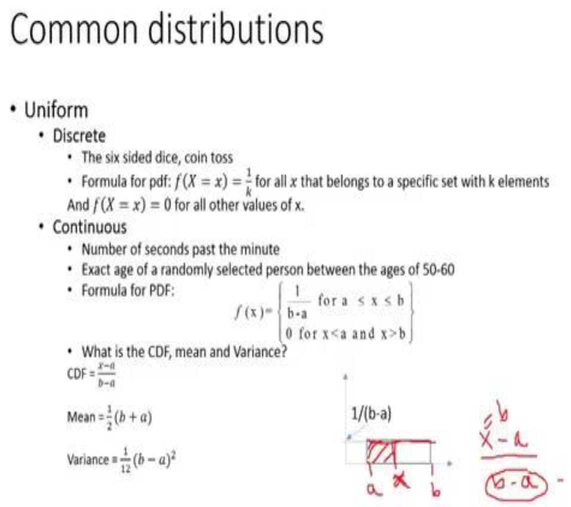
Introduction to Data Analytics
Prof. Nandan Sudarsanam and
Prof. B. Ravindran
Department of Management Studies and
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Module - 02
Lecture - 07
Random Variables and Probability Distributions-2

Hello and welcome to our second lecture in this series on Random Variables and Probability Distributions for a course Introduction to Data Analytics. In the previous lecture we saw, we made a broader introduction to the concept of random variables and distributions, we spoke about the concept of having continuous distributions and discrete distribution, we presented some examples.

And we also spoke about the idea of a probability density function versus cumulative density function. In today's lecture, we are going to pick a 5 or 6 very common distributions and discuss them one at a time. And the first one that and these distributions are going to be both, some of them are going to be discrete, some of them are going to continuous and some of them could be both.

(Refer Slide Time: 01:09)



So, we start with the most common distribution, which is the uniform distribution. The uniform distribution has a discrete version and a continuous version. Now, we will start with the simplest one, which is the discrete version. You already saw examples of this. So, for instance the example that we saw on the last class of the six sided dice, where we were quantifying the probability of getting any particular face values.

So, the face value of the dice is essentially, you throw the dice you get something on top, so you see the 1 or 2 or 3 or 4 or 5 or 6. So, the probabilities associated with these 6 possible outcomes, assuming a fair dice is one sixth for each and, so that is example of discrete uniform distribution. We also saw the case of the coin toss, where here you have only two possible outcomes and as long as they are both equal, it is still uniform.

So, the formula in terms of the PDF the Probability Density Function, which we discussed in the last class, it is fairly straight forward, it is just 1/k when there are k outcomes. So, if this is 6 sided dice it is 1 by 6 for each of those six possible outcomes. If it is coin toss, which is two possible outcomes it is going to be 1 by 2 and the ideas that for each of those k possibilities, it is 1 by k and for the rest of the universe, it is 0.

So, the probability of getting a 7 when you role a 6, a single 6 sided dice is 0, so you cannot get 7 and you cannot get minus 45 either, so that is what this formula says. With the continuous version, here again you are looking at a uniform distribution, so the probability is a uniform, but like we discussed the variable that we are quantifying is continuous. So, the variable in the six sided dice was, what is the number that shows up and that number is either 1, 2, 3, 4, 5, or 6 and that 6 discrete possibilities, but you might have many things that are not discrete and this goes back all the way to our discussions and quantitative variables, which can be continuous or discrete.

So, examples of this and the truth is the uniform distribution, while it is theoretically very intuitive, could be quite convenient in some cases. There are not a lot of examples, real world things that tend to be uniforms, some of them are something like number of seconds pass the minute. So, if you were to randomly, if you have to have random process, which just looked at the clock over the course of date during some random intervals, the number of the seconds pass the last minute could be uniform and that is essentially a space, where you can have 0 seconds pass the minute all the way to 60 seconds pass the minute.

So, your interval is from 0 to 60 and where you find yourself in that interval is described potentially by uniform distribution. And again the idea that it is continuous, just means that you not discretized the seconds as 1 or 2 or 3 all the way to 60, you could be in that in 4.5432 seconds pass the minute. So, it is continuous, the variable time is being monitored continuous as a continuous variable.

Another example could be the exact age of the randomly selected person between the ages of 50 and 60 perhaps in a certain country. Now, again you know you have to come up with fairly specific examples, because something simple like the age of the randomly selected person is not likely to be uniform. You are not likely to find as many people between 80 and 90, as you are between 70 and 80.

And even if you stop the clock at a certain point with typically with things like age, you see that the number of people or the probability of a person in that window decreasing as sometimes age increases and it depends on your state space and it depends on your countries, is the population increasing decreasing, so on and so forth. But, sometimes over a short enough interval, even if the overall distribution is not uniform you can create an interval and say that is my universe.

So, my universe is people between the age 30 to 40, within that interval perhaps the distribution of the exact age of people. So, imagine I take the universe of all people between the ages of 30 to 40 in India and then, I try to create a distribution of the exact age of a randomly selected person from this bucket and that could potentially be a uniform. The uniform is also a great distribution, think of when you want to make absolutely no assumptions.

So, if I want to make the assumption that the country's population is and some of these need not even be assumptions, so it could be a fact. So, something simple like the country's population is increased and so on, then I would be hard pressed to come up with a uniform distribution and I might find richer distribution to represent my data. But, this uniform distribution can be thought of us like, this ground 0 I do not make any assumptions, the probability of everything occurring within a certain interval is equal and, so I could potentially use it.

I have also come across the usage of this a little bit in different aspects like physics and chemistry and so on, for instance the probability of certain types of molecules be in certain locations over certain space could be uniformly distributed and so on. So, we have spoken about, so the formula for the uniform distribution, we spoke about the discrete uniform PDF and we said, the PDF is essentially 1/k. So, with the continuous, the PDF is essentially $\frac{1}{\mathrm{b}-a}$

So, the idea here is that, this is essentially a and this is essentially b and as you can see in the graph, the probability between a and b is uniform and just like in the discreet case, if this distance b minus a, then the probability is 1 by b minus a. And another way, another quiet simple way of thinking about it is, between this interval, over this entire interval from the lower limit to the upper limit to the area under this curve should be equal to 1 and if you look at the simple math of it, $\frac{1}{\mathrm{b}-a}\ (b-a)$, which should be 1, so the height is $\frac{1}{\mathrm{b}-a}$, the length of this rectangle is b - a.

So, $\frac{1}{\mathrm{b}-a}\ (b-a)$, would give you 1 and, so you can essentially think of it that ways well. And; obviously, like we discussed it has to be 0 for if you are less than a if you are greater than b. So, what is this CDF of this distribution? Again, it is, $\frac{x-a}{\mathrm{b}-a}$, we will not be exactly deriving it out. We will give you formulas at the end of this lecture to give you an idea of, how to get to the CDF, how to get to the mean, how to get to the variance.

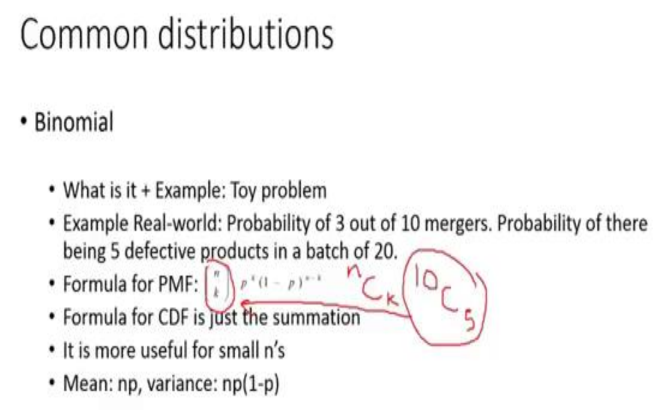
But, the core idea here is that, essentially you take some point and let us call this point x, so this is x, there are better x. And the whole idea is that we know this CDF is essentially the area under the curve to the left of x. So, the question could be if you knew that this height is, $\frac{1}{\mathrm{b}-a}$, how do you go about writing out this area, how do you calculate this area and it can be a function of x. So, that general formula would be the CDF and the idea is fairly simple. If this is x and you are looking at the CDF is a ratio, it is essentially the area, we know that the total area is 1.

So, this total area of the blue rectangle is 1 that we have discussed. So, what percentage of that rectangle have you essentially covered and the idea is, because it is uniform, because this height is constant. If this is x and this is a and this is b, then of the 1, you covered x - a is this area that you have covered and b - a would be the full area that you could be possibly covers. If you think of it as a ratio of, how much have covered and how much I can potentially cover, where this b minus a kind of represents the 1 in some sense of the full area.

Then, I have covered x-a of the b-a that I could cover and therefore, I covered the percentage in some sense, where; obviously, if x was equal to b, then I would have covered the 100 percent and that would be equal to 1, so that is the over all idea behind getting this formula. Again the formula for the mean should be fairly intuitive, if this is a and you know, you have this is b, this central point in some sense is b by you know half of between b and a essentially.

And again with variance we will not derive it, accept to tell you that the core concept is to say how much do we deviate on average from the mean. So, we will talk through some of this formula some of it is to give you an intuition, some of it is to actually give you these problems in the assignment. So, you get a feel for actually figuring out what the mean of variance exactly is.

(Refer Slide Time: 12:03)



Let us move on to the next distribution. The next distribution we are going to talk about is the binomial distribution. So, the binomial distribution is also another distribution, where you have a lot of toy problems associated with it, but by nature of it in the real world sometimes it is more useful to approximate it with another distribution and user, but what exactly is the binomial distribution, let us start there. So, we spoke about this class of distributions and if you did not, then let us just do that right now, so let us.

So, there are these class of distributions where called the Bernoulli distributions and the idea behind that distribution is that, it is very similar to the first example we saw in the previous lecture, where you have an event and it has some probability. So, the 30 percent chance is going to rain, that is what the exact example we use. And, so therefore, the probability that it is not going to rain is 70 percent. What is key about it is that, there are only two possible outcomes and these two possible outcomes, discrete outcomes sum to 100 percent and that class of distribution is called Bernoulli distribution.

So, our standard example of tossing a coin and saying, what is the probability of heads and what is the probability of tails, would be an example for Bernoulli distribution. Now, if just, so that there is no confusion if it happened that the probability of heads and tails are both equal, you can also think of it as a discrete uniform distribution and, but out here we are just saying we saying something different. We saying that Bernoulli distribution are distributions, where there are only two possible outcomes.

The probability of the two outcomes need not be the same as required by the uniform. But, the key out here is that there are only two possible outcomes, two possible discrete outcomes that sum to a 100 percent. So, great, so we spoken about the Bernoulli distribution, what is it happen to do with the binomial distribution, is not that, what this slide is about. Well, if you take the problem and you quantify the probability of heads and tails, then you are talking about Bernoulli.

But, instead if I rephrase that problem and said, what is the probability of getting 5 heads out of 10 tosses. Then, I am describing the problem associated with a binomial distribution. So, an example of a binomial distribution would be, if I said I am going to toss the coin 10 times, I am going to take part in a Bernoulli's process, n number of times and I asked myself the question, what is the probability of getting k successes and success could be defined as getting a heads or it raining or whatever it is. It is one of those two outcomes essentially; you call one of those two possible outcomes of a Bernoulli process as a success.

And then, you say, what is the probability of getting k successes out of n possible trials? So, if I say I am going to toss a coin 10 times, can you tell me the probability associated with 0 heads, 1 head, 2 heads, 3 heads, 4 heads all the way up to 10 possible heads. Now, note if I toss the coin 10 times, I can either get 0 heads or any of those numbers in between till 10 heads, I cannot get minus 1 heads I cannot get 11 heads. So, the probability associated, if I toss the coin 10 times, the some probability that I am going to get 0 heads, the sum probability that I am going to get 5 heads.

But, each of these numbers each of these probabilities when added together should again be equal to 100 percent or 1. So, quantifying the probability of getting k successes out of n trials of a Bernoulli process is the binomial distribution and you can think a various real world example. So, for instance the probability of, you know let say there are 10 mergers the companies considering.

So, in mergers and acquisitions since it is a small enough number, what is the probability of getting 3 out of 10 of them or you might say, probability of having 5 defective products in a batch of 20 products, what is the probability of the 3 defective products, answering questions like that. And one just I noticed, you might have noticed that I have take an example of small enough a numbers.

So, technically the binomial could be answering a question like, what is the probability that out of you know 1 million possible toys. Let us say that I distribute, I am a toy maker and I send out 1 million, what is the probability that I get 500, 5000 toys as broken. And technically that would still be a binomial distribution, but what we will learn probably in the next classes, why that… The computation of that is a little messy, you get very large numbers and you have some other distributions that can approximate something like this to solve problems like that.

But, the core concept is this. You have a Bernoulli process, where something is… You are looking at something that is binary a or b, 1 or 0 and you turn around and you say, what is the probability of getting k successes out of n trails of this process. And, so logically the formula and note that I have used the word PMF of here, that is an important distinction that is worth mentioning. PMF is the same thing as PDF, PDF Probability Density Function, we spoke about that we spoke about probability density functions and cumulative density functions.

PMF is just the standard way of calling a PDF if you are dealing with a discrete distribution. And since this is a discrete distribution, because you cannot get out of 10 tosses of a coin you cannot get 3 and half heads. So, you can only get either 0 heads, 1 head, 2 head or 3 head and so. Since this is a discrete distribution, it is technically called Probability Mass Function or PMF. So, the PMF, this distribution should be fairly intuitive, it is nothing but, n choose k, so this symbol out here just means you might have seen it like this.

So, that is n choose k and that is has to do with a combinations, something you might have studied in the permutations and combinations. The idea is, how many ways are there of choosing n from k trails. So, here I think the words n and k are swapped as suppose to, what I was mentioning earlier here. So, here for instance I was interested in finding out, what is the probability of getting 5 heads, then it would really be 10 would be the n c 5.

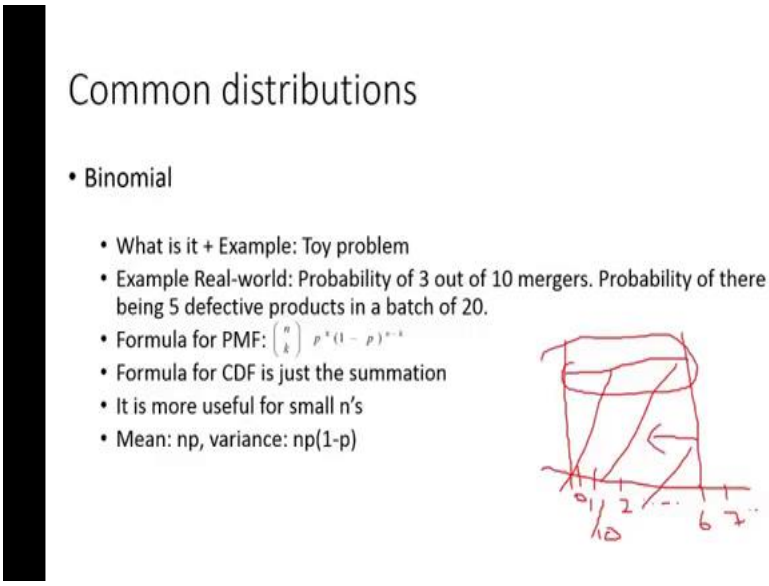
So, how many ways are there of getting 5 of choosing 5 out of 10, so I toss the coin 10 times, there are many ways in which, I could get 5 heads out of 10 tosses. It is either that the first 5 could all be heads and the next 5 could all be tails or you can have 1 head 1 tail 1 head 1 tail. So, that number of different ways in which I can get 5 heads out of 10 tosses is, what is being quantified by this number, which relates to this part of the formula. Once I figure out the numbers of possible ways, this part, which is p power k tells me, what is the probability of getting the 5 heads in the first place.

So, it might be 0.5 power 5, if the probability of head was different from a tails, let us say we were dealing with the problem where 60 percent chance of heads following, because it is an un even coin, then it could be 0., it will be 0.6^5 and this part, which is the remaining part talks about the probability of getting the tails, the remaining 5 as tails. So, essentially the 3 parts of this formula are the different ways of getting those heads, probability of getting so many of those heads, probability of getting the remaining number as tails.

So, that is essentially, what the PDF captures and the formula for the cumulative density function again is the same logic that we were discussing with the uniform, which is that for a given k, which might mean for let us say I am interested in knowing, what is the probability of 6 heads out of 10 tosses. You are essentially looking at nothing but, everything to the left of the curve and to the left of the curve here means, what is the probability of getting 0 tails I mean.

So, let us see what is the probability of getting 6 heads is the question out of 10 tosses and that would be nothing but, the PDF of getting 0 heads out of 10 tosses, 1 head out of 10 tosses, 2 heads out of 10 tosses, all the way up to 6 heads out of 10 tosses, that summation of those probabilities, because that is essentially what would be to the left of the curve.

(Refer Slide Time: 21:44)



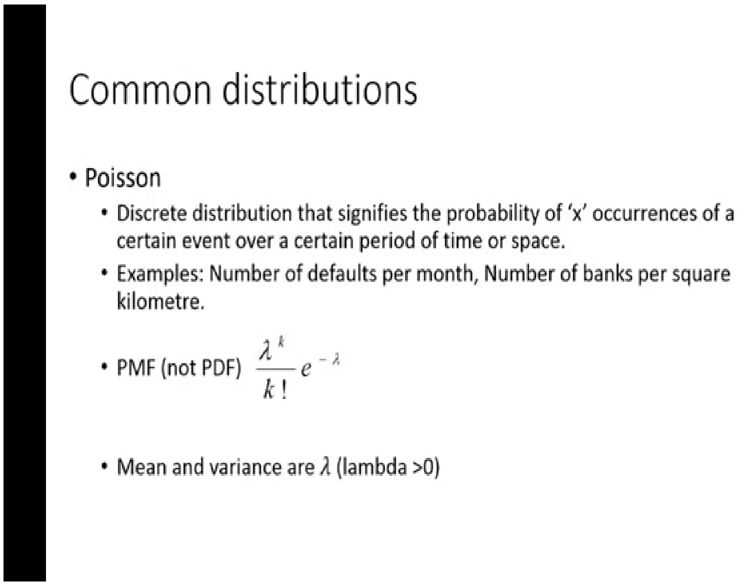
And here I am doing nothing but, thinking of a curve, where 1 out of 10 tosses. We should start with 0 out of 10 tosses and going all the way to, you know 2 ,…, 6 and then, it will go 7 and it will go all the way till 10 out of 10 tosses. So, if this was some kind of PDF, I am essentially looking for this area and the curve and you know, the top part of the line is not uniform I am just representing, which side I am interested in, that is all I am doing there and also it would be discrete.

So, you have discrete rectangles popping out of the 1, 2 and 3 and you just summing them up and the formula also with the summation does not really simplify too much, so you are dealing with the CDF as it is. As I mention it is more useful for small values of n, when you get really large values of n, you took it other approximations. The mean is nothing but, the number of times you are looking at times of probability.

So, if the probability of getting the heads is 60 percent and you are going to toss the coin 10 times, the average number of heads you are going to receive is nothing but, 10 times 60 percent, which would be 6. So, on average I should get 6 heads out of 10 tosses, because this is 60 percent chance of getting heads. And, so the intuition of that should be fairly obvious and again with variance, the intuition itself might not be very obvious. It is really about looking at how much the deviation, how much of a deviation there is from the mean.

So, how much does one out of 10 deviate from the mean and we would look at the probability that you would first of all see of 1 out of 10. And again I think there the formulas for variance might probably help you to get a better idea of it, great.

(Refer Slide Time: 23:41)



So, let us move to our next distribution this is also a discreet distribution and this is very similar to the binominal in many ways both of them are discrete, but essentially while this is discrete, which also counts the number of possible occurrences x. So, it says, what is this quantifies the probability of having getting x occurrences. So, probability of getting one of those occurrences two of those occurrences, but it is over a certain period of time or space see in the binominal we were saying, what is the probability of 5 heads out of 10 tosses.

So, the 5 heads was discrete but, so was the 10 tosses here you are asking the question, what the probability of x is where x is discrete meanings, what is the probability of there being 5 people. But, instead we won't say out of 10 people we will instead say, what is the probability of there being 5 people coming to this bus stop over the next 5 minutes the key difference is 5 minutes is continuous as oppose to 10 tosses, which is discrete the ramifications of that is; however, low the probability is of people arriving in a given time there is still technically a probability, that you have a really large number of people infinite number of people coming over the next 5 minutes, where as I can say with certainty in the case of the binominal that out of 10 tosses I cannot get 11 heads.

And this has nothing to do with the probability of getting a heads in the first place even if

I had a probability of 0.1 of getting a heads in a toss, because of such an unfair coin I still know or 0.9 whatever you know whatever the probability is I still know that its technically not possible to get 11 heads out of 10 tosses.

Here, you are looking at a discrete occurrence such as probability that of n number of defaults in a given month or you know number of people, who are going to arrive at the bus stop in next 5 minutes the number of defaults the number of people, who arrive in the bus stop are discrete you cannot have less than 0 people I mean this in this particular case you cannot have less than 0 and you cannot have two and half people who arrive at the bus stop, so it is a discrete distribution.

But, the distribution is defined over at time or a space, which is continuous. So, it is not how many heads can I get out of 10 tosses its more, how many of a certain occurrence can happen over a certain period of time and that is the core concept there the PMF again note it is not the PDF, because it is a discrete distribution its characterized as lambda power k and lambda here is a parameter. So, it is essentially a number that that you have and it represents like the average rate.

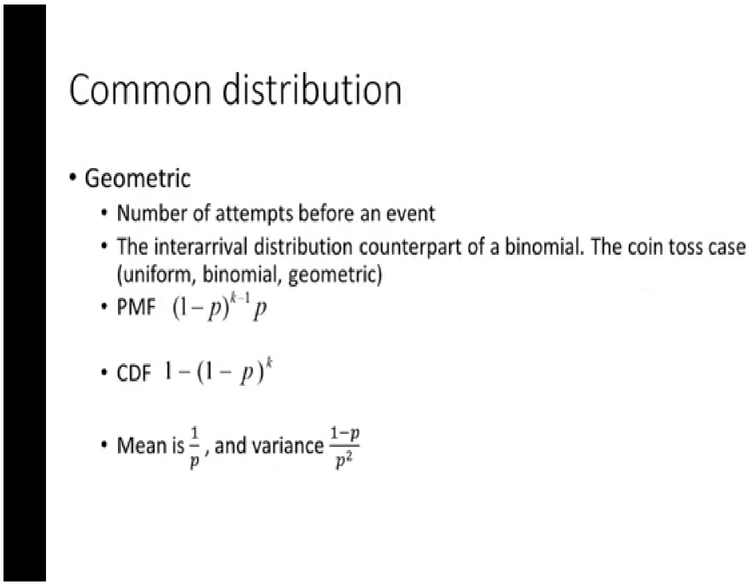
So, 3 people are arriving per minute that would be more like a three out there and k is the variable of interest where you say k is equal to 1 and you get a probability you say k is equal to 0 you get a probability. And technically this k can go all the way up to infinity like I was discussing right and the sum of all those possibilities discrete possibilities is equal to 1. So, if you wanted to know the probability that 3 people will arrive at the bus stand all you will do is you will say, what is the average arrival rate.

So, may be the average arrival rate, which is what is essentially represented by λ. So, the average arrival rate could be something like well two people arrived per minute, so that is two and I want to ask the question what is the probability 0 people will arrive over the next minute. So, I would say on a average two people arrive at the bus stop k is that 0. So, I will put this I will substitute k with 0 and this will this formula will spit out a answer, which is the exact probability of 0 people.

And you will replace k with 1, 2, 3, 4 and so on, all the way up to infinity and the sum of all those probabilities are is essentially, what is essentially the distribution is essentially, what quantifying. Because, of how it is defined the mean is lambda we defined λ as essentially that rate parameter and its interesting property of the that variance also is λ

and λ has to be greater than 0. But, again just to recap this state's space for a Poisson distribution that is the possible value that k can take are always greater than 0 in greater and less than and goes all the way up to infinity. But, it is a discrete distribution, because you can never have two and a half people arriving at a bus stop.

(Refer Slide Time: 28:41)



So, the next distribution we are going to look at is the geometric distribution the geometric distribution is also an discrete distribution, but it is a very interesting counterpart to the binominal distribution. So, take the fact that we said that is a Bernoulli process, which means there is some probability of event happening and some probability that the event will not happen and they add up to one. So, probability of a heads, let us say is 60 percent.

Therefore, by definition probability of the tails is 40 percent that was Bernoulli and then, we said the bi nominal is nothing but, the probability of getting k successes out of n trails. So, what is the probability of getting 2 heads out of 10 tosses 3 heads out of 10 tosses that space is defined by the binominal the geometric defines the probability of the number of times you need to toss the coin before getting your first heads or you can think of it as getting the next heads.

So, number of attempts before an event is what you are looking at and before you can think of it as in inter arrival counterpart to your binominal distribution. So, if you take the coin toss case right essentially it is more like, how many times do I need to toss the coin before I get my first heads or a next heads. So, you start at some point and then, you say so; obviously, anything that I have tossed before it is not influence my future tosses, because they are independent.

And now, I am going to start tossing the coin and tell me the probability that I will have to toss the coin 0 times before I get my first head or 1 you can count you can say, what is the number of tails I am going to see before I get my next heads. So, that could be 0,1,2,3 and technically it can be infinity meaning they could be this really bizarre world, where even though there is a finite non zero probability of getting a heads whatever that number is it could be 0.9 or it could be 0.1 its technically possible that I wind up tossing the coin and infinite number of times I keep getting tails and, so am still waiting for my heads.

So, this is again a distribution, which starts with 0 and goes all the way to infinity depending on how it's defined exactly the geometry is sometimes defined as the number of tails you see before the next heads. And there the distribution starts at 0, because you can see 0 tails another version of this distribution can could start with, how many tosses do you need to make to see the first heads in that case the very you need to at least toss the coin 1 time to see the heads.

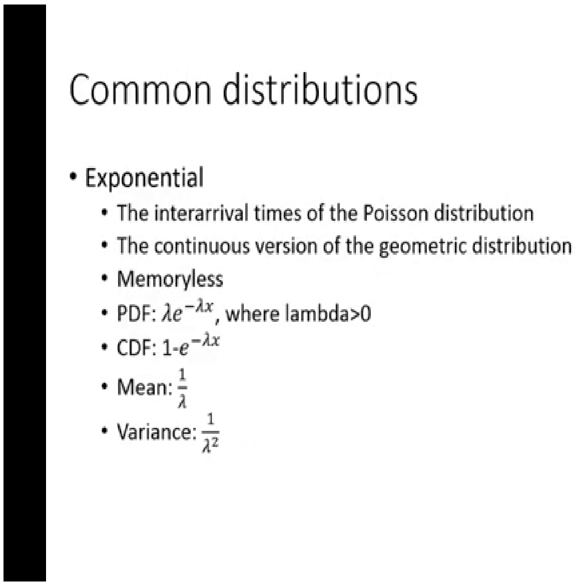
And, so both these distributions you might find in text books the PMF and CDF should be fairly intuitive all we are doing with the probability mass function you are saying, what is the probability of getting k minus 1. So, here you might say k minus 1 is nothing but, the number of tosses before the actual success that you keep getting the tails. So, if p is the probability of getting the heads you are saying, what is the probability of getting k minus 1 tails.

So, if you here we are defining the distribution as on the first toss I get the heads then what is the probability that, so let us say I want to know the probability of getting it taking 3 tosses before I get a heads, then you are saying; that means, for the first two tosses 3 - 1 is 2 the first 2 tosses I should have gotten a tails. So, what is the probability of getting a tail like that and that is 1 - p.

So, $(1 - p)^{(k-1)}$ says before that success happens I need to say I need to say k-1 failures and what is the probability of that and then finally, the probability of that one success. Again essentially the CDF you are doing a summation, but the summation kind of neatly simplifies to the formula that we have shown here and the mean being 1/p should be fairly intuitive meaning if the probability of getting a heads is let us say 10 percent it should be intuitive that it on average should take 10 tosses before I get my first heads, so 1 by 0.1 would be 10 and so on.

And again even if it is not entirely intuitive to you may be working through some problems and formulas, where we will be discussing those can help. Finally, the variance is also $\frac{1-p}{p^2}$ and even that is something that you might have to work through a couple of times.

(Refer Slide Time: 33:45)



So, we come to the last distribution that we are going to be discussing in this class and that distribution is the exponential distribution in many ways the exponential distribution essentially you know how the geometric was looking at the inter arrival time of a binominal distribution the same way the exponential looks at the inter arrival times of the Poisson distribution. So, what does that mean in terms of our examples you know how we spoke about the Poisson distribution as being discrete distribution, where you said, what is the possibility of three people arriving at the bus stop in the next 5 minutes what is the probability of two people arriving at the bus stop in the next 5 minutes, so on.

So, you have fixed the time 5 minutes or ten minutes whatever is of your interest and you looked at the probabilities of various discrete possible occurrences. So, you said what is the probability zero people arrived one person arrives two people arrive three people arrives. Here with the exponential you are describing the inter arrival that is how long should I wait before the next person arrives.

So, it is the exact same thing the exponential is the same thing to the Poisson the way the geometric is to the binomial. The binomial is quantifying the number of the probabilities of getting 3 out of 10 out of 10 tosses the probability of getting 3 heads, where as the geometric saying how long should I wait before the next head heads arrives. The Poisson is quantifying over some time scale or space scale its quantifying the probabilities of different occurrences like, what is the probability of 1 occurring 2 occurring 3 occurring this is how long should I wait before the next thing occurs.

So, you can think of the exponential as moving in time and waiting for that next occurrence and how long should I wait, what you can think of it a space I keep walking and you know I encounter occurrences over some length scale the poisons gives me the probability of seeing n number of such occurrences. But, if I start walking on that scale how long should I walk before I get the next occurrence and long can be in length or it can be in time.

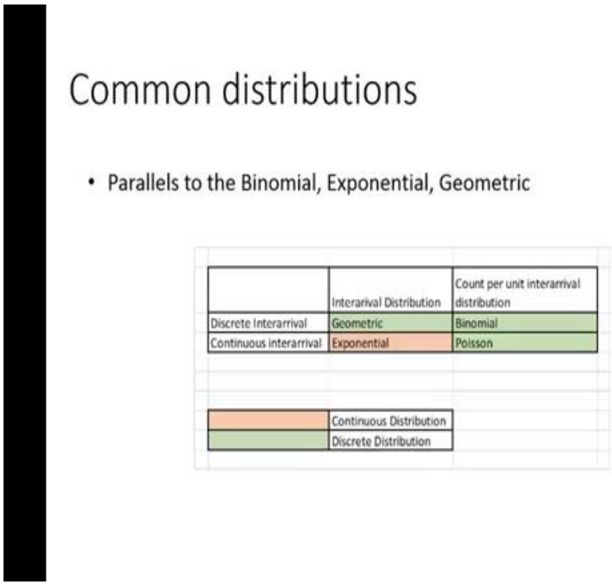
So, think of it is time or space, but essentially the continues version of the you can think of it as the continuous version of the geometric distribution. The more again the PDFs and CDFs it is there for your reference, but the important thing about this distribution is that people call it memory less meaning that the probability of something occurring over a time if the same if you condition that it is not happened yet. So, think of this way I am not saying for instance, so let us think exponential is often used to say to describe may be the failures of a light bulb over time.

So, how long should I wait before this light bulb fails, but as Poisson would describe saying out of 100 light bulbs over a tenure horizon how many would fail, but let us leave that. So, out here an exponential am dealing with one light bulb I am saying how long should I wait before it fails. So, if this distribution were uniform the probability that it would fail between year 1 and 2 would be the same as probability that it would fail between year 5 and 6 that would be uniform what we mean by seeing exponential is memory less is that.

The probability that the bulb would fail between the year 1 and 2 is the same as the probability of the bulb would fail between year 5 and 6 if I tell you at the start of year 5 that the bulb is already not failed. So, if you are standing at time 0 the probability the bulb will fail between year 1 and 2 is very different from the probability the bulb will fail between the year 5 and 6.

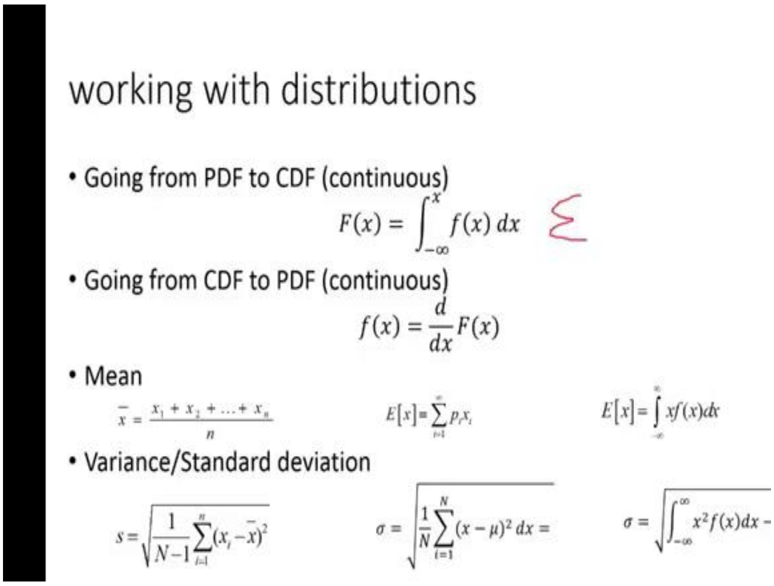
But, the probability that the bulb will fail between year 1 and 2 is the same as the probability of the bulb will fail between 5 and 6 if I go to year 5 and tell you that the bulb has not failed yet same way if I go to year one and tell you that the bulb has not failed yet. So, condition on the bulb not failing the probability is over a future time horizon or length horizon would be the same.

(Refer Slide Time: 38:08)



And just for your convenience we have kind of put together the four distributions that are partly related and saying that these are discrete over the time arrivals versus how the distribution is counted. And I have used the color coding to describe which are continuous distribution and which are discrete distribution and you can see that clearly its only the exponential that is continuous the other three are discrete distributions great.

(Refer Slide Time: 38:34)



Now, before I sign off from this class I just wanted to give you some formulas on how to calculate mean and standard deviation and so on. But, before we do that let us start with the more basic thing, which is given a PDF, how are you going to get the cumulative density function and the idea here is we have always said it is the area to the left side of the curve. So, the left side of the curve is from -infinity, so this is the PDF and the left side extreme is referred to as -infinity.

Now, if you know that distribution starts from another location you just essentially want to get to the starting point all this -infinity means it is the starting point of the distribution if a distribution can technically go to minus infinity it can, but like the uniform for instance it starts at a. So, that would be a you would replace that and x is you leave x as it is, because your cumulative density function is a function of x and this integral is what would solve it.

Obviously, if you are dealing with a discrete distribution you would instead have this summation not integral, but everything else about the formula would stay the same way if you are going from CDF to PDF you will essentially differentiate the CDF and there is not much more to that. So, now, coming to mean and variance the idea here is that the mean x bar is the kind of mean that we have discussed, so far. So, you have some data points you take them up you average them divide by, but if you won’t start thinking of a mean in terms of a distribution.

So, you given instead a distribution not actual data or you use the data and created a distribution out of it you created an f(x) out if it. Then, how do you calculate a theoretical mean of the distribution and that you do by essentially this formula for the continuous distribution and this formula for the discrete distribution and these you can call them expected values these typically gets called μ it is that Greek alphabet μ, that the core idea here is that you take each possible outcome in the discrete case multiplied by the probability that there outcome can take on and we all know that these probabilities multiply all the way to the sum of these probabilities is going to be equal to one.

But, you are multiplying p_i and x_i for each I and; that is the discrete case the equivalent of that is nothing, but f(x) is the equivalent of pi and x is x in both cases. So, this is the equivalent of the continue this is the continuous version of the same problem. The standard deviation is the same here you have the n-1, because you are dealing with x bar, but if you have a theoretical mean that is given to you can use the n and in the discrete case this formula, which is the same formula that you have used to with the exception of one by n takes place.

Now, this actually simplifies and for your convenience on the continuous version I have given you the simplified version this to me this version, which is used for the discrete is more intuitive but this formula simplifies to a formula, which looks more like summation of x^2 and so on. Separately just like on a case of the mean discrete uses a summation and continuous uses an integration we have the same distinction even here, but for your benefit on this one I have shown you the simplified formula.

So, even here originally you have started off with $\int_{-\infty}^{\infty}(x-\mu)^2 f(x)$ and so on. But, that is not what we are doing here and here actually this would not be dx this would just be probability, so it will be p_i out here. But, out here it would be d x and it would be f (x) because it represents that. So, I hope these formulas give you some idea and we will definitely be looking to see you apply some of these formulas on distributions to get answers. But, that’s it for this class next class we will talk about the distribution we have not talked about, so far which is the normal distribution.

Thank you.