Advanced Mathematical Methods for Chemistry

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

Module - 11
Probability, Probability Distributions, Error Analysis
Lecture - 01
Discrete and Continuous Random Variables

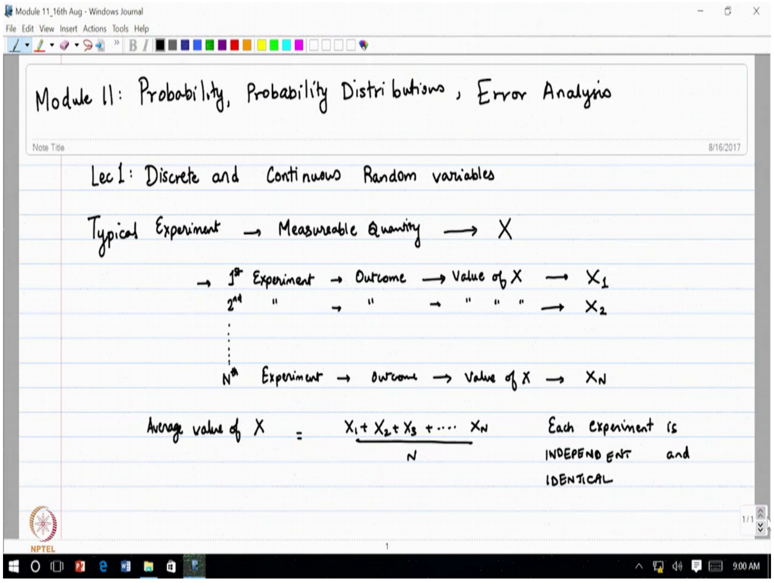
Now, we will start module 11 and in this module I will be talking about probability, probability distributions, distribution functions and error analysis. And now a lot of these you might actually be fairly familiar with, but it turns out that this theory of probability and probability distributions is actually an area of mathematics in itself and it has there are lot of very interesting results and theorems that are part of this branch of mathematics and so it is good to at least put lot of the probability that we use very intuitively in a more formal framework. So, for example we use probability when we are talking about in quantum mechanics; we are talking about the wave function and the square of the wave function we use it to denote the probability that your particle is located at some point in space.

Similarly, in statistical mechanics we talk about the Boltzmann probability distribution and we say that there is a probability that your system is in a certain state or your molecule is in a certain state. So, and we actually use probability a lot in chemistry and in all daily life applications. So, what this module; what I will try to do in this module is to try to put all those ideas in a more formal setting and again I will not be doing too many details, but I will at least give you the skeletal ideas and with this you can actually go ahead and read books and so that you get much deeper understanding of probability.

What I will try to do is to first talk about discrete random variables; where this is something that you are familiar with, but it is in a slightly different language than what you are used to. And then we will talk about continuous random variables, we will talk about probability density, probability distribution functions and so on.

So let us get started, so first I will give a very intuitive example of what you might do in any in any laboratory experiment. So, you might go to a lab let us say say you are doing a titration experiment and you are measuring the concentration or let us say you are measuring the volume of the titrants that you need to add and s you measure the volume and what you typically do is; you measure it many times. And then you measure it maybe 3, 4 times how many ever times you can and you take an average and you report that as the volume of the titrant.

(Refer Slide Time: 03:43)



Now we can; so if I say this more formally; so, what you do is you have and so I will just say if you say a typical experimental protocol; typical experiment it might be any kind of experiment; it need not be a titration, it can be any kind of experiment; you have some measurable quantity. It might be the volume of the titrant or it might be something else you might be measuring some other quantity.

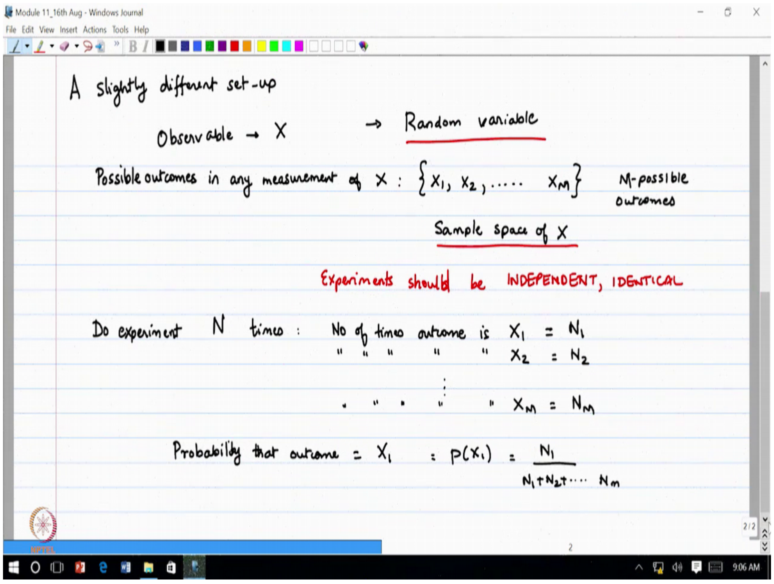
Let me call this quantity as denote by X; X is a measurable quantity; now what you do is; you do your experiment you do let us say first experiment and the outcome of the experiment is some value of X; let us say that value is X 1. Similarly, if you do the second experiment; the outcome of the value of X in the second case, let us say that comes out to X 2; you do this many times; let us say you do it Nth experiment.

The outcome of the value of X is X N, so if you have this is something very very typical situation in a laboratory. Then what you will say is that; the average of X value of X equal to X 1 plus X 2, X N divided by N. So, this is what you would say in any typical laboratory experiment. Now later on, we will see what all the conditions under which you can do this.

But let us assume that you do this there is an assumption; I will just state it here that each experiment is independent and identical. So, each experiment is completely identical and it is independent of each other. So, what; that means, is you do each experiment exactly the same way and each experiment should be independent of each other. So, the second experiment value should not depend on what you got the first time. So, under these conditions you can use the average you, can define it in this way and this is what you report.

Now let us consider a slightly different experiment.

(Refer Slide Time: 07:00)



So, a slightly different setup again we consider the observable X and now we say that the possible outcomes of experiment any measurement of X. So, if you make a measurement of X; then you have only N possible X outcomes and they are X 1, X 2 up to X N; you get this N possible outcomes. So, in other this set of N possible outcomes; so, this is sometimes called the sample space of X; this is the N possible outcomes and I will just jump ahead a bit and I will say that X is called a random variable.

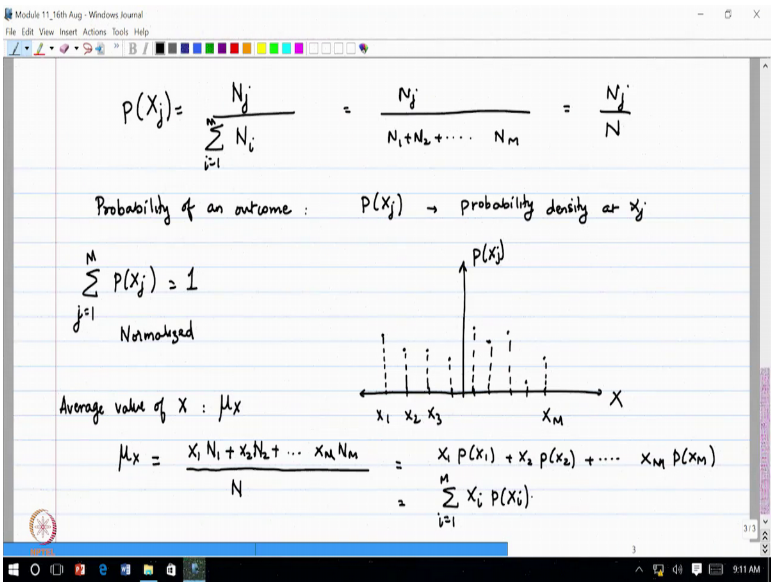
I am calling it a random variable; this is a technical term, it just means that there is some way in which at the experiment you get some result. So, that result we are calling it in general as a random variable. So, this term random variable and sample space of X; so, these are slightly more technical terms, but essentially you can just think of it as you have this observable, you make a measurement and you get this.

One of the conditions when for these to be satisfied is that each of these experiments should be independent and identical. Now there are N possible values of X; now let us say you do the experiment M times, now each time you might get either X 1 or X 2 or anything up to X N. Now let us take the case I mean you can, so if you do it M times; the number of times outcome is X 1 equal to N 1; that means, out of these M times N 1 times; you get X 1.

Similarly, number of times outcome is X 2 is N 2 and so on all the way you get number of times X N equal to N N; this is a slightly bad notation, but we will just stick with this. I will just make a slight change of notation, just to be more typical; we will say that there are M possible outcomes. So, in the sample space of X; we will make it up to X N; so, M possible outcomes and we will do the experiment N times and so you only have X 1 to X M here N M. So, I just change this notation so that you do not have things like N N in this notation. So, the idea is you make a measurement and there are M possible outcomes; that is a sample space of X and you do the experiment N times and you get all these.

So, now you can ask what is the probability that outcome equal to X 1; this is P of X 1, this is denoted by P of X 1 and this is just N 1; divided by N 1 plus N 2; N M; similarly you can do for probability for any.

(Refer Slide Time: 11:50)



So, P of X j is equal to N j divided by sum over i equal to 1 to M N; i. So, I can write this in general this is just N j divided by N 1 plus N 2 N M. So, now this is where we can naturally see probability of an outcome, so we have probability of an outcome. So, this P of X j is also referred to as probability density at X j and what you would imagine is that suppose I have this variable.

Suppose I plot X; the values of X and let us say these are X 1, X 2, X 3 all the way to X M. So, if we have these are the possible values of X on the line and let us say I plot the probability on this axis P of X j. So, P of X 1 I will plot here and let say P of X 1 is some value, P of X 2 is some value X 3 might be something else X 4 might be something else X 5 might be something else.

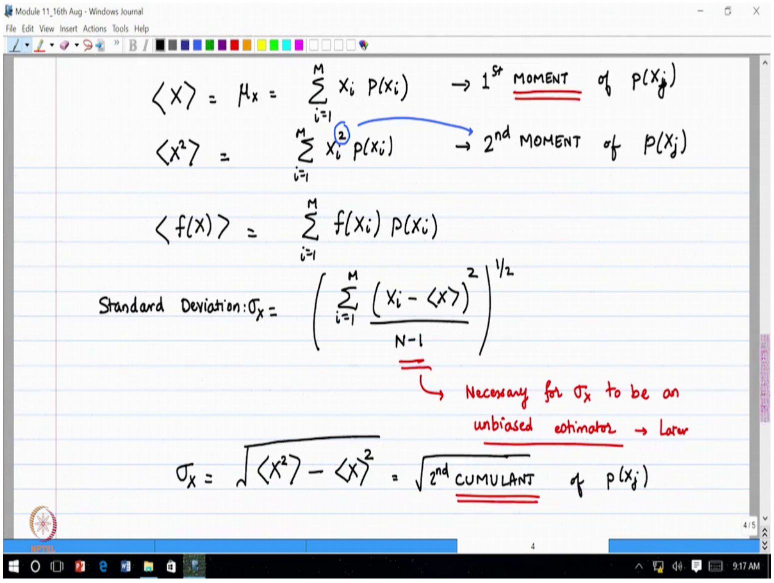
So, these are the probabilities; so on this axis I am showing the probabilities and you can show them as such graphs. This is sometimes call the histogram of X, but essentially what we are plotting is a probability a versus X on the axis. You might just plot them; so. remember the number of if you say number of times you get X j is just probability multiplied by the total number.

So, this is Nj divided by N is a total number of experiments so now way; obviously, if you take the sum over i of j equal to 1 to N; N j or P j P or P of Xj this is should be equal to 1 which is very obvious because in the numerator you will just have some over N j and j equal to 1 to M 1 to M.

So this probability is said to be normalized that means, it adds up to 1 which is what it should be. So, probability did should be normalized sometimes we use unnormalized probability distributions and then we normalize some later. So, now you can ask what is the average value of X; I will just call it mu I will denote it by mu X; so mu X is equal to; so I can write it in the following way; I can write it as X 1 times N 1 plus X 2 times N 2; X M times N M.

So, number of times you get N 1; so that is X 1 number of times you get N 2 you get multiply this X 2. So, this divided by N and I can write this in the following way; I can write it as X 1 P of X 1 plus X 2; P of X 2; X N, X M; P of X M or I can write it as sum over I equal to 1 to M; X i; P of X i.

(Refer Slide Time: 16:50)



So, this is how you define the average and we can define the average of; so we will use the notation, we will use the angular brackets to denote the average of X and you can immediately see from this that the average of X square is equal to sum over i equal to 1 to M; X i square P of X i and in fact, average of any function of X is equal to sum over i equal to 1 to M; f of X i; the value of f at X i; times P of X i. So, you can do this for squares, you can do this for cubes and so on; this is sometimes called the; so we will just go back here probability density or it is called the probability.

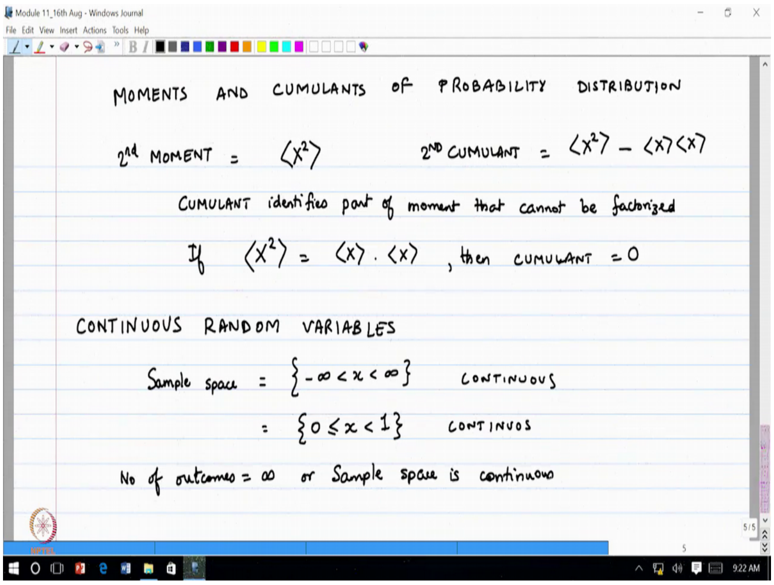
So, P of X i is called the probability distribution of X; so, the set of all P of X i is called the probability distribution of X. And then the average is called the first moment; this is the term this is moment of the probability distribution function that is P of X; X i; X j this is the square is called the second moment, so this 2 here is what connects with the second moment of P of X j. Now a quantity that we use very often is the standard deviation.

This is equal to; so, if you have a finite number of experiments then the standard deviation is basically you can write it as sum over i equal to 1 to M. And you can just write it as X i minus average value of X; the whole square divided by N minus 1. I will go into y; y there should be an N minus 1 and the whole thing to the power half. So, this N minus 1 is actually necessary to have for standard deviation; we will call it sigma of X; for sigma of X to be an unbiased estimator.

We will go into biased and unbiased estimator; so, this will be discussed later; but essentially, you can just take this formula and what you can show after some work is that; you can write sigma X as average value of X square minus average value of X whole square and under root, so this is sigma of X.

So, these are the; so I can also write this as the second moment of X. So, this is the second moment of X, this is the first moment of X squared and in some terminology, this is referred to as th; so it is square root of what it is referred to as second cumulant. So, this is another term; so just as you have used a term moments. So, this is the second cumulant of P of X j, so the standard deviation is or the square of the standard deviation is the second cumulant of the probability distribution function.

(Refer Slide Time: 21:57)



So, I will just emphasize that you have moments and cumulants of probability distribution. So, you can think that like if you take the second moment is equal to average of X square and second cumulant is equal to average of X square minus average of X times average of X. So, what it says is that the second cumulant looks at the part of X square that cannot be written as average of X squared.

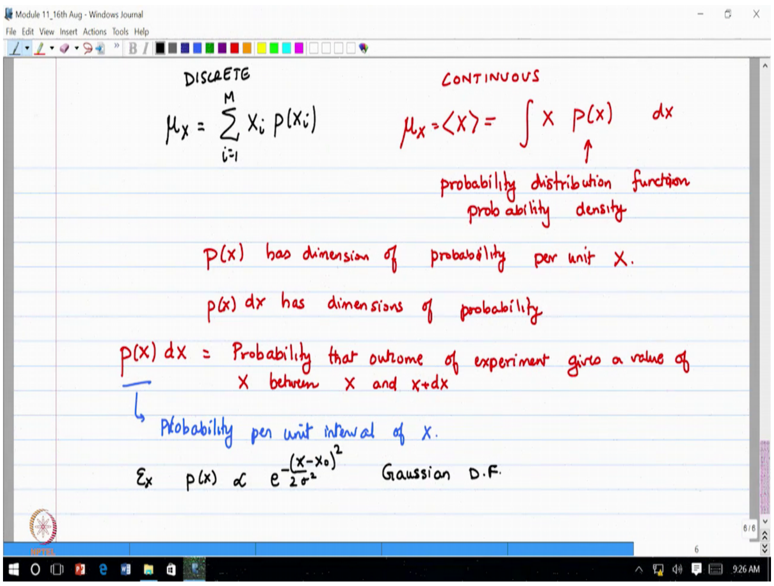
So, it is a part of the second moment that cannot be written as a product of first moments. So, similarly so cumulant identifies part of moment that cannot be factorized, so if you think that if; so if X square equal to average of X into average of X, then cumulant equal to 0, so there is no part that cannot be factorized. Now I will not be doing a general discussion of moments and cumulants, so we will not do a general discussion of moments and cumulants here.

But what I want to do next is to generalize this idea to continuous probability distributions. So, here the sample space for this experiment consisted of a finite number of outcomes. So, in this experiment you had only a finite number of outcomes; now if your space of outcomes becomes continuous. So the basic idea is that your sample space, this becomes basically minus let us say infinity less than X, less than infinity; you need not take that; you can just take; you know this is continuous

Or you could also; I mean, I need not have a range from minus infinity to infinity, I could just have 0 less than equal to X less than 1; this is also continuous. So, the sample space should become continuous, so you do not have only a finite number of values, you have infinite number of values between 0 and 1. So, the number of outcomes becomes infinite; so number of outcomes is infinite or the sample space is continuous.

So, number of outcomes equal to infinity or sample space is continuous; in this case what is done is you define; so let us take the average value of X; mu of X which was defined as sum over i equal to 1 to M; X j; X i, P i; P of X i.

(Refer Slide Time: 25:47)



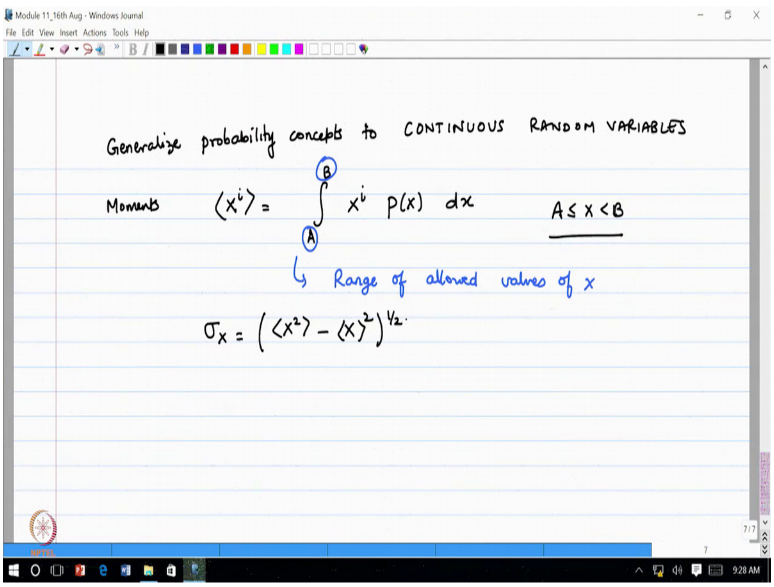
So this is for discrete; now for continuous what is done is you write mu of X equal to average value of X is equal to; now I have some over i equal to M to N; so, the sum goes over all the discrete outcomes all X 1, X 2 up to X M. So, now since my number of possible outcomes of X is infinity, this should be replaced by integral dx and what I have is an X i is replaced by X; which is a continuous variable and I have P of X.

Now this is the disc; this is a continuous version of probability and this is called the probability distribution function or probability density. Now, it is very important that you identify immediately that; if I define it in this way P of X has dimensions of probability per unit of X because what I do is X, PX, DX; this has units of X. So, P X times d X has units of probability, so in other words P X, times d X has dimensions of probability and this is a very important point. So, P of X; d X equal to probability that value that outcome of experiment gives a value of X between X and X plus d x.

So, the value of X is in this interval between X and X plus d X. So, naturally the interval comes and therefore, this is the probability per unit interval of X. So, P of X is a probability per unit interval, so if X is the probability per unit interval of X and this is what we said is a probability density or the probability distribution function. Now this is if you have a continuous variable; this is usually a function of X. So, for example, P of X proportional to e to the minus X square divided by 2 sigma square; this is a Gaussian distribution function; sigma is some constant.

So let me make it X minus X 0 square divided by 2 sigma square; this is a Gaussian distribution function.

(Refer Slide Time: 29:56)



So, what we have seen is we can generalize probability concepts; concepts to; now X is defined as a continuous random variable and what this gives us is you can define moments. So, the Nth moment X N average equal to integral X N or let me not use N again let me use ith moment; X i; P of X; d X and the integral is over whatever the range of values of X is. So, this is the A to B; so what this means is that A less than equal to X less than B this. So, X is in this interval A to B; X can take any values in this interval; so, this is the range of values of X.

And, now you can also define the standard deviation of X as just average of X square minus average of X; the whole square raise to half and using that definition you can define the average values of X. So, what we see in the next class is how to use these probability distribution functions because these are some things that you use very often in chemistry.

So, we will see how to use these different probability distribution functions in various; we will see the common probability distribution functions and then we will see how to do various calculations with them.

Thank you.