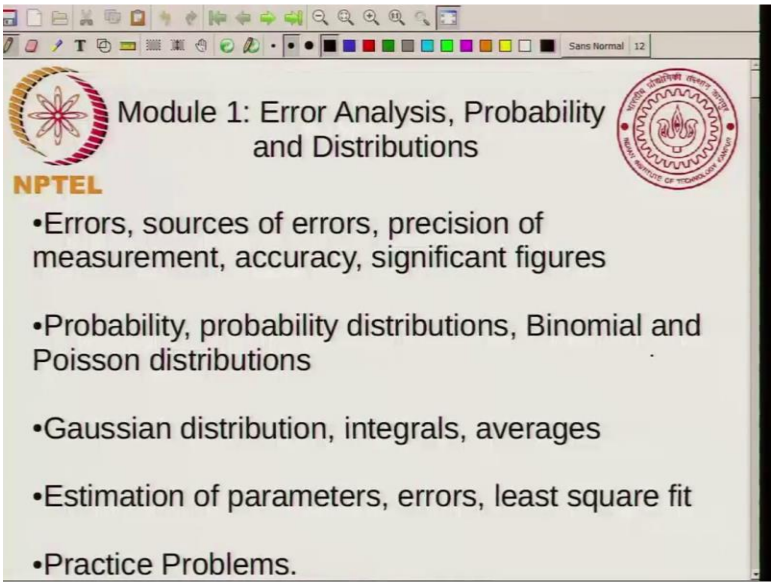
Mathematics for Chemistry
Prof. Madhav Ranganathan
Department of Chemistry
Indian Institute of Technology, Kanpur

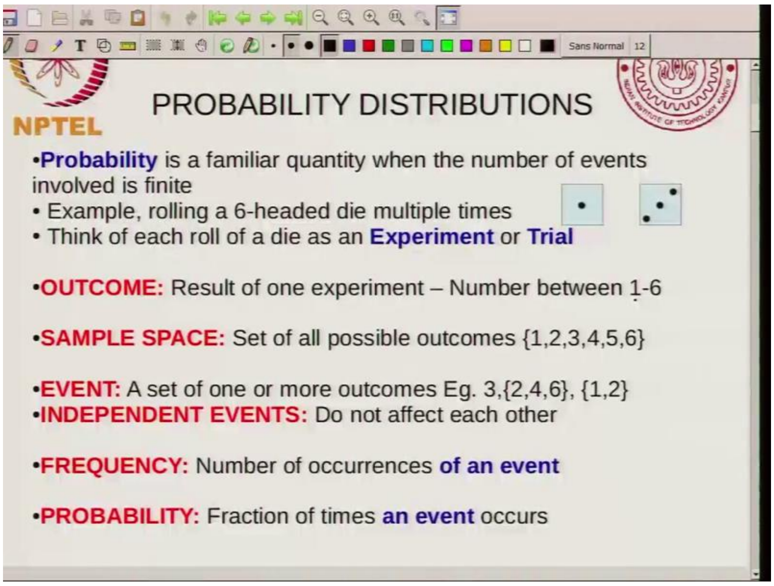
Module - 01
Lecture - 02
Probability, Probability distributions, Binomial and Poisson distributions

(Refer Slide Time: 00:21)



Let us continue that with the second lecture of this course. So, in the first lecture I talked about errors, the sources of errors, precision, measurement, accuracy and significant figures. Today I will talk about probability, probability distributions, the binomial and Poisson distributions.

(Refer Slide Time: 00:32)



So, I am going to talk about probability distributions. Now probability is a very familiar quantity to most of you and we are used to thinking about probability especially when the number of events is finite. For example, you can imagine rolling a dice, so for example, a typical dice that has 6 heads. So, 6 sided dice you roll it, imagine you roll it multiple times then you can ask, how many times you get a certain number like how many times you might get a 1 or how many times might get a 3 and you might ask the question what is the probability of getting a 1 or a 3? And just to make the language clear you can think of each roll of the dice as an as a single experiment and the number that you get when you roll that is the outcome of that experiment.

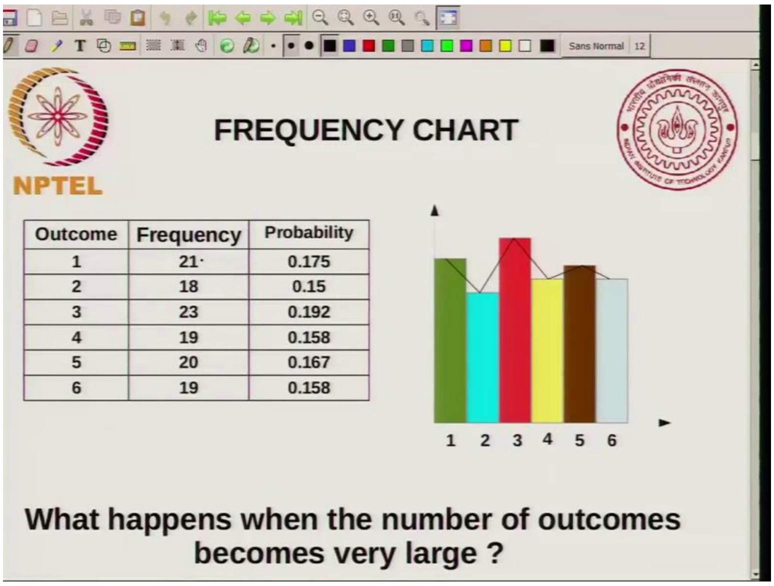
So, we will just define a few terms, just so that later on when we discuss other topics you will be comfortable. So, the outcome is a result of 1 experiment and in this case, if you are rolling a die the outcome is a number between 1 and 6, so it can be 1, 2, 3, 4, 5 or 6. The sample space is the set of all possible outcomes, so here the sample space consists of exactly 6 numbers 1, 2, 3, 4, 5 and 6 and event, event is a set of 1 or more outcomes. So, you can say an event can be suppose you get a 3 on the dice that is an event, you can have an event like you get a 2, 4 or a 6 you can define an event like that. So, an event can be either a single outcome like 3 or it can be multiple outcomes like 2, 4, 6. So, you can have an event of getting 2, 4 or 6. So, any of those contributes to the event or you can have like only 1 or 2 you know you can have event with just 2 possible outcomes. So, an event is a set of 1 or more outcomes.

So, there is a slight difference between event and an outcome, outcome is result of a single experiment. So, it can be just a number between 1 to 6. For example, you can define an event what is the probability you can define an event as an even number outcome as an outcome of an even number. So, that defines an event and that has 3 possible outcomes 2, 4 or 6.

Now, and I should emphasize that in many cases an event is just the same as an outcome. So, many cases we treat an outcome as an event, but you can treat events as set of more than 1 outcome. Then what are independent events? Independent events are events that do not affect each other. So, if you roll the dice 2 times, then the result of those 2 dice are independent of each other. So, you can define independent events. What is the frequency is a number of occurrences of an event. So, the frequency is always defined with respect to an event. So, it is the number of occurrences of an event. Now if your event was just a single outcome then how many times you can ask what is the frequency of 2 that is how many times 2 appears in a roll.

The probability in this case is defined as the fraction of times an event occurs. So, you always ask the probability of an event and that is basically the fraction of times the event occurs and this is the general concept of probability that you are familiar with and what we usually say is that you do the experiment many times then your statistics are very good, if you do the experiment very few times then you know you might not get good statistics. So, if you do a experiment a very large number of times then you would say that the probability of getting 1 and this probability of getting 3 on the dice are the same and each of them is 1, 6. So, this is what you are familiar with, with probability distributions at least of when the number of events is finite.

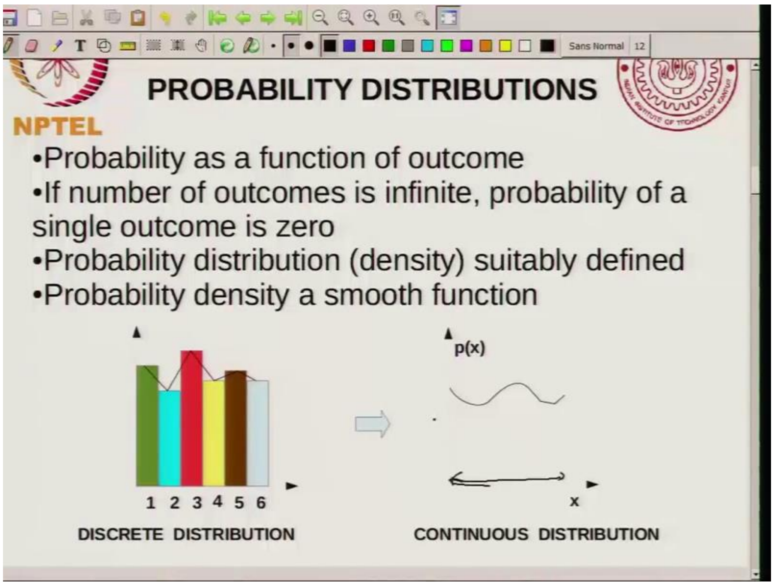
(Refer Slide Time: 04:51)



Now, you can make a frequency chart and this is something you are all familiar with. So, you have 6 possible outcomes and you look at the frequency of each outcome, outcome is the same as event in this case. So, let us say you roll the dice 100 times and 21 times you get 1, 18 times you get 2, 23 times you get 3, 19 times you get 4, 20 times you get 5 and 19 times you get 6 and from this frequency, you can construct the probability of each event. So, there are 120 roles that I have done here, not 100; there are 120 experiments. So, if you divide each number by 120 you will get the probability of each event and what you can do? You can take this frequency and you can plot it on a chart like this, on a diagram like this where you have a 1, 2, 3, 4, 5, 6, 1 has appeared 21 times, 2 has appear 18 times, 3 has appear 23 times and so on.

So, you can make a chart like this of frequency and this is something all of you have seen before. Now the question is what happens when the number of outcomes is very large. So, suppose here I had only 6 possible outcomes, suppose I had an experiment where the number of outcomes is extremely large, it is not just 6 possible there are many possible outcomes then what happens to this chart and what happens to this graph, and to answer this we will introduce the concept of probability distributions.

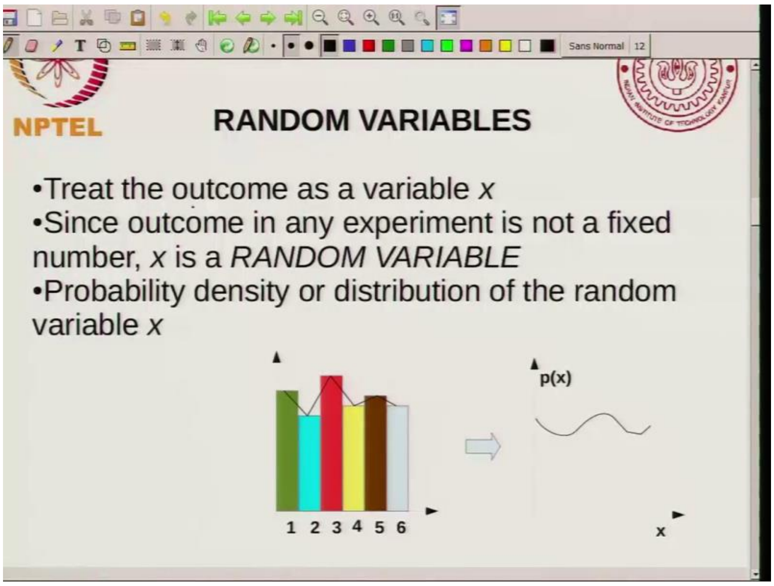
(Refer Slide Time: 06:24)



So, unlike probabilities of events these are probability distributions. So, here we think of the probability as a function of outcome, if the number of outcomes is infinite then the probability of a single outcome is 0. So, here you have only 6 events. So, the probability is discrete there are only 6 events and the probability of each event probability of 1 is finite, probability of 2 is finite, 3 is finite and so on. But if you had the case where instead of 6 you had infinitely many outcomes, for example, if this is the p then you imagine this is called a probability distribution p of x and this is x is the x axis and what you have is instead of having these discrete jumps you have a smooth function and because x can take any of these values for each of those values you have a certain value of the probability function. I will come to explain what this probability function is this probability function is in fact, a probability distribution or a density.

Let us suitably define and this is a smooth function. So, it is a function of x because x can take any of these values. Now, x can take any value from here to here and for all of these you have this probability defined this way. So, this is what happens when you go from a finite number of outcomes to infinitely many outcomes and so, since x can take any of these values the number of outcomes of x is actually infinite because x can be any real number. And what happens if you ask what is the probability of a single outcome, now it turns out that that is actually 0 because the you know if you do an experiment many times the chances that you will get a single real number is actually 0 mathematically, but you can always ask what we will shows, that you can ask what is the probability of finding and finding an outcome within some certain interval.

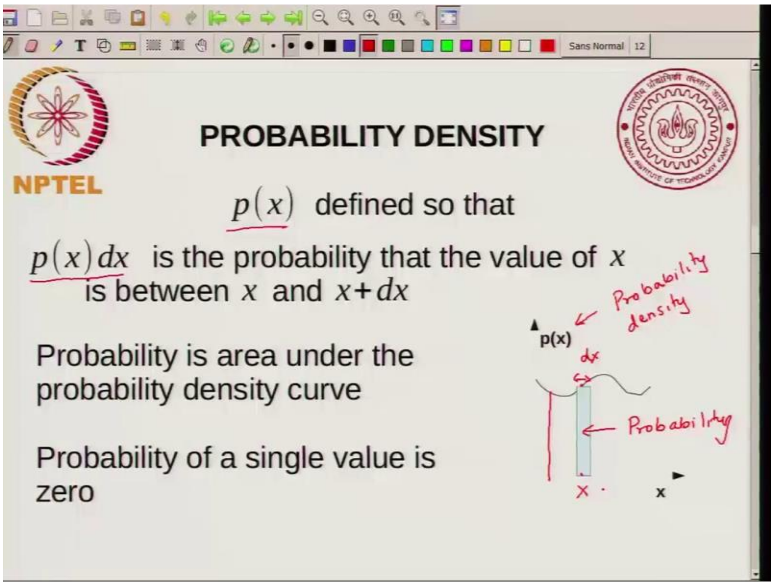
(Refer Slide Time: 08:44)



So, let us talk more about this probability distribution and just to remind ourselves this is a discrete distribution whereas, this is a continuous distribution. So, underlying all this is the idea of random variables. So, here you treat each outcome as a variable x. So, you treat the outcome as a variable x and since outcome in any experiment is not a fixed number then you call x a random variable; that means, when you do an experiment like a roll of a dice you get an outcome. Now that outcome is not fixed it is not determined if I roll it you do not know what you will get, so you think of this outcome of the experiment as a random variable. And now you can talk about probability density or probability distribution of that random variable. So, as we said before we talked about probability of the outcome and here you have probability distribution of the random variable x. So, again this is the function.

Now again this terminology of random variables is fairly standard in statistics and we have deliberately chosen this terminology because we can talk about probability distributions and actually the whole basis of lot of things that you do when you are reporting experimental data is based on this theory of random variables.

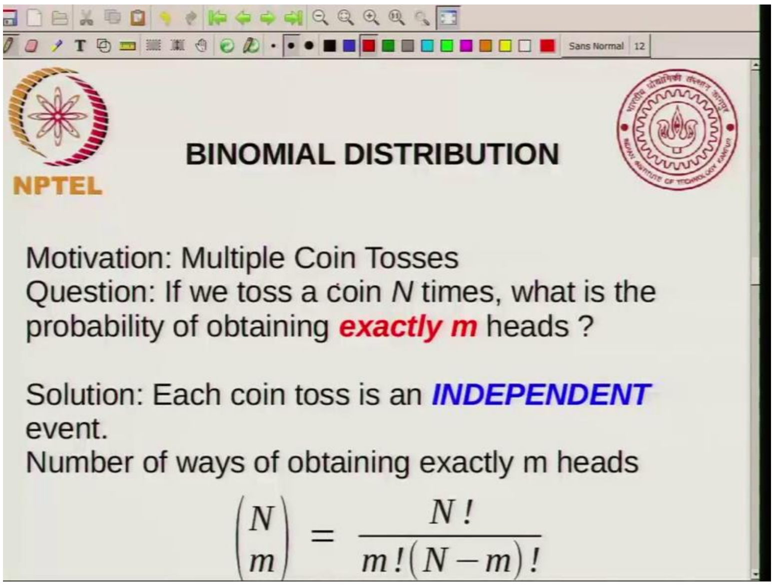
(Refer Slide Time: 10:00)



So, let us look at the probability density in a little more detail. So, the probability density is actually this function p of x and p of x is actually defined in the following way p of x times dx, dx is some small interval in x, which I have shown here. So, this is my dx and p of x times dx is basically this area, so this is the point x and this is this interval d x. So, p of x times dx is the probability that the value of x is between x and x plus d x. So, this is the definition of the probability density. So, p of x itself is not a probability, but p of x multiplied by dx that gives you a probability and this is how we define the probability density. So, p of x is defined such that p of x multiplied by dx is a probability that the value of x is between x and x plus d x. So, you can think of probability as a area under the probability density curve.

So, what we are plotting here this p of x is a probability density and the area under it this area is the probability. So, the probability that x is between x and x plus dx is this, you can ask what is the probability that x is between any 2 points is just the area under the curve under those points and this explains why the probability of x being a single value is 0 because if you just have a single value then there is no area, there is no area under it. So, there is no area to see. So, the probability of a single value is 0. So, what I want to emphasize is that when you go from discrete events to continuous set of events then your probability is replaced by a probability density.

(Refer Slide Time: 12:05)

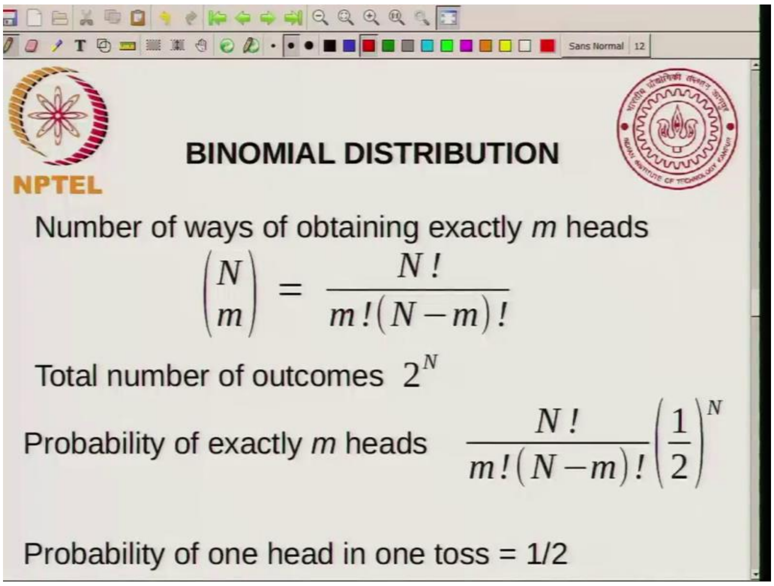


Now, there are many types of probability distributions. So, the probability distribution that refers to what is the probability of different events. So, the set of probabilities of all events defines a probability distribution and there are many common distributions that are that are seen and I will just talk about a few probability distributions.

So, the first one and this is the simplest one. So, that is called the binomial distribution and to motivate this you imagine that you take a coin a 2 headed coin. So, it has a heads or tails and you toss it, you toss it multiple times. So, let us ask a question suppose I take a coin toss it exactly N times; what is the probability that I will get exactly m heads. So, N and m are 2 numbers; obviously, m is less than N and what you are doing is you imagine that you take a coin you toss it N number of times, N can be some large number and you ask what is the probability that you will exactly m times you will get heads exactly m, the word exactly is very important in the binomial distribution. So, not more than m not less than m, exactly m heads.

And if you want to answer this question then what you will see, what you will immediately notice is that each coin toss is an independent event, each toss of the coin. So, the first time I toss the coin there is 50 percent chance of getting heads, second time I toss it again there is 50 percent chance of getting heads. So, since each coin toss is completely independent if I want to get exactly m heads then how many ways can I obtained it; that means, out of my total N tosses m of them have to be heads. So, I have to choose exactly m out of N which are going to be heads and that is exactly this N choose m which is N factorial divided by m factorial N minus m factorial. So, this is exactly the number of ways you can, exactly the number of total ways in which you can obtain m heads.

(Refer Slide Time: 14:18)

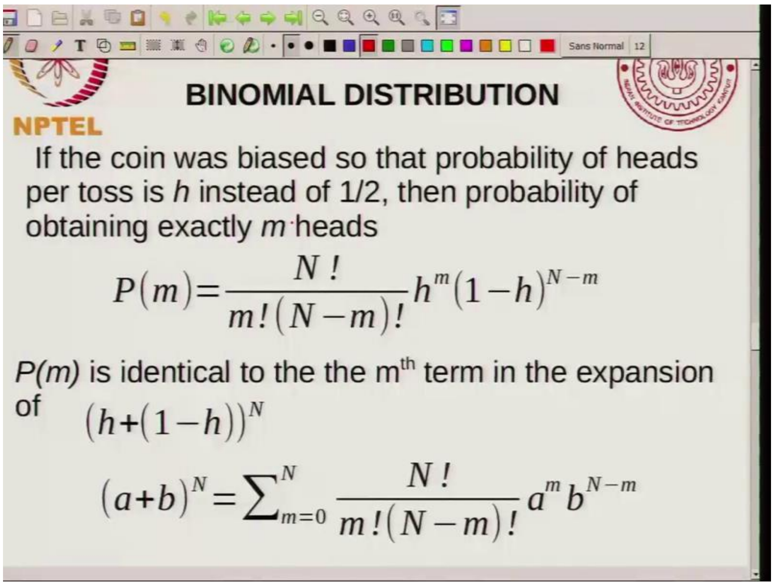


So, the number of ways of obtaining exactly m heads is equal to N factorial divided by m factorial N minus m factorial and the next thing we can ask is; what is the total number of outcomes? So, the total number of outcomes is 2 raised to N because each time you toss the coin you can get 2 outcomes. So, if you toss it 2, 2 times you can get 4 outcomes. So, you can get 2 heads, 2 tails, first one head second one head tails or the second one head and the first one tails. So, there are exactly 4 possible outcomes. So, similarly if you toss it N times there are 2 raised to N outcomes.

So, now this is the total number of outcomes and this is the number of outcomes in which you have exactly m heads. So, the probability of getting m heads is just a ratio of this to this and so that is written in this form. I have deliberately chosen to write it as 1 over 2 raised to N multiplied by this combinatorial factor. And why I chose to write it as half raised to N is that the probability of 1 head and 1 toss is just half. So, in any toss the probability that you get a head is half. So far this is very good because what you can say is the probability of getting m heads is given by this combinatorial factor times half raised to N.

So, now notice that you can think of this as this half raised to N factor in a slightly different way, you can think of it as m times you have to get heads the probability of that is half raised to m and then N minus m times you have to get tails. So, the probability of that is half raised to N minus m and these 2 multiplied will give you the factor of half raised to N and this is useful because now we can generalize this case to a case where the coin is biased.

(Refer Slide Time: 16:04)



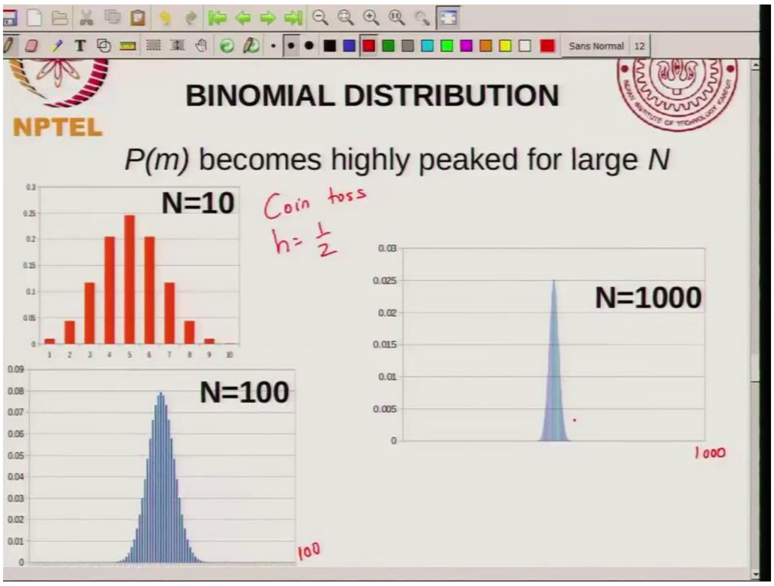
So, suppose the coin were biased. So, the probability of heads per toss in each toss you do not have 50 percent heads, but you have h, h is a fraction which is less than 1. So, you have some fraction. So, it h it might not be exactly half it might not be 0.5, it might be say 0.6. So, you have a coin that gives that is slightly biased, it might be slightly misshaped or for some reason it gives heads more often than tails. So, h is the fraction of times it gives heads.

Now by the same argument you can show that the probability of obtaining exactly m heads is this factor, now you have instead of half raised to N you have h raised to m and 1 minus h raised to N minus m. So, 1 minus h is nothing, but the probability of getting tails, of tails. So, the probability that to your coin when you toss it shows the tails is just 1 minus h and notice that you have N minus m tails, if you have a exactly m heads you have m heads and N minus m tails and this is what turns out and this is called the binomial distribution. So, binomial distribution means you have 2 possible outcomes and you repeat the experiment many times then what is the probability of obtaining exactly m of 1 possible outcome and that is given by this by this function and why it is called the binomial distribution because this is exactly identical to the m th term in the expansion of h plus 1 plus h raised to N.

So, suppose you have a binomial expansion a plus b raised to N then you will immediately write this as sum over m equal to 0 to N, N factorial divided by m factorial times N minus m factorial a raised to m, b raised to N minus m. That is what you would write now instead of a plus b raised to n if you have h plus 1 minus h raised to n then these terms would exactly correspond to your P of m. So, all these terms in the binomial expansion will look just like this, if you had h plus 1 minus h raise to N and that is why this is called the binomial distribution.

Now when you notice that this is nothing, but the m th term in this expansion then we know that if you sum over m equal to 0 to N of all the probability you will just get 1 because h plus 1 minus h is nothing, but 1. So, if you sum over all these you will just get 1 and so that is expected because the sum of probability should add up to 1.

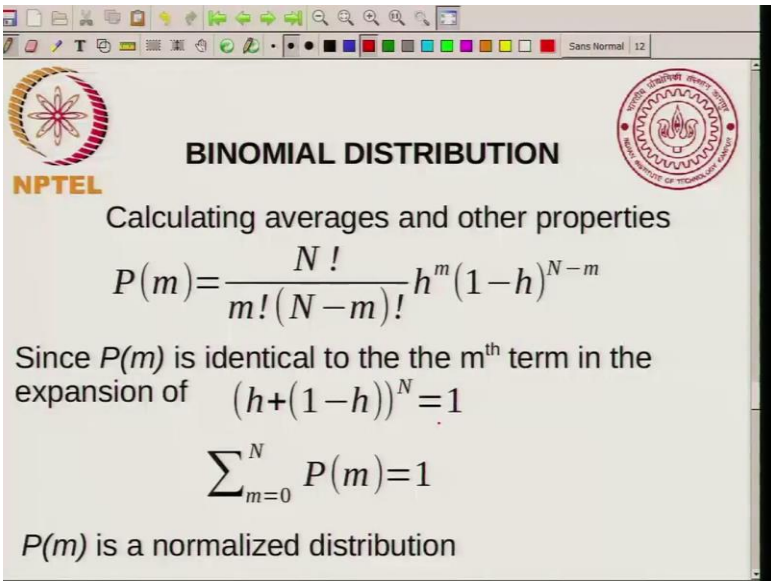
(Refer Slide Time: 18:55)



So, what does the binomial distribution look like? So, suppose you had N equal to 10, suppose you had only 10 and this is what I am plotting is P of m. So, the probability that you get exactly 1 head, so this is a coin toss; so this is coin toss h equal to half. So, we are considering the case when the probability of getting heads is just half and if you do the experiment 10 times. So, if you do a total of 10 events then the probability of getting 5 heads is this much probability of getting 6 and 4 is this and you know this is about 0.246 and 4.2545 and then it goes down probability of getting exactly 1 head is very small, probability of getting exactly all 10 heads is also very small, so what happens is that this P of m becomes highly peaked for large.

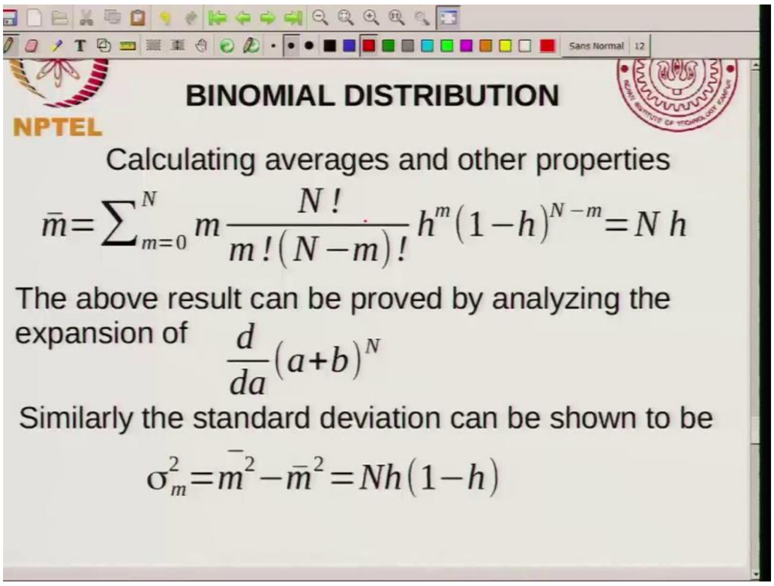
Now suppose you had N equal to 100 then the probabilities now become more sharply peaked you see that this is 0, this is 100, so this is 100. So, you see that the probabilities of getting 1, 2, 3, 4, are all very small it is only probability is getting from 35 to about 65 which are actually relevant which are finite and when you see that it becomes very sharply peaked. If N is 1000 then this becomes even more sharply peaked. So, what you see is that it becomes very narrow there are only a small range of small fraction of the values where the probability is actually significant in all the other cases is almost 0. So, this is a property of the binomial distribution that it becomes highly peaked; that means, it becomes very very narrow as you make enlarge.

(Refer Slide Time: 20:49)



Now, what can you do with the distribution? You can calculate averages and other properties knowing the distribution. So, as we said earlier that since P m is identical to the m th term in the expansion of h plus 1 plus h raised to N, this is just equal to 1. So, you can immediately see that sum over m equal to 0, P m equal to 1. So, in other words p is a normalized distribution; that means, all the probabilities add up to 1.

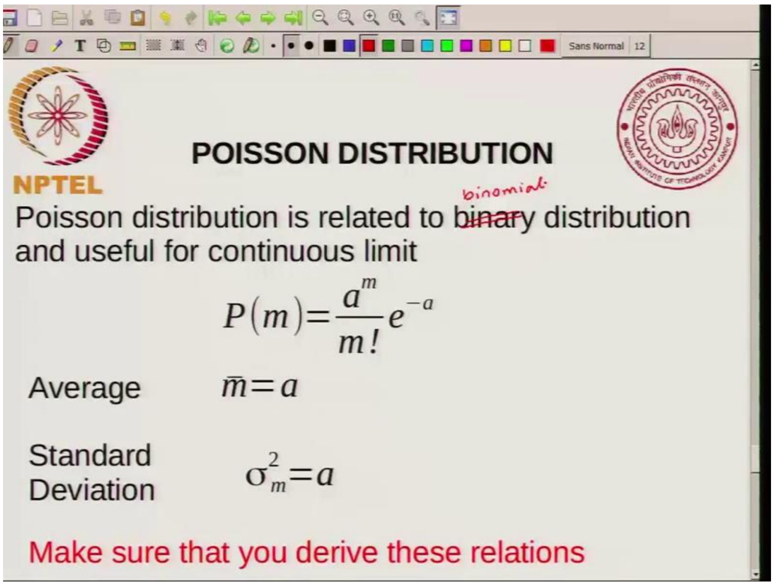
(Refer Slide Time: 21:15)



Now, let us calculate averages. So, what is the average value of m? You can calculate that by doing a sum over all m, m equal to 0 to N, m multiplied by P m. So, this whole thing is just P m. So, multiplied by m and you sum over m equal to 0 to N, that is the definition of the average value of m and if you work this out you can show that this result is equal to N times h.

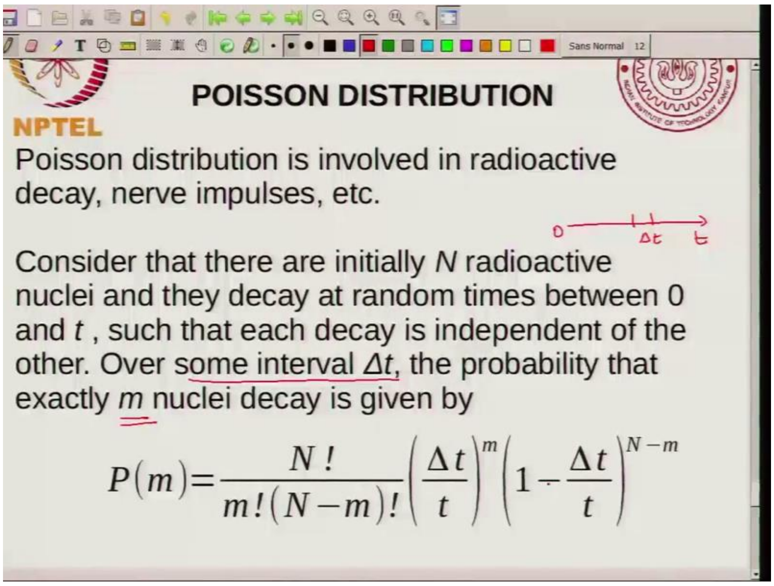
Now it is not it is not actually very simple to show this result going from this expression to this is not that straightforward, but you can prove it by analyzing the expansion of this. So, you can analyze the expansion of a plus b raised to N times d by da of that and you can finally, work it out and show that, and I leave this as an exercise for you to work out. Similarly the standard deviation can be shown to be. So, the standard deviation is defined as the average of m square. So, instead of m you put m square and you take the average and then you take the average of m which is this quantity N h and you square it, and you take the difference of those two you will get the square of the standard deviation, and this can be shown to be equal to N h times 1 minus h. Again I leave both these as exercises for you and they are not absolutely simple, but you need to know a few tricks to in order to derive them.

(Refer Slide Time: 22:54)



Now, the binomial distribution is also you know there are other distributions that are related to the binomial distributions which are fairly useful, one of them is which is very well known as a Poisson distribution and this is related to the binomial distribution you should read binomial, binomial distribution and it is useful for taking the continuous limit, the limit where things become continuous and the definition is that P of m is given by a raised to m divided by m factorial times e raised to minus a. So, this is what is called the Poisson distribution. And if you calculate the average in a Poisson distribution this average of m using this distribution you will get you can show that this is equal to a, this is fairly easy to show. Similarly you can calculate the standard deviation or the variance sigma m square and that is also equal to a. So, again I request you to make sure that you actually derive these relations.

(Refer Slide Time: 23:54)

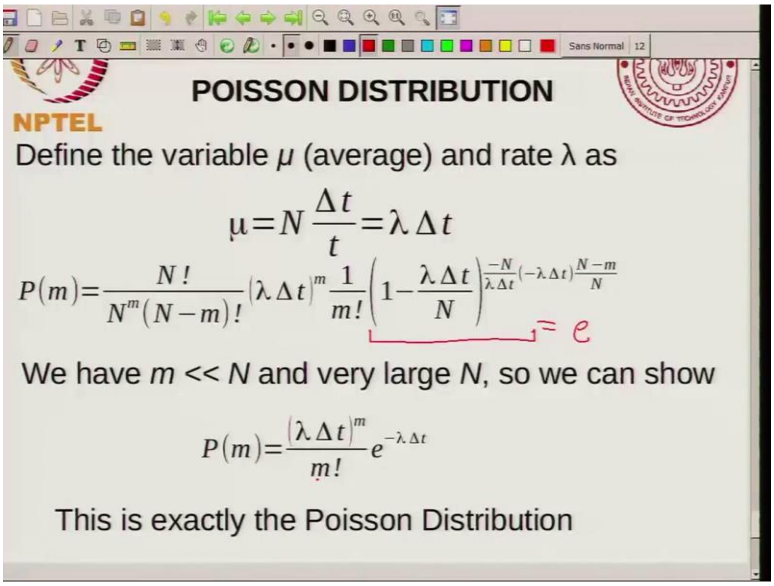


Now, how do we understand the Poisson distribution? Poisson distribution some of you might have heard is something that you see in radioactive decay it is also seen and you know when you are analyzing nerve impulses and we will just take an example suppose you consider initially that there are N radioactive nuclei and these decay at random times between 0 and t such that each decay is independent of the other. So, at time t equal to 0 you have N radioactive nuclei then what happens is as time goes along at some point one of the nuclei will go, we will decay, at another point some other nuclei will decay and this goes on. And each nucleus is decaying independent of each other.

Then you can ask a question over some interval of time delta t. So, in this time from 0 to t you can imagine some interval of time delta t. What is the probability that exactly m nuclei decay during this time delta t. So, just to show you have you have 0 to t and you have some interval delta t and you are asking what is the probability that in this interval there are m nuclei that decay there are m nuclei the decay in this interval. So, m nuclei decay in over this interval delta t. So, the probability of this is given by, it is you can justify this, so the probability of any nuclei, any 1 nuclei decaying in this time delta t is just delta t divide divided by t because the nuclei decays any time between 0 and t probability that it decays in this interval is just the fraction of this interval divided by the total time that is delta t by t and so since each nuclei is independent of each other this is nothing, but you can say that the probability of m nuclei decaying is just given by the binomial distribution where the probability of any 1 nuclei decaying is just delta t by t and the probability of a nuclei non decaying is 1 minus delta t by t.

So, once you have this P of m in this form what you need to do to get a Poisson distribution is to take this and consider the case where N is very large.

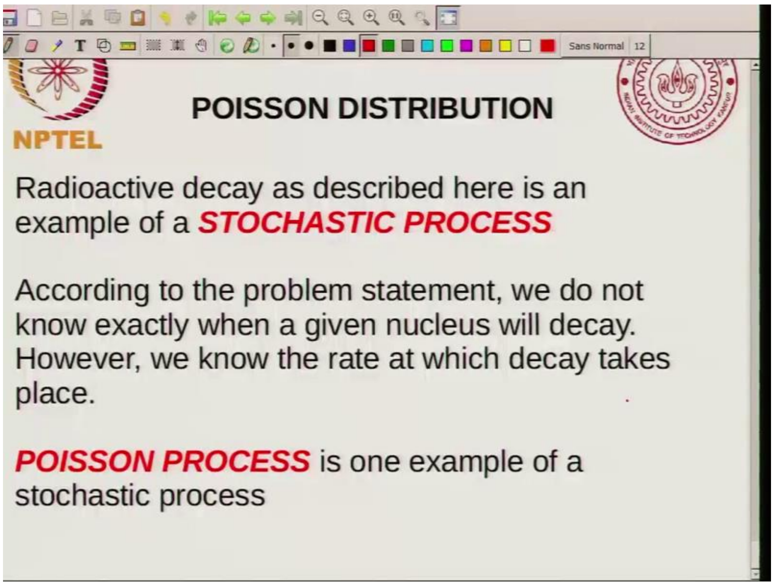
(Refer Slide Time: 26:12)



So, first we defined the variable mu which is the average and the rate in this form. So, mu is defined as n times delta t by t which is nothing, but N times the probability of 1 nucleus decaying which is like N times h in the binomial distribution. So, it will be equal to the average and just we defined a quantity lambda. So, we defined this average as N as lambda times delta t and with this you can write the Poisson distribution in this form. So, what I have done is I have just replaced delta t by t as lambda delta t and divided by N. I have taken the N raised to m outside and I have just chosen to collect terms in this form and what I have is 1 over m and I have 1 minus lambda delta t by N raised to minus N divided by lambda delta t times minus lambda delta t. So, so basically I had I mean you can think of I had N minus m, but what I wrote it as N minus m by N times N by lambda delta t and lambda delta t.

The reason for writing it this way is that you immediately identify that when N is very large this quantity is equal to e that is that is the definition of e, e is the transcendental number e. So, then this whole thing just becomes e and so this becomes e raised to minus lambda delta t and when N is large and m is less than less than m this factor goes to 1 and again this whole factor will go to 1. So, finally, you will just be left with this. So, this is exactly the Poisson distribution and so what we did is we got the Poisson distribution as a limiting case of the binomial distribution.

(Refer Slide Time: 28:00)



And just to conclude what I will say is that the Poisson distribution that we are shown here and we discussed radioactive decay this is an example of a stochastic process. So, the radioactive decay is a stochastic process that occurs at any random instant in time and according to the problem statement we do not know exactly when a given nucleus will decay; however, we know the rate at which the decay takes place and this Poisson process is one example of a stochastic process.

So, in the next lecture I will talk about the Gaussian distribution and so I will end, I will complete this lecture here.