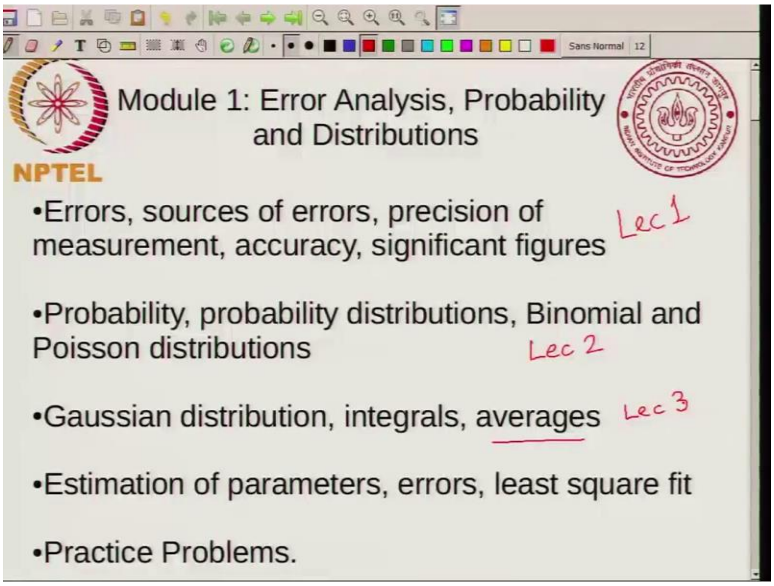
Mathematics for Chemistry
Prof. Madhav Ranganathan
Department of Chemistry
Indian Institute of Technology, Kanpur

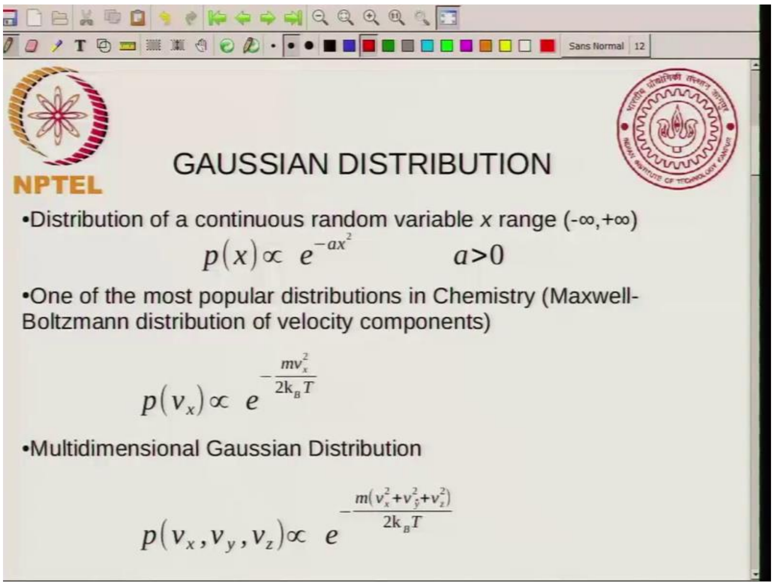
Module – 01
Lecture – 03
Gaussian distribution, integrals, averages

(Refer Slide Time: 00:16)



We have seen errors; sources of errors, precision measurement etcetera, this was in lecture 1. Then in lecture 2, I talked about probability, probability distributions, binomial, Poisson distributions and now today's lecture we are going to talk about the Gaussian distribution, integrals and averages. So, this is the topic of today's lecture.

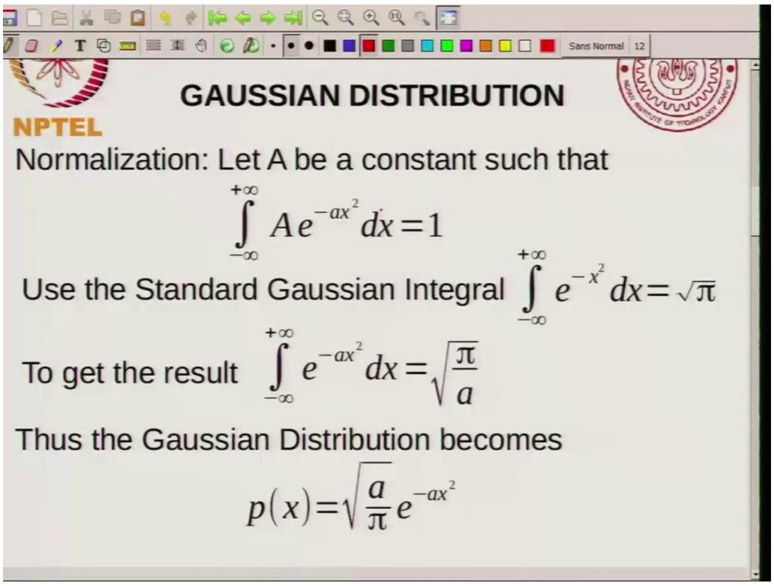
(Refer Slide Time: 00:37)



So, what is a Gaussian distribution? It is a distribution of a continuous random variable x and we will start by defining it in the range x going from minus infinity to plus infinity. So, its x is goes over the real line and the Gaussian distribution, we can understand this way; the p of x which is the probability distribution function is proportional to e to the minus a x square where a is the number that is greater than 0. This is actually one of the most popular distributions in all of physical sciences. In chemistry you have seen it in the Maxwell Boltzmann distribution for velocity components of a gas. So, for example, the probability distribution of the x component of velocity of a gas is proportional to e to the minus m v x square by 2 k B T. So, this is what is known as the Maxwell Boltzmann distribution.

You can think of a multi dimensional Gaussian distribution. So, here I have just one variable x in these 2 cases x or v x, now what you have seen in the kinetic theory of gases is that you have a probability for v x, you can have a probability of v x, v y and v z, and in the case of the kinetic theory of gases this is just proportional to e to the minus m v x square plus v y square plus v z square by 2 k B T. So, you can have one dimensional and multi dimensional Gaussian distributions.

(Refer Slide Time: 02:03)

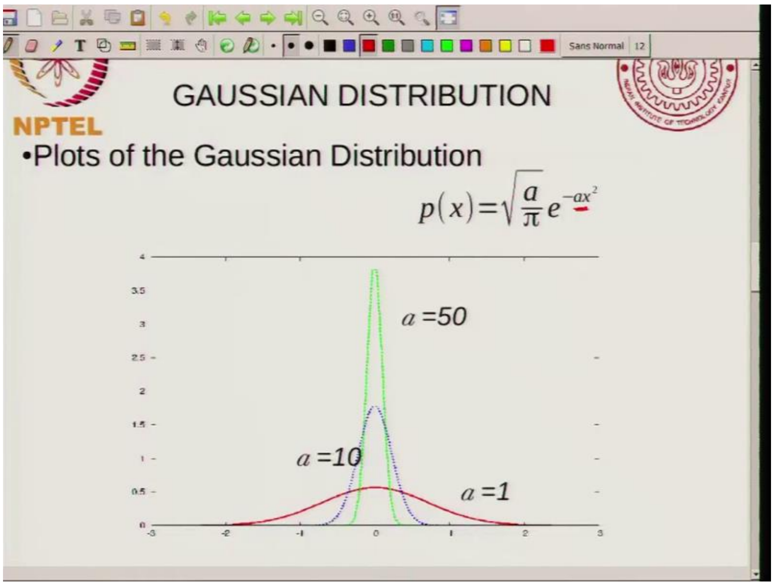


So, what can you say about the Gaussian distribution? You can normalize this, so if A is a constant such that the integral from minus infinity to infinity A e to the minus a x square dx is 1. So, why should this be 1? Because what we said is that e to the minus ax square is a probability density and, so e to the minus a x square times dx is the probability that x is between x and x plus d x, and if you integrate from minus infinity to infinity you should get 1.

Now what we said is in the previous case is that probability is proportional to this. So, the constant of proportionality is chosen A is chosen such that this integral exactly equals 1. So, what should the value of A be? So, we use a standard Gaussian integral. So, the standard Gaussian integral is A integral from minus infinity to plus infinity e to the minus x square dx this is root pi. So, this is a standard Gaussian integral and you can show this in using 2 dimensional polar coordinates, but we will not bother about showing this will just take this result. Now if you have instead of e is the minus x square you had integral e to the minus a x square then instead of just having root pi you would have root pi by a.

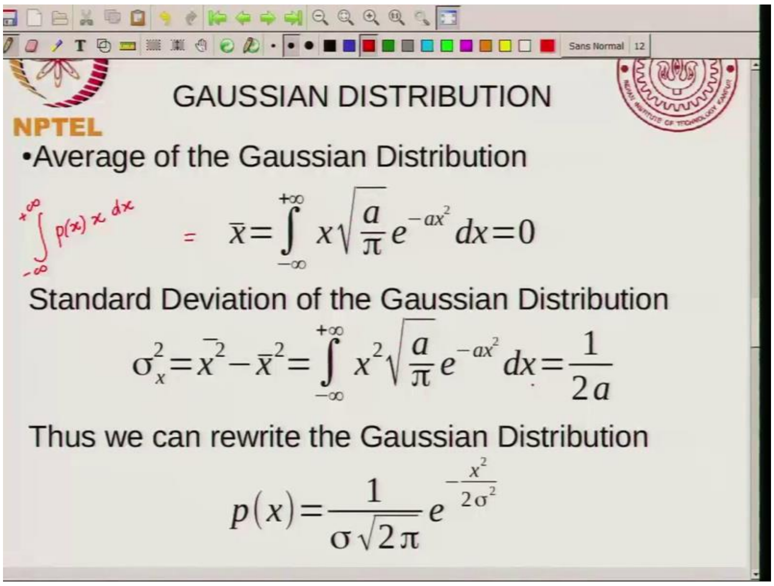
And now the Gaussian distribution since, the Gaussian distribution this capital A should be square root of little a by pi. So, that integral of A e to the minus x square dx will becomes 1. So, your p of x Gaussian distribution can be written in this form. So, this is the normalized Gaussian distribution. So, what does the Gaussian distribution looks like.

(Refer Slide Time: 03:50)



So, suppose you plot the Gaussian distribution. So, p of x is square root of a by pi e to the minus a x square. So, if a equal to 1 then you have this red curve, this red curve right here this is for a equal to 1. For a equal to 10 you get the blue curve, notice that the blue curve it is slightly higher peaked at x equal to 0, but it decays faster. In all the curves you will see that the maximum occurs at x equal to 0. So, this function is maximum at x equal to 0, when x equal to 0 the value is square root of a by pi. As you increase a, this maximum value it increases, but the rate at which it decays increases. So, it goes to 0 faster. So, when a equal to 50 this maximum value is much higher, but it goes to 0 very fast. So, the Gaussian distribution what you realize is that a, a tells you about the width of the Gaussian. So, if a is smaller than the Gaussian is wider, if a is larger the Gaussian is narrower. We will keep this in mind and we will formally see how to connect a to properties of the Gaussian distribution.

(Refer Slide Time: 05:04)

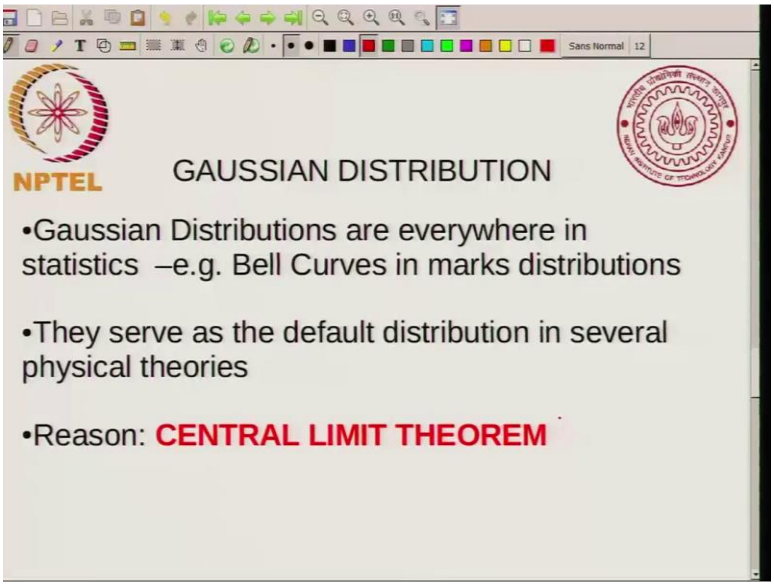


So, let us ask the question what is the average of the Gaussian distribution. So, to calculate the average, what you do is you write your average as integral p of x times x dx from minus infinity to plus infinity. So, what you do is, if you did not have this x then this integral would have been 1, but now if you have an x then you get the average value of x and this average value of x I can write in this form and you can immediately see since x is an odd function, this is an even function of x you are integrating from minus infinity to infinity this value will give you 0. You can ask; what is the standard deviation of the Gaussian distribution?

So, the standard deviation is average value of x square minus average value of x squared, so the average value of x is 0. So, the standard deviation square is just x square average. So, average of x square means you have p of x and instead of having x you have x square. So, x square times p of x dx and this you can show is equal to 1 by 2 a, I will let you work this out you can show this value of this integral as 1 by 2 a. So, this is my sigma for the Gaussian distribution this is the average and so what I can write; I can write my Gaussian distribution as instead of a I can write a as 1 by 2 sigma square and, so I can write my Gaussian distribution in this form. So, where I had square root of a by pi I will have 1 over sigma root 2 and where I had a x square I will have minus x square divided by 2 sigma square.

So, this is the usual form of the Gaussian distribution and what we notice is that if you know sigma you can write the Gaussian distribution. So, sigma measures as we said in the earlier that sigma measures the spread and what you see from the graphs is that as you are a value becomes larger your spread becomes less and you can see from here that sigma and a are inversely related to each other. So, your spread is proportional to 1 by 2 a or the square of the spread is proportional to 1 by 2 a and so what you can see that increasing a reduces the spread of the Gaussian distribution.

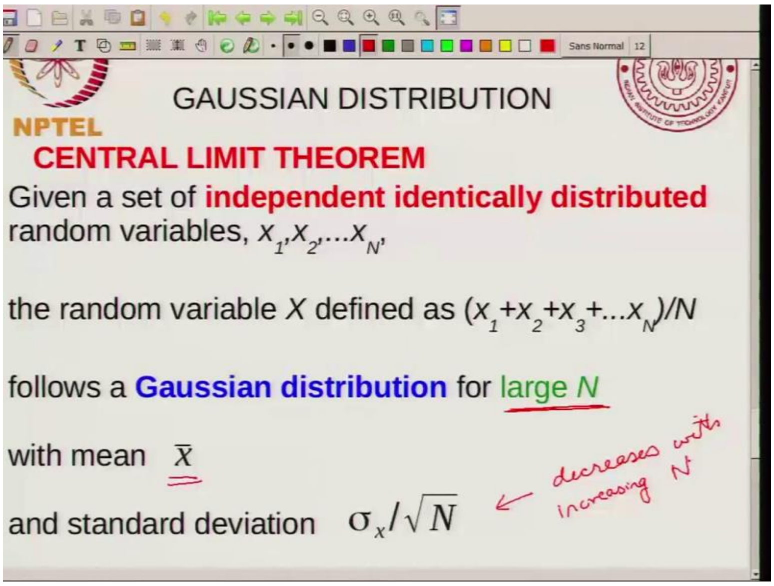
(Refer Slide Time: 07:37)



Now, Gaussian distributions are everywhere in statistics, I mean where you have heard about bell curves you know that if I take all the suppose there are very large number of students taking an examination and you take all their marks and you make a histogram of the marks you get something that looks like a bell curve and it looks like a Gaussian, Gaussian distribution and these serve in fact in lot of physical theories.

The Gaussian distribution serves as the default distribution you know if you do not know what a distribution is you by default you will assume the Gaussian distribution, there is a reason for this and that is related to something called a central limit theorem in statistics and I will just go that briefly I will just give you the idea.

(Refer Slide Time: 08:19)



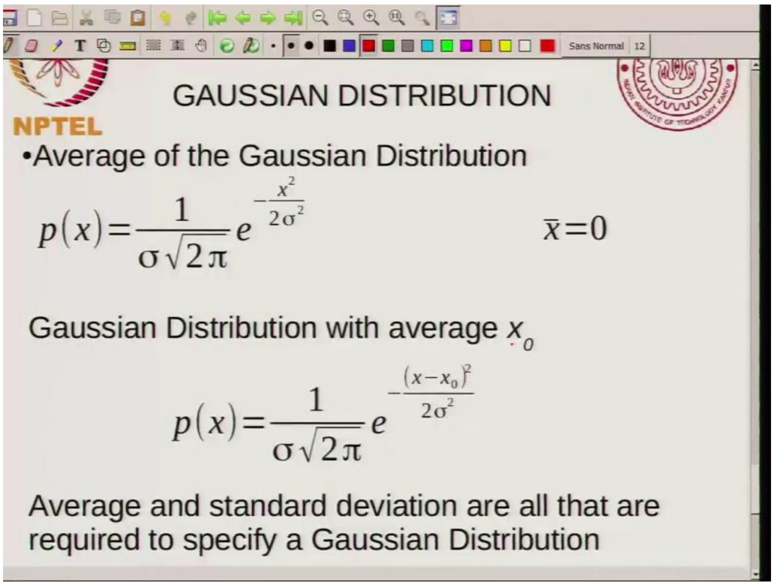
So, central limit theorem is sort of the reason why Gaussian distributions are everywhere. So, what it says is that suppose you are given a set of independent identically distributed random variables. So, x 1, x 2 etcetera are random variables each of them is independent of each other, each of them has the same probability distribution. So, that is why they are identically distributed.

So, the probability distribution for x 1, x 2 etcetera as a same. Each of these is a different random variable then you define a new random variable x as the average of all these variables. So, you have N random variables you define a new random variable as the average of all these. Then you can show then according to a central limit theorem then as N becomes very large for large N then x follows a random; a follows a Gaussian distribution with mean x bar. So, since each of these each of these variables they have the same probability distribution. So, all of them will have the average of x bar and this is exactly the average of capital X and the standard deviation of X is sigma x divided by square root of N, is a standard deviation of small x of any of these is the same sigma x and you divided by square root of N that is a standard deviation of capital X.

So, this is the central limit theorem that says that you know if you have many sort of independent identically distributed events and you take the average of them you will get a variable that follows a Gaussian distribution provided you take large enough number of them, then the mean will just be the mean same as a mean of any of those variables and the standard deviation actually goes down. So, if you take larger N your standard deviation will be dropped. So, the standard deviation decreases with increasing N.

Now immediately you realize something that you can think of this as doing an experiment many times and then taking the average. So, you can immediately see the connection if you do the experiment more times then you know that your standard error of mean goes down and this is a direct connection with that idea right here.

(Refer Slide Time: 10:54)

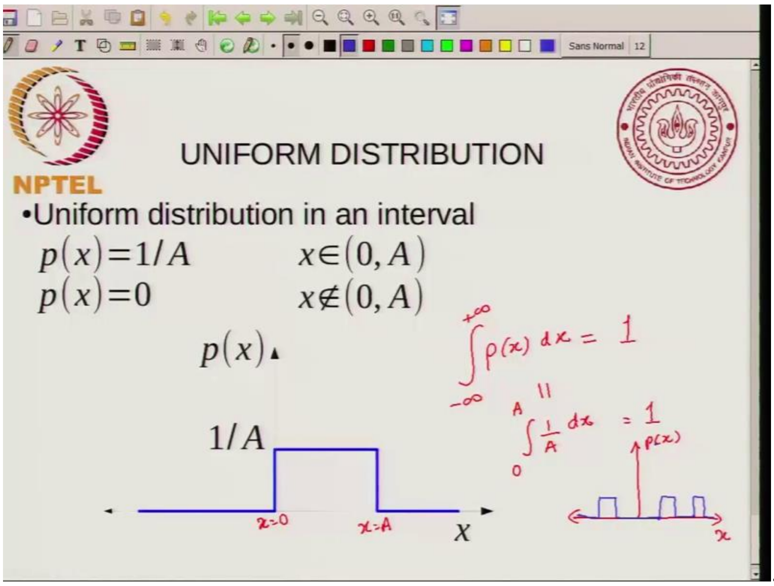


So, what is the average of a Gaussian distribution? So, the Gaussian distribution is defined in this way and average is clearly 0, now you can also have a Gaussian distribution with average x 0. So, the average of a Gaussian distribution in this form is 0 that is very clear we already saw that, now if you want to have a Gaussian distribution with average x 0 then what you do is instead of having x square here you have x minus x 0 square and keep everything the same.

So, this gives you a Gaussian distribution with average x 0 instead of 0 and what this illustrates also is that if you know the average and you know the standard deviation if you know x 0 and sigma then you can write a Gaussian distribution. So, the Gaussian distribution you need only 2 things if you know x 0 and you know sigma you can write p x. So, to specify a Gaussian distribution you just need the average and the standard deviation.

So, the point is that whenever you describe something in terms of averages and standard deviations you are inherently assuming some sort of Gaussian distribution is there, you are assuming that your data is distributed like a Gaussian distribution. Now there are other distributions a Gaussian distribution is not the only continuous distribution. So, this is a continuous distribution going from x going from minus infinity to infinity you can have many other continuous distributions because what we said is that your distribution is just represented by a function. So, if you go back to what we how we define. So, you have a p of x a function is a function of x and that defines your distribution and you can have many different kinds of functions. So, you can have many different kinds of distributions and the Gaussian distribution is one very popular distribution, but you can have other distributions also. So, one other distribution which is a simple distribution is what is called a uniform distribution.

(Refer Slide Time: 12:49)



Now a uniform distribution, you can have a distribution that is uniform in an interval. So, p of x equal to 1 by A when x is contained in 0 to A or it is 0 otherwise. So, you have p of x, p of x need not I mean it needs to be continuous but it need not be smooth, it can be it can show derivative discontinuity. So, this is an example of a distribution. Now if you plot this. So it is 0 when x is less than 0 and it is in this region this is x equal to A, x equal to 0. So, in this region the value is 1 by A everywhere else it is 0, that is what it looks like.

Now why did I choose this value of 1 by A here? Because I want integral p of x dx from minus infinity to plus infinity this equal this should be equal to 1. So, I want this integral to equal to be equal to 1. Now I chose this as 1 by A for this interval, so that this integral which is nothing, but integral 0 to A, p of x p of x is 1 by A dx and this is equal to 1. So, this is a uniform distribution I took it in the interval from 0 to A, I can take it for in the interval 0 to 1 I can take it from minus 1 to 1, I can define many such uniform distributions over different ranges and I choose the value to be suitable so that that particular distribution is normalized.

Now, I can take a distribution I showed this case, I can also take distributions that look something like this. So, if this is my x, my p of x might look something like take a different color. So, it might look uniform here 0 here, uniform here, I could have distributions that look like this. So, what I mean is that the range 0 to A is not a continuous range, but rather it is split into 2 or 3 different pieces that is also valid p of x can be essentially it can be any function you can have any function and that should that can satisfy this p of x.

So, what we have seen in this lecture is the Gaussian distribution that is one very popular distribution, but we should keep in mind that probability distributions can be any functions and so you can have many different kinds of probability distributions. So, and you can calculate averages, you can calculate averages of squares etcetera for Gaussian distribution these things are easy to calculate and you have seen some of these exercises when you do let us say kinetic theory of gases, when you do kinetic theory of gases you calculate the root mean squares velocity or you calculate the average velocity. So, these are all related to calculating averages and of the Gaussian distribution.

So, in the next class I will talk about estimation of parameters and least square feet and then finally, I will end this module by solving some numerical problems. So, I will stop for today here.

Thank you.