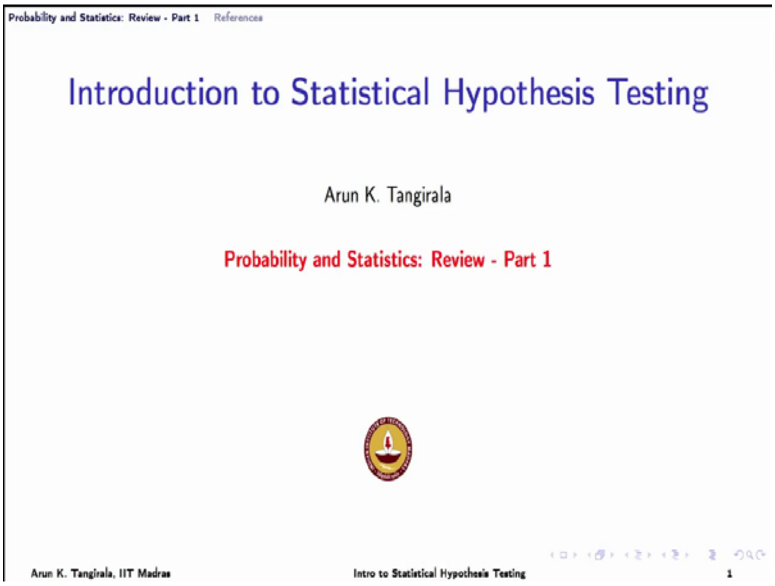
Introduction to Statistical Hypothesis Testing
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 02
Probability and Statistics: Review - Part 1

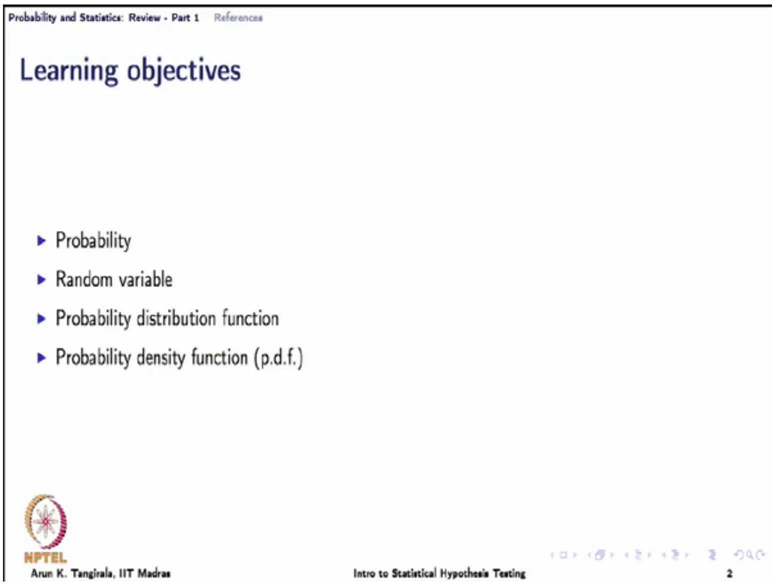
(Refer Slide Time: 00:11)



Welcome to the second lecture or second topical lecture in the course on Introduction to Statistical Hypothesis Testing. What we are going to do is, review certain basics of probability and statistics. To recap what we did in the previous lectures, we looked at why one would look at hypothesis testing and what are the different situations and what generally constitutes hypothesis test and of course, got some familiarity with some terminology in hypothesis testing.

Now, a point that I did forget to mention in the previous lecture is while we are learning the theory, what is more important, what is equally important is to learn how to take this theory into practice? So, in that regard from time to time in the course outline I am going to show you how to implement certain concepts that we learn in a software package called R. In that respect, we will load tutorial introduction to the usage of R and then subsequently as we learn the concepts, we will also learn how to implement those concepts? How to carry out hypothesis test in this software package called R? Before we get to the tutorial, maybe it is a good idea for you to install R and R studio on your desktops or laptops, so that when you are looking at the tutorial you are ready to work along with me, but anyway that is later lecture.

(Refer Slide Time: 02:01)

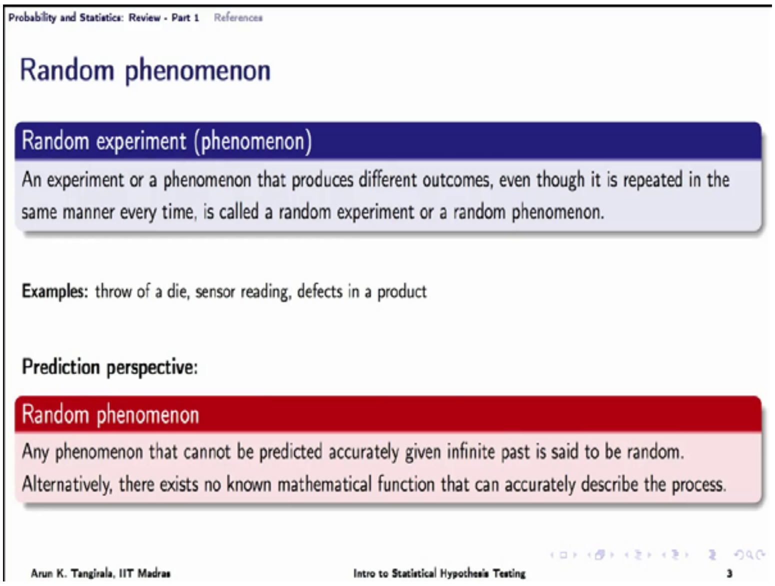


In this lecture, what we are going to look at is some basics of probability and define, what is a random variable? What is a random phenomenon? Learn and review the concept of probability distribution function and density functions. Now, lot of you must be familiar with these concepts. For those of you, who are familiar with this concepts it serves as a refresher and for others who are just about to begin the journey in random data analysis this is a quick review I should say. There are hundreds and hundreds of books and articles written on this subject of probability theory and statistics. What we are going to do is a very quick round up. I would strongly advise for those of you are new to this topic to refer again to the book on probability and statistics for engineers by Montgomery and Runger, essentially the book that we are following or the other book by Ogunnaike.

Now, before we flung into review of these concepts, we have to be clear as to why we are looking at these concepts. The primary reason is as I mentioned in the previous lecture, any data that we collect, any experiment that we perform is always going to be characterized by some uncertainties that is, factors beyond our control and these uncertainties by the very nature are unpredictable. So, scientifically or statistically the way we handle this random phenomenon or uncertainties is through the use of probability theory.

It has a simple example, suppose I want to predict whether it is going to rain tomorrow. Now, obviously I do not have sufficient knowledge of the physics, of the process to be able to exactly say or accurately say whether it is going to rain tomorrow or not. So, the natural recourse is to list all possibilities and then assign a chance to each of the possibility. So, for instance I would say that there are 2 possibilities; it may rain or it may not rain and I would say there is a 50-50 chance or 60-40 chance depending on the season of the year, the weather conditions prevailing today and so. Also taking into account all of this I assign certain chances to each of this outcomes which we call as probabilities. So, remember that probabilities are assigned based on our prior knowledge and experience and so on, but we still have to satisfy certain rules and there are certain axioms associated with probability which we are going to review today. So, the bottom line is that whenever we are dealing with random phenomenon or random phenomena we take recourse in probability theory and that is what we are going to talk about.

(Refer Slide Time: 05:03)



So, let us go through some technical definitions. All of us do understand qualitatively what is a random phenomenon? Or what a random experiment is? Technically, any experiment or phenomenon is considered random, if upon repeated operations in the sense, if I have a process and I am going to repeatedly operate it under same operating conditions I am not going to change anything. Whatever is in my control I am going to keep it fixed and then observe its outcomes? If I keep getting different outcomes each time I operate this process by keeping all the factors that can be fixed in a fixed manner, then we basically say there is something in this process that is beyond my control.

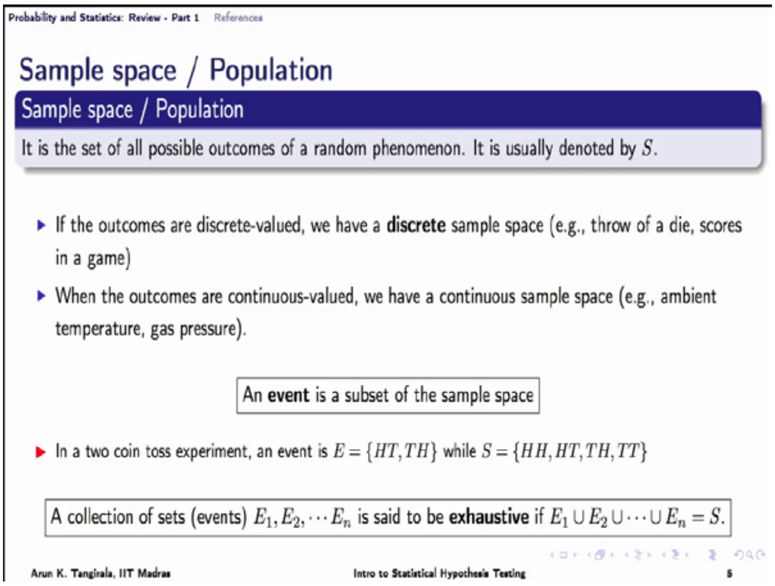
There are some uncertainties that are governing the outcomes and they are unpredictable and therefore, I call this phenomenon as random for example, if I am throwing a die or a pair of dice, it is a same pair of dice, it is a same surface, it is a same me who is throwing it. Let us say, I am able to throw it at a same angle, I can have a machine which will throw the die exactly at the same angle almost exactly at the same angle and so on. On the same surface everything else held constant, still the face value on each role is going to be different. So, it is not that I have truly held everything constant, but whatever I managed to hold fixed, I have managed and then I throw this pair of dice and the face value is different in different trails. Now, we kind of give up in terms of trying to write a first principles model or model based on physics and then say that look there is some unpredictability, there is something that is unpredictable in this and therefore, I will call this as random phenomenon and then what I do is, list all outcomes. So, essentially it is helplessness, you can say it is an ignorance, you can say on my part to be able to predict that makes a phenomenon look random. That is the most important thing and we will quickly talk about that a bit later and you can give numerous examples of this random phenomenon.

From a prediction view point, any phenomenon that cannot be predicted accurately given the knowledge of all other factors qualifies to be called as a random phenomenon or a random process. Practically speaking, there is no process out there which can be predicted accurately, we know very well. So, what is this telling us, is it telling us every process is to be treated as random? Well, yes and no depending on the extent of knowledge that I have, if I know the process fairly well. So, if I hold a chock for example, hold it in air or any object in air and leave it. We know very well at least on the earth it is going to fall right that is because I know that gravitational force will pull it down heavy, rarely it can happen if a magician or something weird set up has been there, where it will float in the air, but the probability of that is so low that we may as might as well call it as a deterministic process. I know for sure that object is going to fall; in that case it may a bit meaningless to call the phenomenon of the object falling in the air as random that is the important thing. So, we have to be practical when we deal with data.

We would have a mix of determinism and randomness, when we call something as random. We are saying predominantly it is random that is what I know very little about the process mostly I am ignorant about it and when we call something as deterministic predominantly it is deterministic that means, I know most of what is happening like the object falling in the air, I know for sure it is going to fall of course, there may be some 10 to the minus 5 percent uncertainty or 10 to the minus 10 uncertainty and so on, but predominantly it is deterministic. So, we have to be clear in our mind as to when we call a phenomenon random.

In this course, we are going to assume that; I do not know much about the physics of the process and I am going to treat all the variables as random. There is a field where you can deal with mixed effects deterministic and stochastic effects, but we will not deal with it, we will keep things simple and we will also deal with single variables and we will talk about it a bit later.

(Refer Slide Time: 09:45)



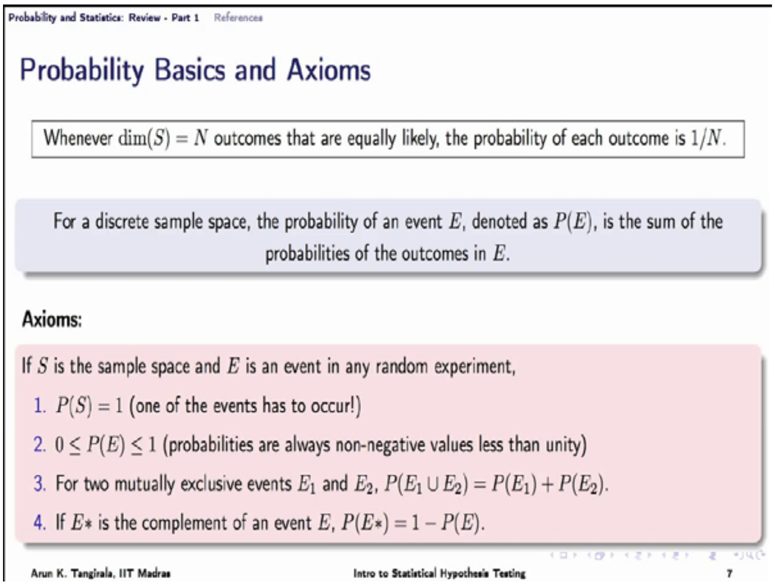
So, now that we are in the world of random phenomena. We did say that the way we deal with this is, we list all the outcomes and then assign probabilities or chances to it. So, there are some technical jargons, technical terms entities that participate in this theory and the first technical term that we encounter is sample space or population. We will interchangeably use this terms; sample space is nothing, but a set of all possible outcomes. So, in the rain fall example case, the sample space is yes-no, that is will rain or not and you can even assign a number to it and say one-zero or you can assign some other number and when we do that then we enter in the world of random variables again we will review that. So, sample space is usually denoted by S and that is the notation that we will follow.

If the outcomes are discrete value then we call this as a discrete sample space and likewise, if the outcomes are continuous valued then we call it has a continuous sample space. When do we encounter discrete, well in the pair of dice example or the rain fall example, the outcomes are discrete value. You cannot have 1.5 coming in appearing on the face value of a dice. Therefore, it is the outcome span, discrete sample space. When do we have a continuous sample space, well many, many situations a classic example is sensor reading or the ambient temperature. If I am looking at the ambient temperature, the temperature outside this room or even within this room there are numerous possibilities, it spans continue. Although, there may be interval that you believed that outcomes fall in within that interval it is continue. So, then we say the outcomes space is continuous value or it is a continuous sample space.

The next technical term that we encounter is event. An event is simply a subset of the sample space. So, for example, it will rain, it will not rain. So, the outcomes there just 1 and 0 or yes or no and the event is yes. For example, I may be interested only in the yes-part not the no-part and an event need not contain a single element. So, for example, if you take the pair of dice or even a single die, when I throw single die the outcomes are 1 to 6. I can say I am interested in all events, where the face value is less than 4. So, I have 1, 2, 3 that constitutes an event and another common example that one encounters is a coin toss experiment and so on. So, I have given one example here, suppose I am looking at a 2 coin toss experiment, the sample space is given here HH, HT, TH, TT and I may be interested in those events where at least 1 head not at least where exactly 1 head shows up and. So, the possibilities are HT and TH. We say that the collection of events is exhaustive; when all these events add up to the sample space, add up in the sense not in a mathematical addition, but the union of all of them.

Now, we will learn certain probability basics and axioms as I said, the assignment of probabilities is the individual spread, but the probabilities themselves have to satisfy certain criteria, certain axioms. So, that we can build a unified theory, otherwise each will have his or her own theory.

(Refer Slide Time: 13:26)

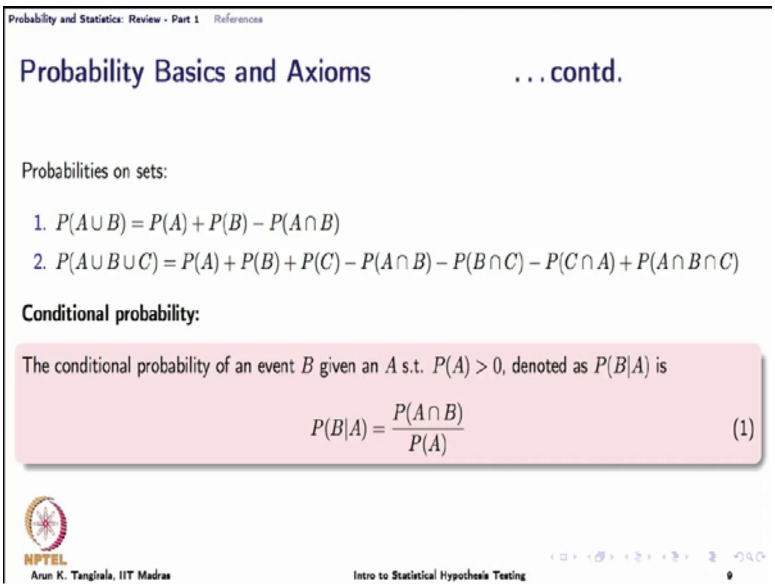


The first thing to remember and this will encounter frequently in the context of random sampling is when I have a sample space whose dimension is N and all the outcomes are equally likely, then simple way of assigning the probability is that the probability of each outcome is 1 over N. For a discrete sample space, the probability of an event is denoted as P of E. Now, for a continuous valued sample space, there slightly different treatment is necessary because we cannot count the outcomes. There we do not talk of probability of the particular event taking on the variable, taking on a certain value. As an example, if I look at the room temperature, it is not appropriate to ask, what is the probability that the temperature is exactly 25 degree Celsius? That is because the probability for a continuous valued variable, the probability at an exact point by definition is 0. There is no way we can actually ascertain whether random variable has exactly taken on a particular value.

So, for now we will talk of discrete sample space and then move on to continuous valued sample space. Now, when I look at an event E in a discrete sample space, the probability of that event E is the sum of the probabilities of the outcomes in E of the E. So, in the previous case for example, we had 2 possibilities HT and TH. If I ask, what the probability of E is, I simply add up the probabilities of these 2 outcomes. Of course, assuming that these outcomes are independent they are not influencing each other, but we will talk about that as well certain basic axioms in probability. Assume S is the sample space and E is an event in any random experiment or random phenomenon. Then we know this basic definition probabilities by definition should be a value between 0 and 1 both 0 and 1 being possible and the probability of S being 1 means that some event should occur, right. So, what is the probability that some event will occur, it is 1 and when I have 2 mutually exclusive events which means, if E 1 occurs, E 2 should not occur. That is what we mean by mutually exclusive events, then the probability of E 1 union E 2 is probability of E 1 plus probability of E 2; that means, there is no way they can occur together and very often we do encounter this compliment of an event E.

For example, in the coin toss experiment, I said let E correspond to the case, where exactly 1 head appears. So, we had HT and TH, the compliment of that is whatever is left out in the sample space and obviously, when E occurs E star does not occur. So, by definition E and E star are mutually exclusive and by applying our third axiom probability of E union E star which is nothing, but E union E star which is nothing but a sample space and which by axiom 1 is 1. So, the left hand side is 1 when E 1 is E and E 2 is E star and on the right hand side I have probability of E plus probability of E star. So, therefore, probability of E star is 1 minus probability of E. So, that is a very standard result that we use to compute probabilities depending on the questions, we do compute the probability of the complimentary situation and then answer the question.

(Refer Slide Time: 17:12)

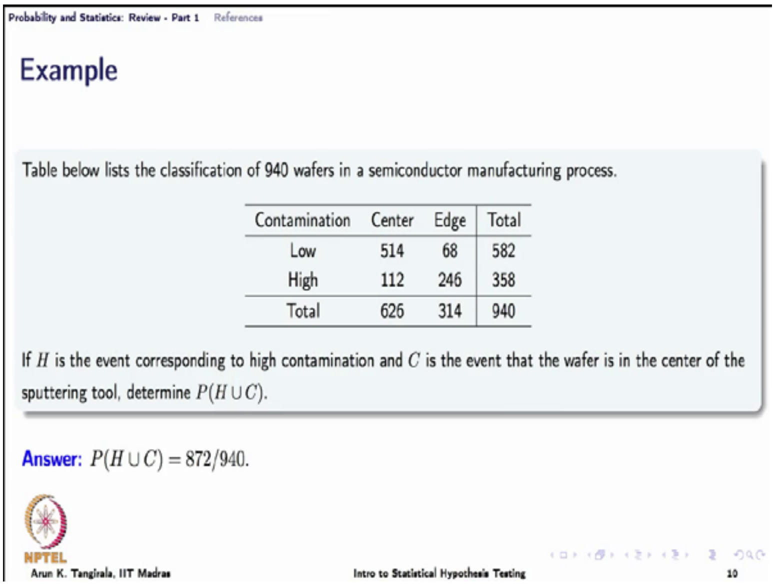


So, that is axiom 4 is quite useful in computation of probabilities and when I have 2 sets or 2 sets of events and I have 2 events A and B, the probability of A union B that is, collectively when I put all the events in A and all the events in B, then the probability of A union B is probability of A plus probability of B minus the probability of A intersection B. Qualitatively, what we are saying is we would have counted the common elements in A and B twice when we add up and therefore, we have to discount for that in our calculation and that is why we have third term, which is probability of A intersection B and now, we can extend this to 3 events and so on; because I am not going to go into the fourth event, but again you can explain the second result also using the same argument that used earlier. Whenever I have 2 events A and B, we can also talk of conditional probability. So, suppose let us say this person, Rahul, I want to know what is the probability that Rahul is carrying an umbrella, A is Rahul carries an umbrella, B is the event it rains. What is the probability that, A is the event that it rains and B is the event that Rahul carries an umbrella? I can ask this question has to what is the probability that Rahul carries an umbrella in general. There is a certain probability associated with it and we can come up with this answer depending on Rahul's habits.

Then there is another question that I can ask, what is the probability that Rahul will carry an umbrella even it is raining? That is what we call as conditional probability and that probability hopefully will be different from the unconditional probability. The unconditional probability is probability of B. The conditional probability is probability of B given A and the answer or the computation of conditional probability comes from base formula or base theorems, which says the probability of B given A is the probability of A and B occurring together, divided by the probability of a occurring in itself that is called as an unconditional probability. So, of course, this definition is valid when probability of A is greater than 0.

So, if I know the probability of rain that is whether it will rain, what is the probability and I know the probability that Rahul of A intersection B, that it rains and Rahul carries an umbrella. How do I compute these probabilities? In practice I have to only rely on observations. It is very hard to at least calculate what is the probability that Rahul carries an umbrella using a mathematical model? I have to know the history of Rahul's habits and may be Rahul himself can tell me what is the probability that he will carry an umbrella and also the probability that he carried an umbrella, whenever it rained that is A intersection B. Once, I know that, then I will be able to say, what is the probability that Rahul carries an umbrella given it is raining? Now, from this conditional probability stems a notion of independence and we will come to that, but just to break the monotony of theory here is a simple example.

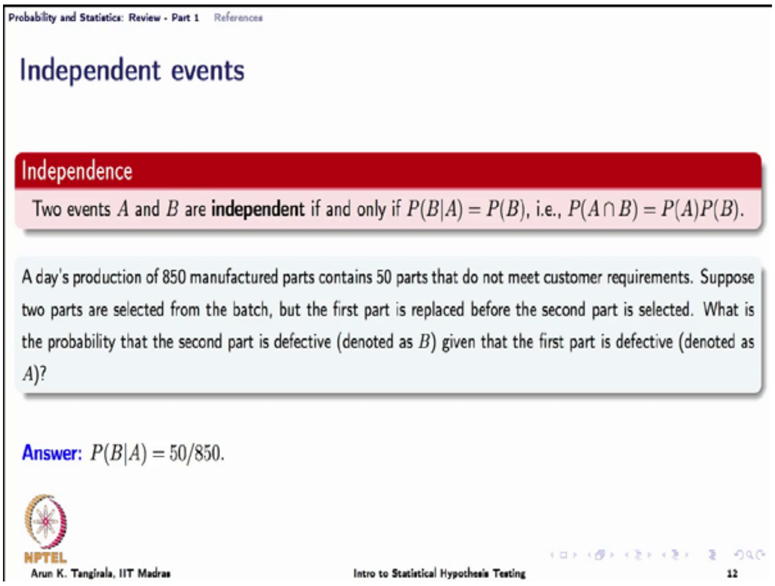
(Refer Slide Time: 20:40)



Where I am asking you to compute the probability of the union of two events, what is the process here. We are looking at some semiconductor manufacturing process. Please do not think that I am really obsessed or I like semiconductor manufacturing process a lot, somehow we keep running into this example, but it is a very powerful industry today as you know. Let us say that, there is a semi conductor manufacturing processes and I picked 940 wafers and I looked at how many where contaminated? This is an example form Montgomery's book and let us say, we classified this contamination levels as a low and high and this center and edge has got to do with the operational details, as to know where the particular machine was placed to make this wafer at the end or center or at the end and we have done a classification of this 940 wafers and the table essentially tells you this classification. That is when the contamination is low and among those when the wafer that the machine was placed at the center location, 514 occurrences where at the center location and the 68 occurrences where at the edge location and likewise for high. So, all these details are given.

Now, if H is the even corresponding to high contamination and C is the event that the wafer is in the center of the tool. Then what we like to know is the probability of H union C, this is different from H intersection C, H intersection is H and C occurring together. H union C is collection of all events in H and all events in C and what one could do is now go back to this probability expression that we have given in one probability of A union B and apply that here. So, what you need to do is compute probability of H which is 358 by 940 and probability of C which is 626 by 940 and probability of H intersection C that is given here as 112 and well that is the number of occurrences. So, the probability is 112 by 940 and then you apply this from formula probability of H plus probability of C minus probability of H intersection C, you will get the answer. So, it is a simple exercise. So, now, we will get back to the theory.

(Refer Slide Time: 23:35)



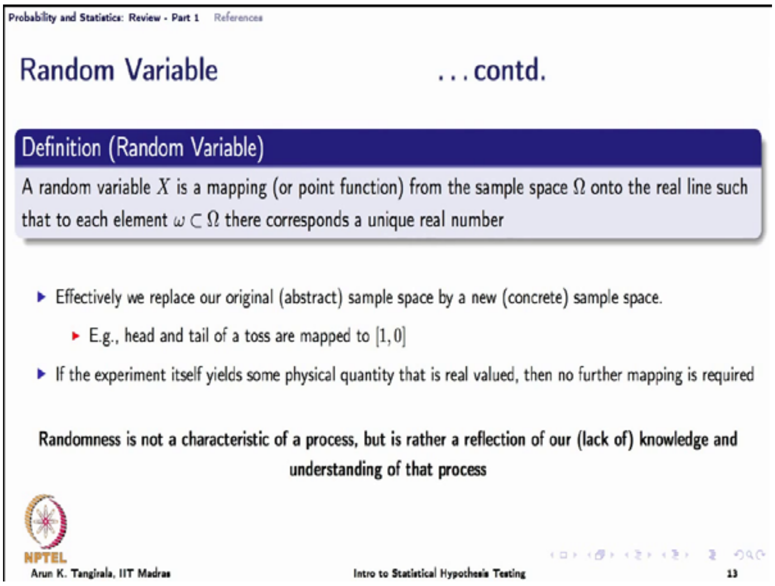
We said from the conditional probability stems in notion of independence why because we said that, when there are two events we can look at the probability or then of B given. Suppose Rahul is carrying an umbrella has got nothing to do, whether it rains or not simply Rahul loves to carry the umbrella. Then does not matter whether it rains or not, the probability of Rahul carrying of umbrella will not change at all. Then we say that those two events are independent. So, 2 events A and B are independent, if and only if the conditional probability is same as unconditional probability. The other way of saying that, is that probability of A intersection B is simply the product of the respective probabilities and of course, where does this second result come from, it comes from the conditional probability result itself because probability of B given A is probability of B and we know that probability of B given A from equation one is probability of A intersection B over probability of A. Simply plug into this first condition and you will get the second condition.

So, as an example again to drive home the notion of independence, this example is again from Montgomery's book. Let us say there is a process industry which is manufacturing certain parts and in a day it produces about 850 parts and let us say we know that there are 50 parts of this 850 that do not meet the customer's requirements, whatever these requirements are. Now, suppose I select two parts, two specimens, let us say at random from the batch, but in such a way that the first part is replaced before I take the second part and I want to ask a question what is the probability that the second part is defective? This we call as sampling with replacement. So, the question is as follows. Suppose, the first part is defective then is it going to alter the outcome in the second part, clearly it would not because I am replacing it.

So, the population remains the same I have not alter the population nor have I alter the probability distribution therefore, the probability that the second part is defective in this kind of a sampling is going to be invariant to the outcome of the first part. Then we say that, if A is the probability that the first part is defective and B is the probability that the second part is defective, A and B are independent, but if I were to, not replace then they are dependent on each other. Then we say A because what happens is if A is defective and I have not put it back then I only have 49 defective parts. When it comes to the second sampling then we will have to calculate the conditional probability in a different manner. So, here the probability is B is same as probability of B given A, which is 50 over 850. So, hopefully drives on the notion of independence. So, until now we have talked about discrete valued outcomes or discrete sample spaces, but as I mentioned earlier when you have continuous valued sample spaces, it is hard to talk of probability of this variable taking on a specific value in a continuous valued sample space.

What is appropriate to ask is, what is the probability that the temperature for example, in this room is a within a certain interval. However, small that interval is, the length of the interval cannot be 0. Now, very often we encounter situations where the outcomes are categorical that is one story is that of the continuous random variables, where we do not necessarily talk of probabilities of x equals something, but x falling within a certain interval we can answer that now the other situation that we encounter quite often is. So, until now we have looked at situations where the outcomes are discrete valued and so on that and particularly where we talked about numbers. We converted even the categorical outcomes like rainfall, yes or no into numbers and spoke about it. Now, we generalize this idea and also move to the field of continuous valued sample spaces in that context. So, as to deal with the universal theory it is convenient to introduce the notion of random variable and that is what the main purpose here is.

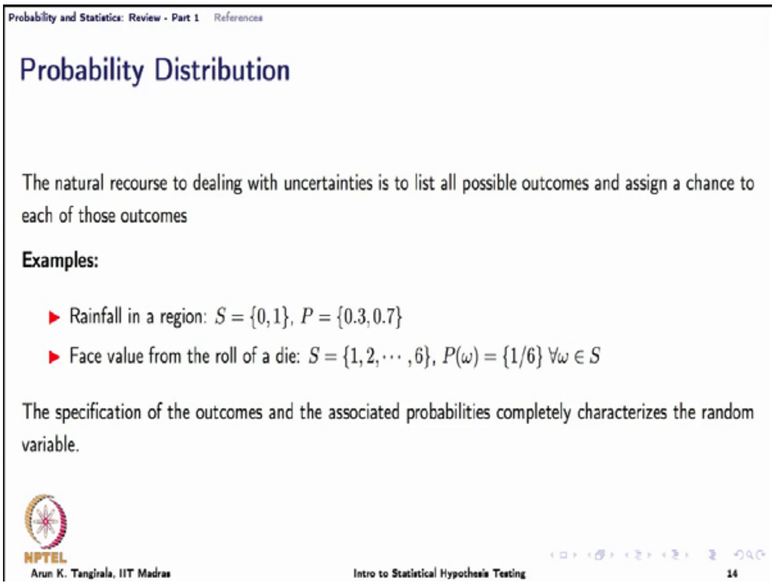
(Refer Slide Time: 28:28)



So, random variable is nothing, but mapping of the outcomes which could be categorical generally to a number line, a classic example rainfall being, yes or no is mapped to 1 or 0, it could be even 1 or minus 1, the mapping choice is yours. Of course, outcomes are already in numerical in nature then there is no need to do this mapping. Then directly the outcome itself is the random variable the sample space.

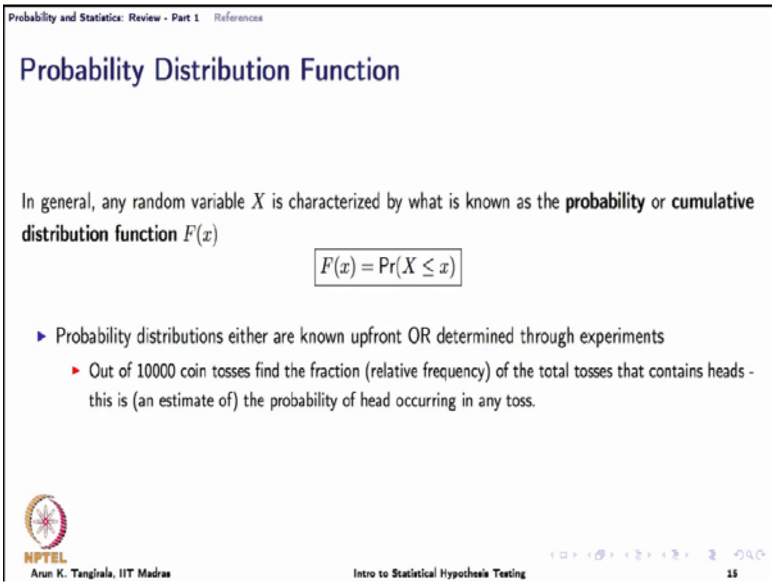
We say that for each element in the sample space, we assign a numerical value and once you are done that now you have a random variable. So, essentially what we are doing is we are replacing our original sample space, which is probably having categorical variables by a new sample space, and of course, this mapping can have some impact on the analysis therefore, the mapping has to be chosen carefully. Again just to reiterate, randomness is not a characteristic of a process, but it is rather of reflection of our ignorance or of lack of understanding time to time. It is important to tell ourselves this fact that the randomness is a matter of perspective and knowledge of the user. So, now, that we have introduced the notion of a random variable the next in line is the concept of a probability distribution.

(Refer Slide Time: 29:55)



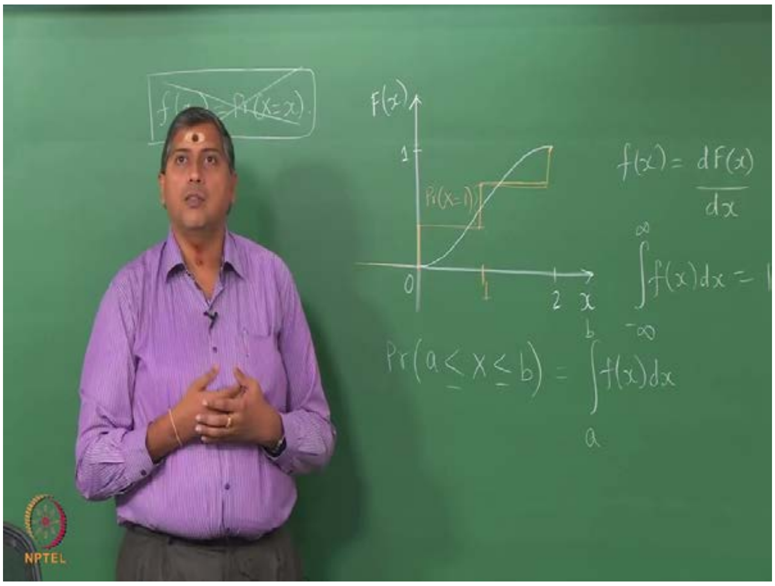
Earlier, we said the idea in probability theory is to list all the outcomes and then assign chances to those outcomes, but we cannot do that for a continuous valued random variable. For example, if I cannot ask a question what is the probability that the temperature in this room right now is exactly 25 degrees Celsius, rather we are qualified to ask, what is the probability that the temperature is between 25 and 25.01? So, the interval can be very small, but the length of the interval cannot be 0. In such cases, it is convenient to introduce a function call probability distribution which applies to both discrete and continuous valued sample spaces. By definition, the probability distribution function is defined as a probability that a random variable x will take on values less than pre specified value small x.

(Refer Slide Time: 30:48)



Now, just a quick note on a convention, by convention we use the upper case for the random variable to denote any random variable and the lower case for the value that it takes on. And also, the probability distribution function known as a cumulative distribution function because you are looking at a probability of x taking on any value less than or equal to x, which means you are looking at probability of x taking on any value from the left extreme to the pre specified outcome that we denote as big F or the upper case F of this variable x, that is the random variable x because it is a mathematical function like any other mathematical function, but additionally it has the features of the probability distribution function. And, a probability distribution function has to satisfy certain mathematical properties for it to be called as a probability distribution function. For example, at the left extreme the probability F of x should be 0 because there is nothing happening beyond the left extreme value and at the right extreme the probability distribution function has to be unity because is again nothing happening beyond the right extreme of the random, of the value that the random variable takes and there is also another important requirement that the probability distribution function has to satisfy that it should be monotonically non decreasing between the left and right extremes.

(Refer Slide Time: 32:33)



So, as a quick example, let say the x-axis is the value the random variable takes and the y-axis is the probability distribution. Assume that, 0 is the left extreme possibility and let us say here some value maximum here, let us say 2 is the value that this random variable can take, the important requirement is let say we call this here as 1.

The points that I made earlier, means that f of x that x equals 0 should be 0 and F of x that x equals 2 should be 1, but between these 2 points. It cannot vary as it likes because probabilities only add up they do not subtract and therefore, you F of x has to necessarily be monotonically non decreasing. This is if x is a continuous valued random variable, if it is a discrete random variable then we have the distribution not being continuous function. In that case, what we would have is this kind of a situation where suppose, x takes on only values 0, 1 and 2. Then the probability that x takes on any value less than 0 is 0, at x equal to 0 exactly, here it remains 0 here and at x equals 0 there you can talk of the probability that x can take on a value of 0. In that case, the probability distribution looks like step like function or a staircase like signal. Now, because x is discrete valued it cannot take on any values between 0 and 1, the next value it can take is 1 and at this point the probability that we add up, the probability of x taking on values of 0 and 1 because it is a cumulative distribution function.

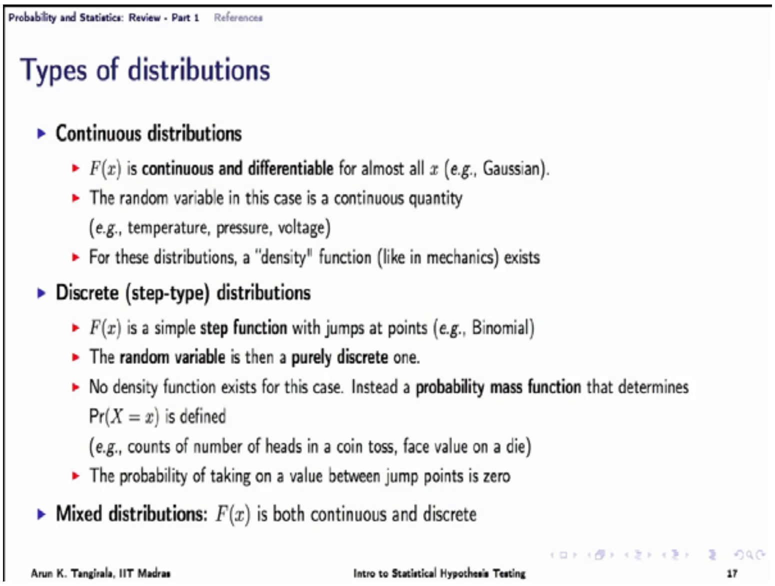
So, at this point F of x may be here because it is an addition of probabilities here and we say it remains constant because the probability that x can take on any value is 0 here. So, the F of x remains flat and jumps to this new value. This is the probability that x equals 1 and then likewise by extending the same argument. So, it has to reach 1 because F of x at 1 is the probability that x can take on any value less than or equal to 2, in this case right extreme value. So, this is the difference between a continuous distribution and discrete distribution, that is discrete valued random sample space cased for a continuous valued random variable F of x is continuous and for a discrete valued random variable F of x is a staircase like signal and immediately, now we can think of defining a density function for a continuous valued random variable because F of x is continuous.

I can think of a density function analogous to the way we define densities in mechanics, we talk of mass distribution and so on. Here, this density which is mass per unit volume, we can talk of such quantities only when we are dealing with a continuum, when the mass is only located at discrete points in this space then there is no notion of density. So, here we can define density. For example, as the derivative of the distribution, this is one definition and by this definition also it follows that the area under the density curve between two extremes is 1. So, when I write here minus infinity to infinity, do not tend to think that always we are dealing with random variables that take on negative values and up to minus infinity and likewise, here for the upper limit the way this lower and upper limits should be interpreted is these are the this is the left extreme and this is the right extreme possible value and the area under that should be 1. So, for any function to qualify a probability density function the first requirement if this is not then you have to normalize or scale F of x accordingly multiplied to the suitable factor. So, that the area always remains 1, this is to say what is the probability that any event will occur and by this definition we can say that using the density function I can calculate the probability that x takes on any value within the interval A and B inclusive A of and A and B as the area under the density curve evaluated, this is evaluated in that interval A to B. So, this is the standard way of using the density function.

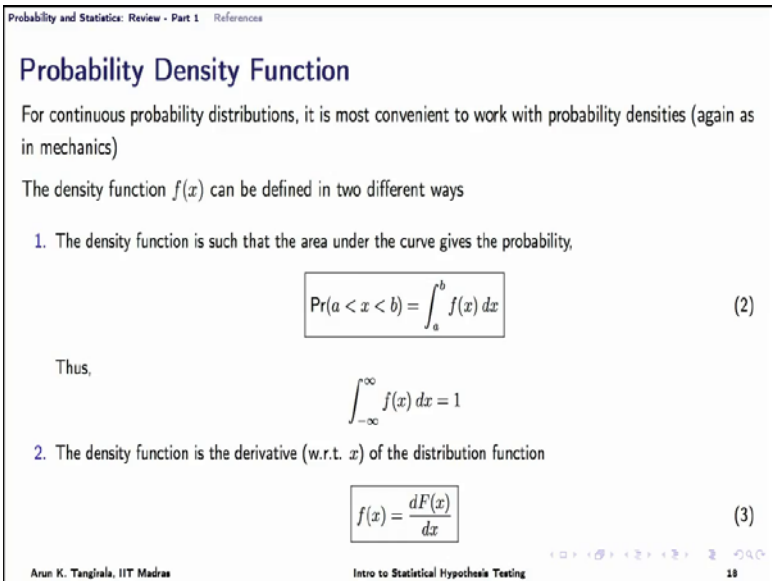
A common misinterpretation of the density function is that its value at a specific value of x is the probability that x equals x this is very common to actually write, misinterpret this way. This is wrong for a continuous value random variable; it is wrong for a discrete valued random variable. We will define what is known as a probability mass function analogous to density by density, but it is not a density by it is called a mass function and in that case this holds you can then say that exactly the value of that mass function at that value is a probability that x will take on that value. So, this difference between distribution functions and density functions.

Density functions only exist for continuous valued random variables and distribution functions exist for both discrete valued and continuous valued functions. Now, what remains to be reviewed is, what are the different density functions or distribution functions that one encounters in practice and we will review that and close the lecture we have just spoken of these distribution functions.

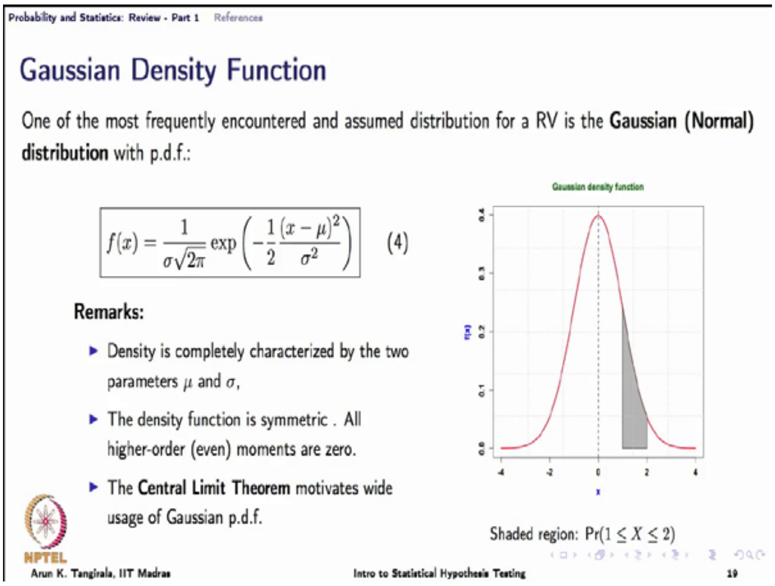
(Refer Slide Time: 38:39)



(Refer Slide Time: 38:42)



(Refer Slide Time: 38:45)

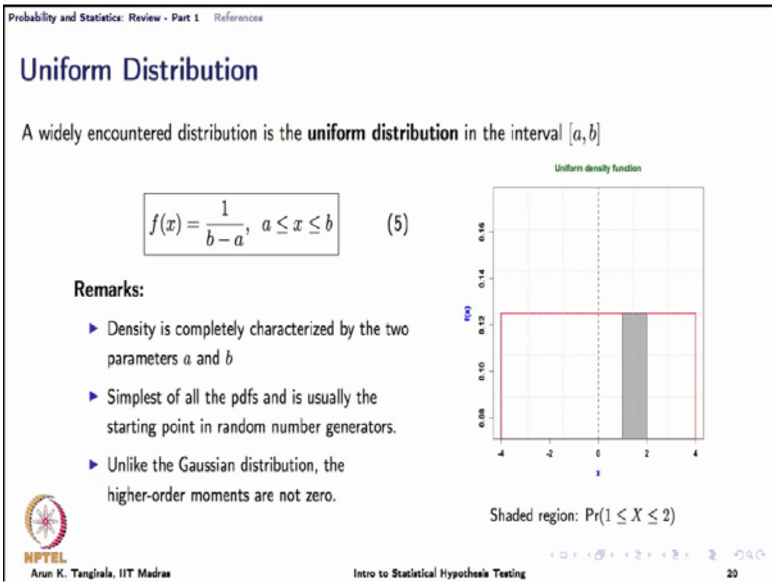


You can also have a mixed distribution therefore, I will move on and we have also just gone through this. So, now, what remains to be seen, what are the different probability distribution functions that come about? Remember these distribution function are not fictitious, it is not an imagination they are all based on the observation of the natural or the man made phenomenon and depending on the nature of the phenomenon, you will get particular distribution or density function.

In fact, if you look through, read through the first few chapters of the book by Ogunnaike, there is a way that you can derive the probability density functions from first principles. If you know something about the physics of the process, it is possible to derive the density functions, but very often this probability density functions or distribution functions as a case may be, are derived from experimental data from observations and what many people have done in the past is they have observed this particular class of phenomenon. For example, averages whenever you are looking at averages of some variables then these averages are known to follow. What is known as a Gaussian distribution function or a bell shape function? A function that we would have encountered for the maximum number of times in our life and this function fortunately has a mathematical expression, it is called a Gaussian or a normal distribution or a normal density function.

When we say p.d.f, here in our course stands for probability density function and typically, that is the case in the literature and the expression is given in the equation four, where there are 2 parameters that have to be specified. To know the density function, this 2 parameters has we very well know are called the mean and standard deviation mu and sigma and there are many nice properties of this density function. We are not going to review it, but the immediate thing that we point out is that the density function is symmetric. So, in the figure I have a shaded area for you that is the area corresponding to the probability that x will take on any values between 1 to 2 and what makes a Gaussian density function very popular. As I said, averages in average of bunch of random variables tend to follow a Gaussian distribution that is one of the primary reasons why the density function is very popular. We will keep using this density function quite frequently in the course therefore, you may want go through some additional reading and then comes another most important distribution.

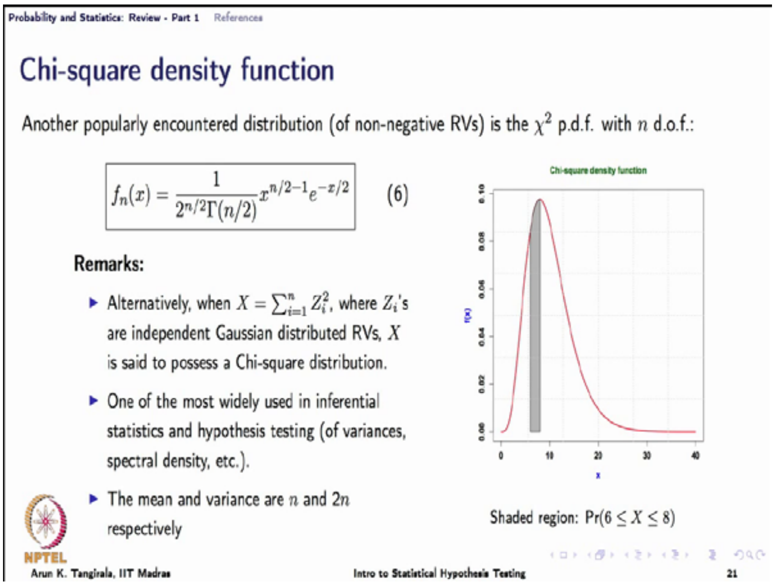
(Refer Slide Time: 41:18)



Earlier, we said in random sampling when I sample n values out of sample or let us say I take P outcomes randomly, I sample them out of a sample space of n outcomes then the probability of each of that outcome would be 1 over P. What is the point there, the point there is that all those P outcomes were equally likely or even if I take a sample space of dimension n, earlier we said if all outcomes are equally likely then the probability of each outcome is 1 over m then we say that the random variable follows uniform distribution then this applies to both discrete valued random variables as well as continuous valued random variables.

So, what I am showing you here is for the continuous valued case, where the random variable takes on any value between the interval A through B and by definition f of x has to be 1 over B minus A, so as to satisfy the legitimate requirement of a density function, which the area should be 1 and the density is completely characterized by these 2 extreme values A and B and of course, one can calculate the mean and variance of this distribution. We will talk about mean and variance later, this is the simplest of all the p.d.f's. Very often you assume either Gaussian or uniform distribution, if you do not know anything and you think all outcomes are equally likely then what you are appealing to was uniform distribution. Otherwise, if you know that you are actually looking at averages and so on, then Gaussian distribution is a natural choice. Once again the shaded area represents the probability of x taking on values in a certain interval.

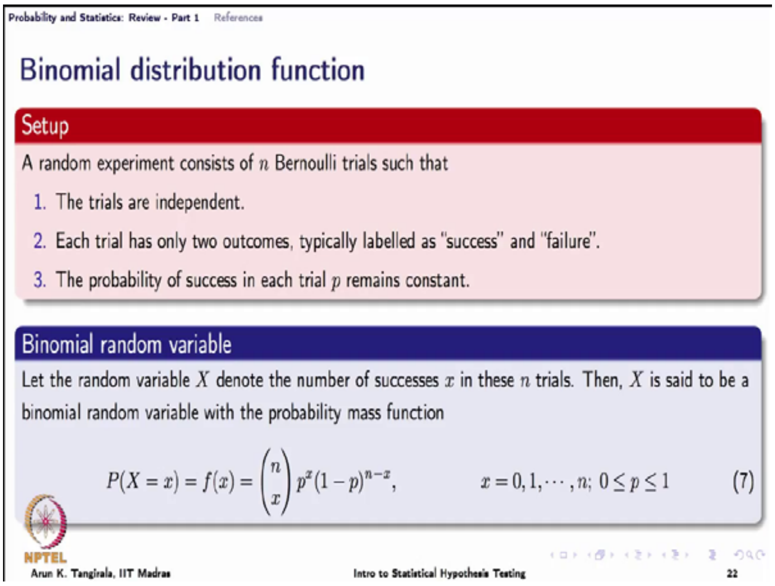
(Refer Slide Time: 43:07)



And, the next comes the chi-square density function that we will encounter. The chi-square density function is some kind of a density function corresponding to variables that are constructed from squared random variable. When I sum up n squared independent Gaussian distributed random variables, the resulting random variable that is, let us say there are these random variable sets, which are Gaussian distributed and they are also independent and I construct a new variable x which is sigma z i square i running from 1 to n. This new variable x has a chi-square distribution with n degrees of freedom. Where do we encounter the situation? When we calculate sample variances or when we calculate spectral densities then we run into these kinds of a situation for our course.

The reason that is important is we will look at sample variances and we know how the sample variances are calculated 1 over n minus 1 sigma z i minus z bar square and so on, but we will of course, talk about that later. Just to tell you that this is the reason why we are going to look at the chi-square density function and the p d f has an analytical expression. In fact, you can see there is a gamma function appearing in the chi-square. The chi-square density function is the special case of the gamma density function and you will be able to see that in the literature as well and in this case, the mean and variance are the n and 2 n, respectively. The main characteristics, the parameter that characterizes a chi-square density function is a degrees of freedom n and when you look up tables or when even you look up when you use software packages, you have to specify this degrees of freedom and you can show that as a degrees of freedom become very large, the chi-square density function will tend to Gaussian density function and you can see from the shape. The last in discussion is a binomial distribution function.

(Refer Slide Time: 45:13)



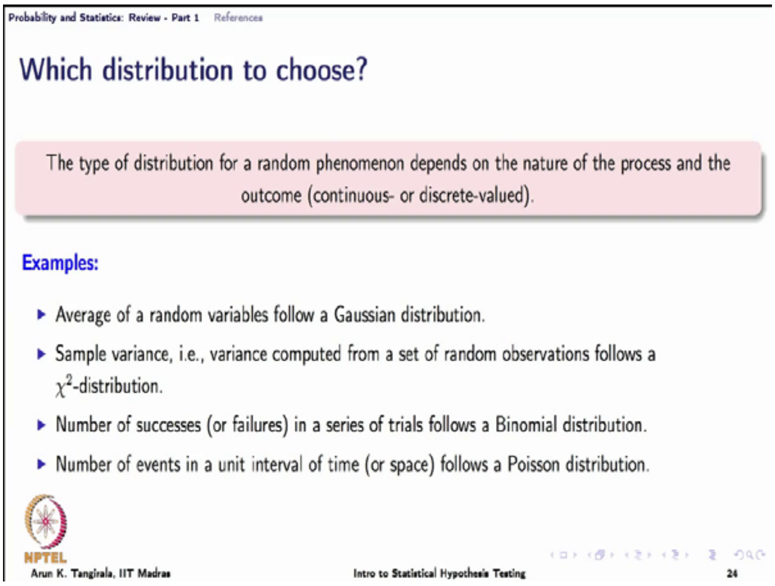
Here, we are looking at discrete valued outcomes. The earlier three density functions we were considering continuous valued random variables, but now we are looking at discrete valued random variables specifically which have only two outcomes like the rainfall case or the outcome of a game were only victory or loss can be the possible outcomes or a part being defective or not defective and so on. So, we call these outcomes categorically as success or failure and the frame work for defining a binomial random variable is that of a Bernoulli trail, where the trails are independent that is each time I perform the experiment. It is the outcome of that experiment has got nothing to do with the outcome of the previous or the future experiments and the probability of success in each trail remains constant. Under these conditions, when look at a random variable that denotes number of successes x in n trails. So, for example, toss a coin, there are going to be two possibilities, head or tail and we call head as success, let us say tail as failure and I toss this coin n times and I am looking at the number of trails, number of heads that show up in this n trails. Then that random variable follows what is known as a binomial distribution and it is called a binomial random variable and of course, now you use, combinatorial mathematics to compute these probabilities.

The density function, here we do not call this as a density function. We call this as a mass function, probability mass function, where now we can ask what is the probability that x takes on a specific value of successes, specific number of successes. That of course, depends on the number of different ways in which x can occur and therefore, you have n c x that is the notation that is meant by this parenthesis n c x and then P is as we know the probability that a success occurs, this is a very familiar distribution.

We will use this to calculate proportions of example defective items in a lot and so on and just to give you a feel of how binomial mass functions look like I have simulated 2 different or calculated the mass functions for 2 different situations. As I have shown here, the difference is in the success rates and the number of trails in this 2 figures, as in the left hand side of the success rate is 0.5, the success in each trail, the probability of success in each trail is 0 point 5 and on the right hand side figure, the success probability is 0 point 2, the number of trails is indicated.

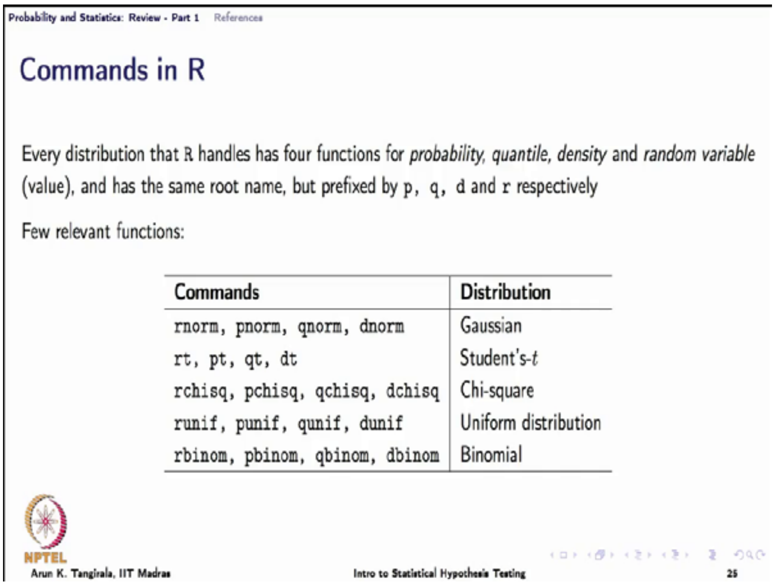
So, as you can see as the probability of success decrease then the shape becomes more and more asymmetric.

(Refer Slide Time: 48:42)



In the end, we want to ask a question, which distribution to choose? I have already said that it depends on the nature of the event, the type of distribution for a random phenomenon, exclusively depends on the nature of the event. For example, if you look at average of random variables, it follows a Gaussian distribution or if you are looking at number of events in a unit interval of time. Let us say, I am looking at number of accidents occurring in an industry. If 2 accidents do not occur simultaneously, only 1 occur accident occurs in a unit interval of time and the rate at which this accidents occur on the average does not change then we say that the number of events in a unit interval of time follows what is known as a Poisson distribution, which I have not talked about it. It is a discrete value random variable. So, read through the literature and learn more about this, but whatever review I have given should be a sufficient for the course and at a later time I will show you how to for example, simulate density functions in R.

(Refer Slide Time: 49:40)



Once, we go through the R tutorial, as a part of the tutorial I will show you how to, for example, simulate a Gaussian density function and just for your reference for those of your familiar with R, I have given you the commands associated with the concepts that we have learnt in this lecture to be used in R. You can look up the help on each of this, if you want to learn more or wait for the tutorial. So, we will meet again in the next lecture.

(Refer Slide Time: 50:14)

Where we will talk about statistics and which looks at the practical aspects of a random variables. Here we have talked of probability distribution functions and the density functions, but the more practical way of looking at random variables is to look at statistics. What are the statistics that we are going to look at is what we will discuss in the next lecture. Alright, and then see you in the next lecture.