails. These this is these are these are numbers which are useful to know 1.96 because these are the numbers that we would use when we will do error analysis of the data that we collected in the last lecture.

Now, one another way of representing probability distribution is we are their cumulative probability distribution, what are cumulative probability distribution is simply let me call that as capital F x the function this represents what is the probability that my random variable x will take up a value which is less than or equal to some number small x ok. What it means is that if I have a number small x here what is the probability that the random variable take up values below this, what is the probability, that is simply this area all the area from minus infinity to small x which means this is simply minus infinity small x.

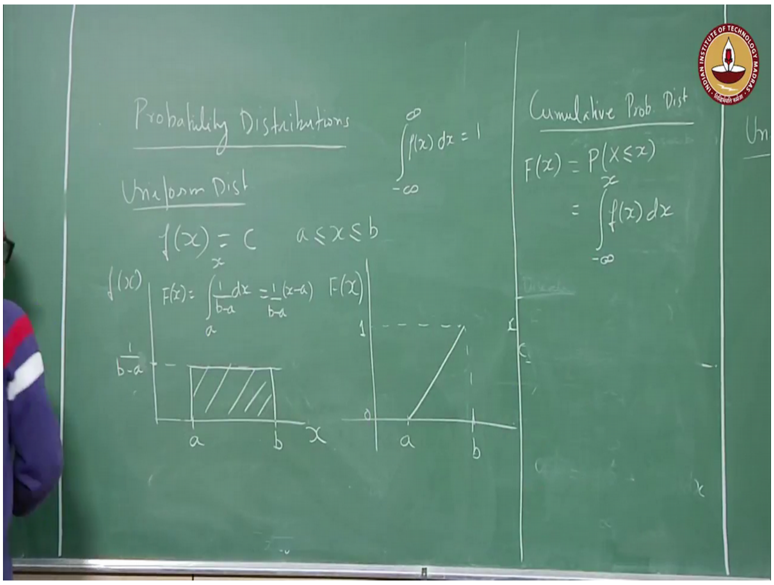
Probability density function f x times d x this is what you would get for the continuous case for the discrete case also you would have a cumulative probability distribution f x for a discrete random variable and instead of integral you replace it by summation that is all. Now, just like I had represented mean value and standard deviation and variance by functions for the discrete random variable for the continuous random variable. What is the mean value equivalent relations? But instead of summations you will have integrals.

So, this would be integral minus infinity to plus infinity x f x d x. So, instead of summation all you are seeing is that these are integrals this is for the continuous case.

The variance the square of the standard deviation would be minus infinity to plus infinity x minus mu square f x d x that is all.

Now, one important distribution I want to introduce before I end this lecture for the continuous random variable.

(Refer Slide Time: 25:00)



And that is what is called as the uniform distribution just like you had normal distribution. So, I will introduce the distribution uniform distribution and what is a uniform distribution, uniform distribution has the probability density as a constant let us say over the interval over some interval a to b. So, if I were to plot this.

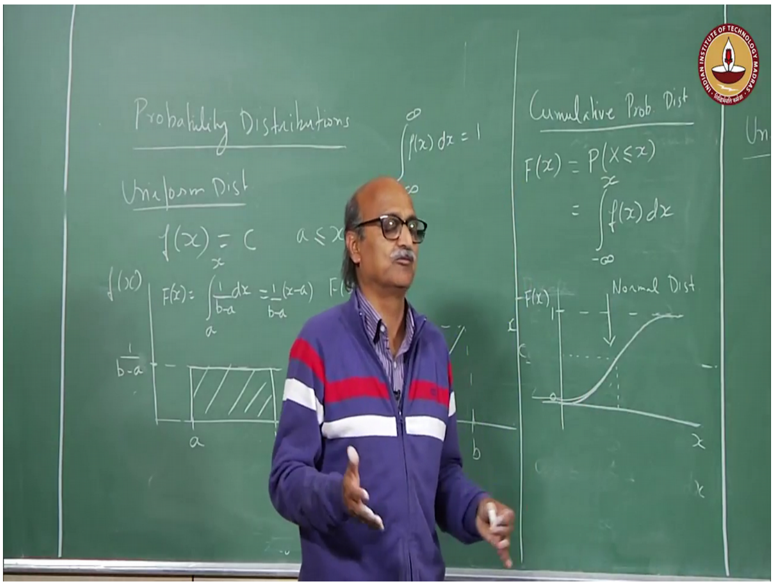
So, let us say I have a here and b here, there will be a c here. So, this is how my distribution will look like , what should be the value of c, but do you think there will be some specific value of c given a and b, one important property for all distributions minus infinity to plus infinity f x d x should be 1 which means area here should be 1.

Now, this is simply a rectangle so, c should be 1 upon b minus a right. So, this is the width the length is b minus a you get 1 , how will the cumulative distribution look like for this, well you integrate from a to x and instead of c 1 upon b minus a d x and this will turn out to be 1 upon b minus a times x minus a.

Now, let us look at this cumulative distribution when x is a the value is 0 and when x is b the value is 1. So, one of the properties of the cumulative distribution is that whichever whether it is this one or this one it will increase monotonically from 0 to 1 always. So, if this is a, this is b, this is 1, this is 0, in this case this is a straight line for the uniform distribution is this simply a straight line equation.

Similarly, if I looked at the cumulative distribution for the normal distribution if this density function f x was following a normal distribution.

(Refer Slide Time: 29:12)



Then the cumulative distribution f x versus x this will also vary from 0 to 1, but it will not be a straight line it would be and it will become asymptotic here. So, that is for a normal all of these distributions the cumulative distribution. If I take up any value of x that this probability value is telling me what is the probability of having a random variable x less than or equal to small x that is all it means and here the x is stretching from minus infinity to plus infinity. So, this would be going further down like that for the normal distribution ok.

So, this was a quick review on distributions I have introduced geometrical probability and based on this we will quickly evolve I will show you how to how to analyze this data that we collected in the last class, want to introduce what is called as confidence intervals that suppose I in the point fraction problem I have calculated an average value of point fraction. Now, after what are we trying to do by calculating those what we are trying to do is, estimate the true value of point fraction assuming that there is some true value that exists.

So, when I estimate an average value of point fraction and then there is some true average existing; I want to give some confidence to my estimate, I want to put an interval to that estimate that I expect that my if I will keep repeating these experiments my average will lie in a certain interval and how much confidence I have in that interval I will assign it a probability. For example, I can assign a 95 percent confidence interval that there is a 95 percent chance that my mean value of all the experiments I do will lie in that range and I hope that in that range my true value also exists.

So, with this will terminate the lecture here and we will continue in the next class.