NPTEL

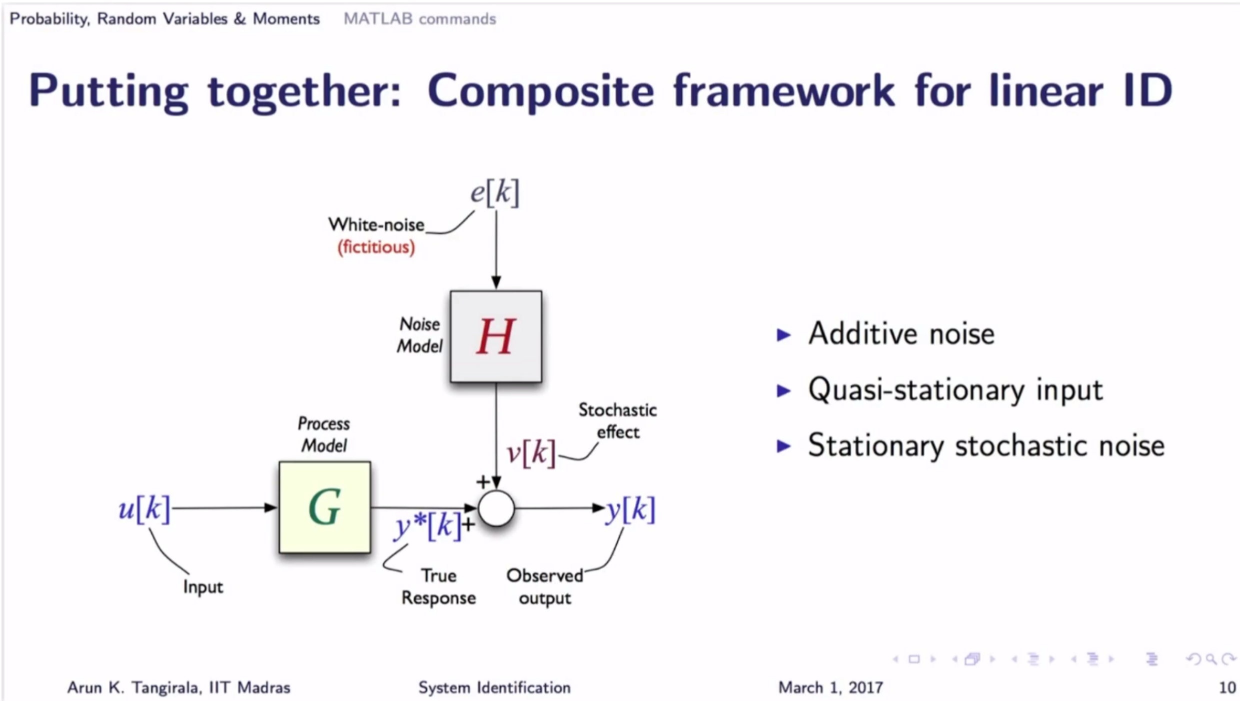
NPTEL ONLINE COURSE

CH5230: System Identification

Probability, Random variables and moments: Review 3

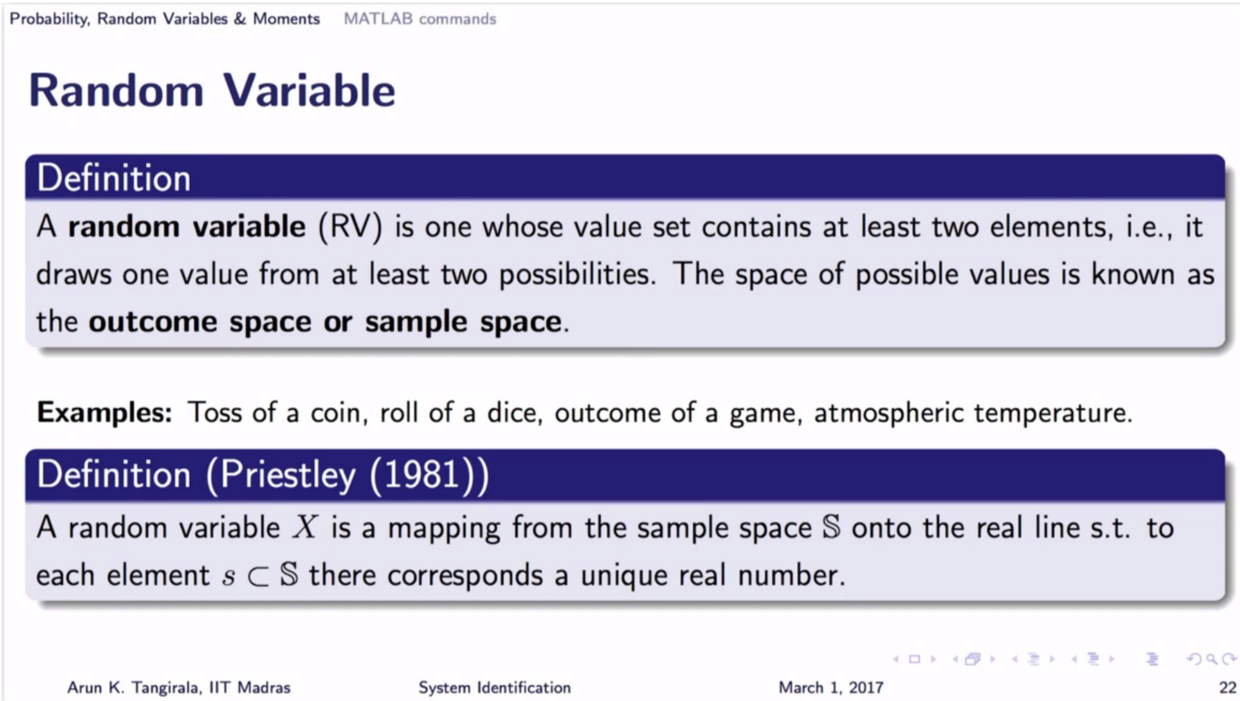
Arun K. Tangirala

Department of Chemical Engineering
IIT Madras



Arun: Very good morning. Let's get started. So what we learned yesterday is how random signals are challenging to model and what is the general framework for system identification if you recall. We are actually focusing on this kind of setup and where now we are learning, we are going to just review the concepts that will help us model VK and we have already said that this VK is going to be modeled as a white noise passing through a linear filter. And we have also stated conditions namely stationarity and spectral density. So it's time to review those concepts that will allow us to understand these conditions, characterize the signal which is the random signal VK.

Now yesterday I did mentioned to you that random signal is actually an ordered collection of random variables. So this is one realization and at each point in time we think of this as a random variable. So first we understand how to describe the signal at any instance in time, and then quickly review those concepts that will allow us to put them together. In the end what we are going to do is we are going to work with tools that will allows us to [00:01:59] the signal in time but you should always realize remember that this random signal whatever descriptions that we give whether we describe it at any point in time or we look at the entire signal itself we are actually looking at multiple realization. That is the collection of realizations. So the difference between a deterministic signal and random signal is this additional dimension called outcome space which will become clear once we review the theory of random variables. So let's now move on and ask how the random signal is defined at any instance in time. And as I have always been saying when we think of random variables we let go the notion of time. We don't worry about time.

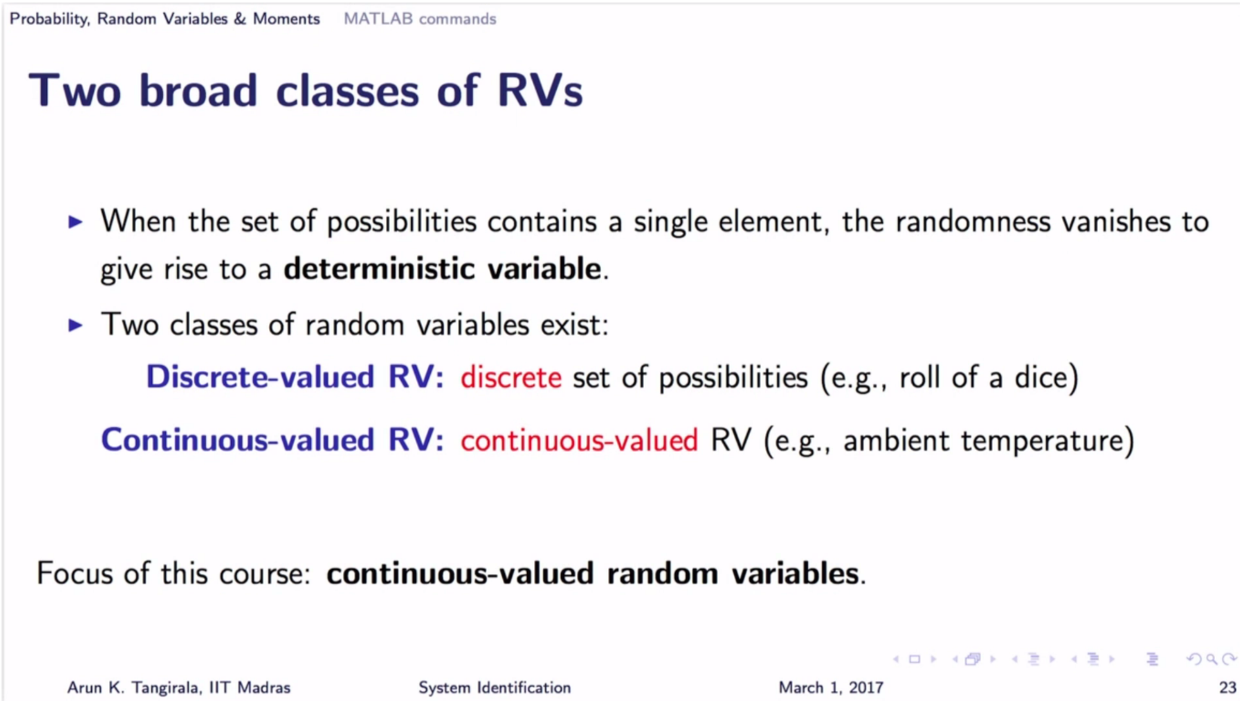


Now there are different ways of defining a random variable. A simple way of remembering it is that it's a variable that has multiple possibilities in itself. Given all the conditions that generate the random variable there are multiple possibilities unlike a deterministic variable. Like I said suppose there is a function sine of X or sine of whatever it is and I give a specific value of X is only one possible value. Whereas with the random variable there are multiple possibilities again recall our discussion yesterday this is just our imagination that there are multiple possibilities from which one of the possibilities will manifest as outcome. That you should keep telling yourself until it is reinforced well in your mind.

Now the set of all possible values is called a sample space. And there are numerous examples that we have been hearing, we have been coming across since our high school days. There is another definition of random variable which is more formal. It basically says that a random variable is essentially a mapping from the possibility set to the real number space. Why was it defined this way because there are many outcomes that are not necessarily numerical. You can have a categorical outcome.

Now in order to have a unified theory of probability our variable, our mapping had to be introduced and that is the random variable. So for example in a game you may have victory, loss and draw. But we want to convert them to numbers. And this notion of random variable allows us to map the categorical outcomes to numerical values and the choice is yours.

Now as far as the working definition is concerned a random variable is that variable which has many possibilities and this possibilities are numeric. Out of which it will take on only one value. And how does it take on and so on what is characterized by the probability distribution function. But before we talk about that it's important to note that there are two classes of random variables; discrete valued random variables and continuous value random variables.



We will not worry about the very rigorous definitions of these discrete value and continuous value random variables but it suffices to know that random variables that take on only discrete values that is not all numbers on a continuum they are call discrete valued random variables. You can think of for example population, or when you roll a dice what number shows us and so on. These are all discrete valued random variables and the other thing that I keep telling all my students is when you think of something as a random variable you need to have a right justification in your mind. Why are you treating that variable as a random variable and the justification that is generally very convincing is that I cannot predict it accurately. So if I roll a dice the value that shows up on the face I can treat it as a random variable because there is no way I don't have any formula to predict it unless I am Sankuni but otherwise I cannot. Unless I am doing some match fixing for example but we will rule out those situations. In a general scenario there is no way I can predict accurately what shows up when I roll a dice. Therefore it justify to think of it as a random variable.

So each time you think of something as a random variable it is important for you to have this justification at the back of your mind. That's a very good habit and it's an important habit to get into because very often people just assume signals to be stochastic, variables to be random and so on without necessarily thinking why they should be treated as random.

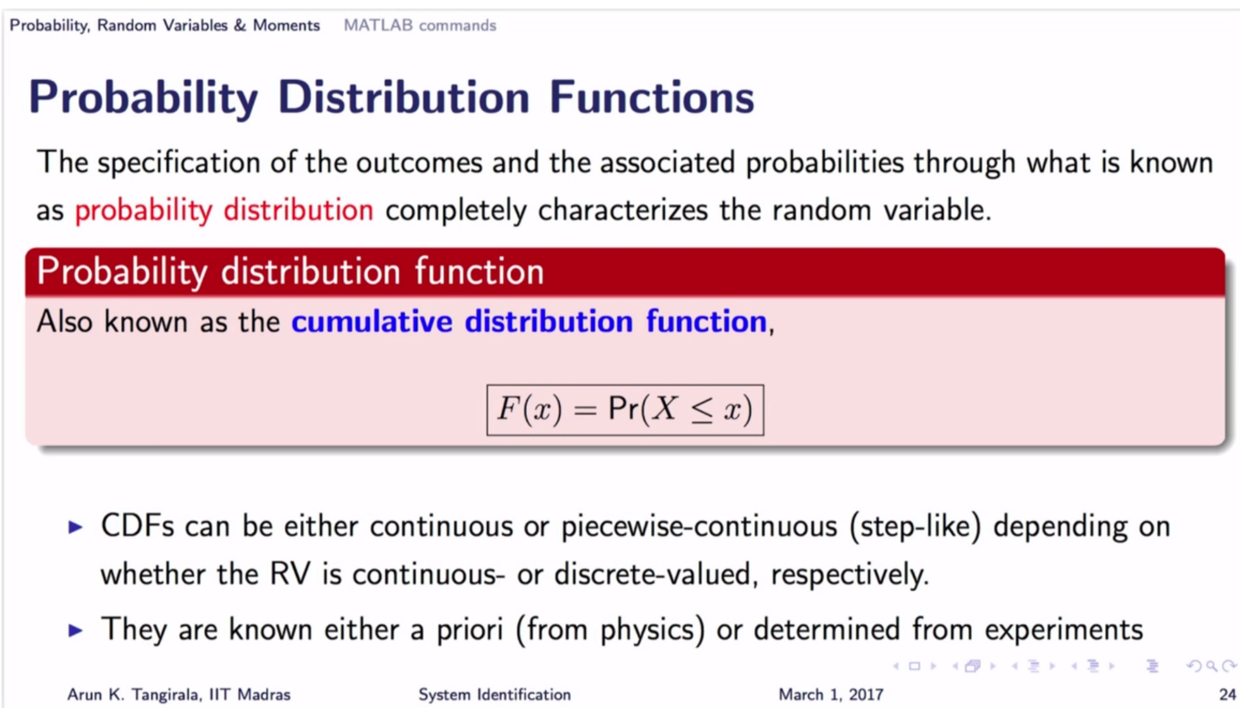
Continuous valued random variables by their very name indicates that this random variable takes on continuous values like the temperature, reading from sensor and so on. In fact, if you think of it reading from a sensor strictly cannot be treated as continuous valued random variables. Why? Suppose I am looking at temperature sensor. On the face of it appears that the random that I can think of this temperature reading as a continuous valued random variable. Again you have to ask why am I treating the reading as a random variable? What is the reason? Better way of stating?

Student: [00:08:12]

Arun: Why?

Student: [00:08:19]

Arun: So you have to be very clear. There is going to be error in the reading and I am not going to be able to predict the error. In other words, when you thinking of some variable as a random variable not only you need to have the justification you also have to – have some idea of the source of uncertain. So the source of uncertainty in temperature reading is a sensor error. If not anything you have known all disturbances and so on. So there is this source of error, the uncertainty that you have to keep in mind. Now just now I said on the face of it, it looks like a continuous valued random variable but if you look at it closely it isn't and that's because for all practical purposes these readings are quantized. So you can argue that no it has to be discrete valued random variable. But we assume that the quantization is very fine. And for all working purposes we will assume it to be continuous value. But strictly speaking all your readings are actually discrete valued random variables due to quantization. Keep that in mind. In this course we will primarily deal only with continuous valued random variables but a lot of theory that you learn more or less applies with some differences. So one of the things that characterizes a random variable completely is probability distribution function. We know that. Why does this come into picture because we have said already random variable is that variable which has multiple possibilities out of which we will assume one value.



Now which one assume – which one does it assume, what is – does it have equal chances of assuming any value or are there certain values that are more likely than the other ones and so on. That is what we mean by chances. We usually assign chances. And those chances we call as probability and how these chances are distributed or probabilities are distributed across outcomes is what is characterized by probability distribution function. And to formulate things we introduce this cumulative distribution function also known as a probability distribution function which is nothing but the probability that this random variable you know the notation the upper case is used to denote the random variable and the lower case is used to denote the value. So the cumulative distribution function also known as a probability distribution function is a probability that X will take on some value any value from the left extreme. We are looking at univariate random variables, from the left extreme to the point of interest, to the value of interest. So it's a cumulative one. And for a continuous valued random variable F of x would look like this, generic F of x. Assume that I am going to clip the left extreme here. And let's say the right extreme is here. It may not be symmetric, please don't get the impression that always the outcomes are symmetric about the origin. typically you will see this kind of a – of course the value extreme value at the extreme right extreme the value of F of x is one and at the left extreme the value is zero. This is X min and this is X max.

By definition cumulative distribution functions are always positive, non-negative value and they are monotonically non-decreasing. And there are couple of other conditions and so on. So at any point your CDF cannot exceed the value of unity for obvious reasons. And it cannot decrease. And the other obvious thing is at the left extreme F of x is zero and the right extreme it is one. So this is – these are the features of any CDF suppose I give you a function and claim it to be CDF you have to do all these checks if necessary to convince yourself that yes it qualifies to be a CDF. Not all functions qualify to be CDF necessarily.

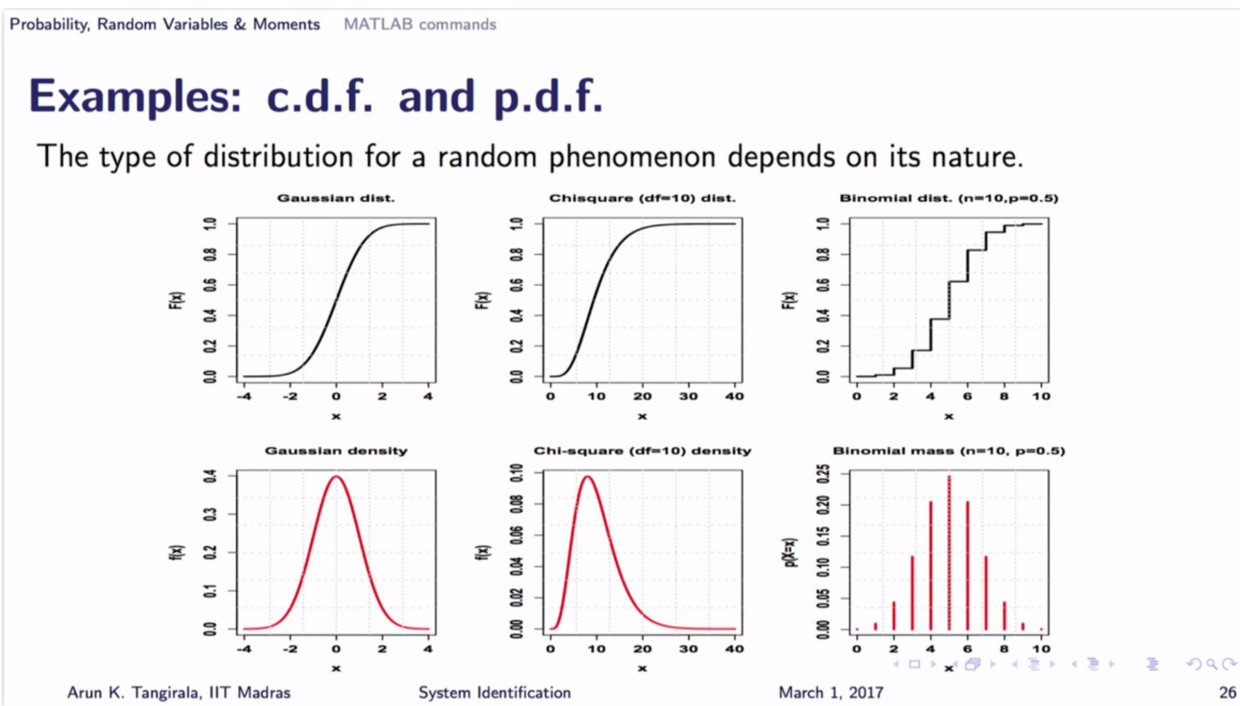
Now since we are dealing with continuous value random variables it is convenient remember this word, it is convenient to work with the probability density function. If you are uncomfortable with distribution probability distribution as I always say turn to mechanics and think of mass distribution functions. You must have read how mass is distributed and so on. There we come across the motion of density. It will – it could be mass per unit length, mass per unit area, mass per unit volume and so on. But we can speak of such density only when the mass is distributed contiguously. There has to be contiguous space. There cannot be gaps. Since we are looking at F of x that is in fact you can – that is stricter definition of a continuous valued random variable. Random variable is said to be continuous valued if its CDF is continuous and differentiable.

Earlier I said I gave you a working definition but this is stricter definition. Anyway for all such random variables now we can think of a probability density function and there are two definitions of this density function. One is that the area under the density gives me the probability. That is how you also see definitions of densities in mechanics. So the area under density gives me probability. The other definition is the derivative of the CDF will give me the density. So if I were to draw a standard Gaussian PDF notice that I use lower case f and this is dimension to denote densities. You have to understand that by looking at the notation. So once again here let's say I have X min and X max. A Gaussian PDF would look like this the most familiar one. Of course it tapers off, it doesn't exactly [00:14:40] but just to give you a rough sketch. I will show you a few sample PDFs by this one.

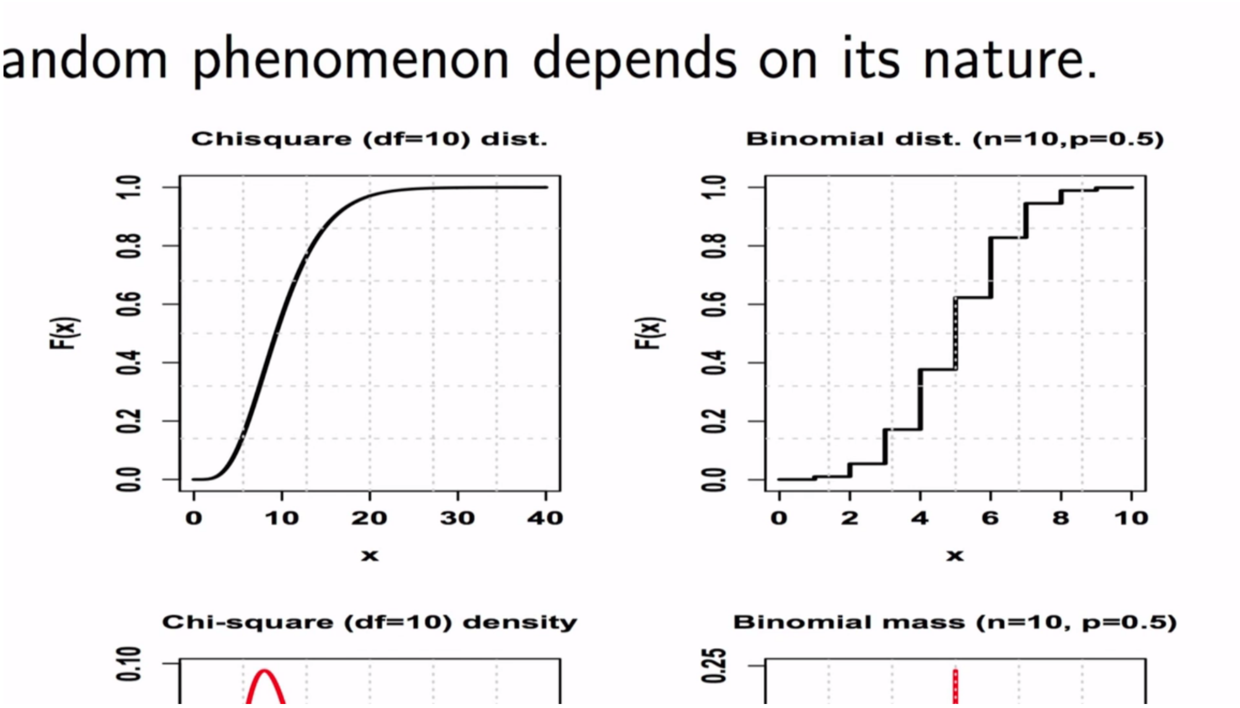
Now the area under this PDF gives me the probability. You should always remember that. If I am looking at an internal AB, this is the probability that A takes on values. X takes on values between A and B. therefore by definition area under the density function is unity because it's a probability that something should occur.

Now one of the misinterpretations that people generally end up having is that the value of the density at any point is the probability of X taking on that value. So quite often at least beginners make this mistake that F of x is equal to probability that X equals X. Now unfortunately, for a continuous valued random variable this is wrong. For discrete valued random variables we do not work with density functions, we work as what are known as mass functions. They are also derived from the CDFs. For discrete valued random variables this interpretation is okay. for continuous valued random variables this interpretation is not correct because for continuous valued random variables the probability that X will take on a specific value is zero. Now that is the irony that you can say some kind of a peculiarity in probabilities that the probability will be non-zero over an interval but zero at a point and that's because probabilities are measure functions. They are measures. They are actually [00:16:33] let's not get scared but they are essentially measures. So if I ask you what is the area of a point. Area is a measure. And you would come up with answer saying zero. So probabilities are also measures and the measure of this – this measure at any point is zero.

So in a space of continuous or continuum if you are to ask what is the probability that X will exactly take on some value that is zero. On the other hand if the interval is infinitely small then the probability is non-zero. So in other words I can approximately say that for example F of x times dx is roughly the probability or you can say Delta X take on values between X and X plus dx. This is an approximation that we roughly use and this becomes better and better as dx becomes smaller. But it cannot be zero. Dx I mean arbitrarily small. But it cannot be zero. That's why it is infinitely small. So there is something to remember, do not interpret density as a probability at a specific value because for continuous valued random variables there is the probability that X will take on a specific value is zero.

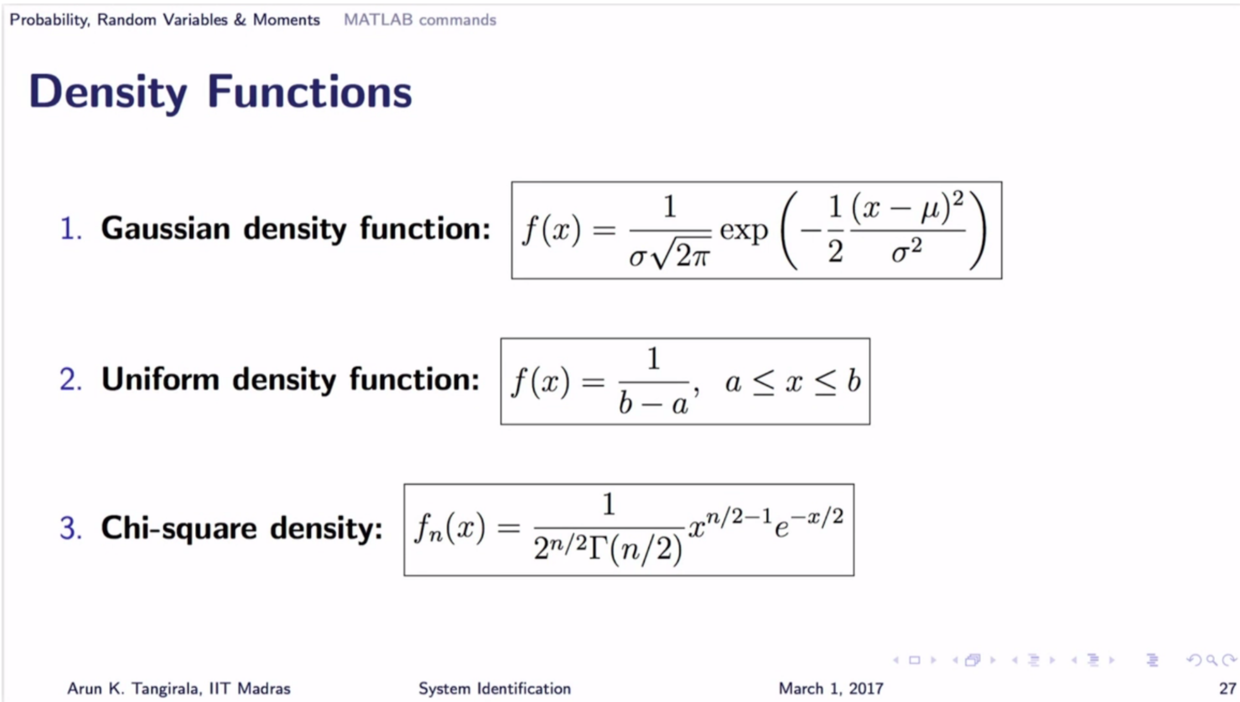


Now these are some examples of CDF and PDF. Commonly used ones. On the left extreme by the way the top panel here shows the CDF and the bottom panel shows you the PDFs. On the left you have Gaussian density at the bottom and the Gaussian CDF at the top. And in the center you have chisquare PDF at the bottom and Chisquare CDF at the top. And on the right you have not a density function but a mass function. And this is for a binomial distribution. That is for binomially distributed random variable.



I am not going to go into the definition but some of you maybe familiar. So if you look at this the CDF unlike for the continuous valued case it's staircase like function because X cannot take on values within I mean between two points. It can only take values at specific instance. So the probability that X will take on any value between two points is zero. Therefore the increment is zero. And remember F of x is , the big F of x is incremental function. It is increasing but we said before it is monotonically non-decreasing. We are very careful. Mathematicians are extremely careful. They are as precised as possible. So hopefully now you understand.

Now PDFs if you look at PDFs there are many many density functions and don't think that these are fictitious. Generally, there are a class of phenomena associated with the PDF. It is not that suddenly people decided that okay let us prepare syllabus on statistics where we will propose some 100 different density functions and whoever liked whatever came to one person's mind the density function did not show up just like that. These density functions have come by way of observing the phenomena around. It could be natural phenomena. It could be man-made phenomena and so on. The Gaussian or the normal PDF literally you can say normal I mean that's what you see everywhere being assumed and also counted is one of the most common ones for various reasons. And we will perhaps learn those reasons shortly but you know very well that when it comes to anything large you think of a Gaussian PDF but that need not be true all the time.

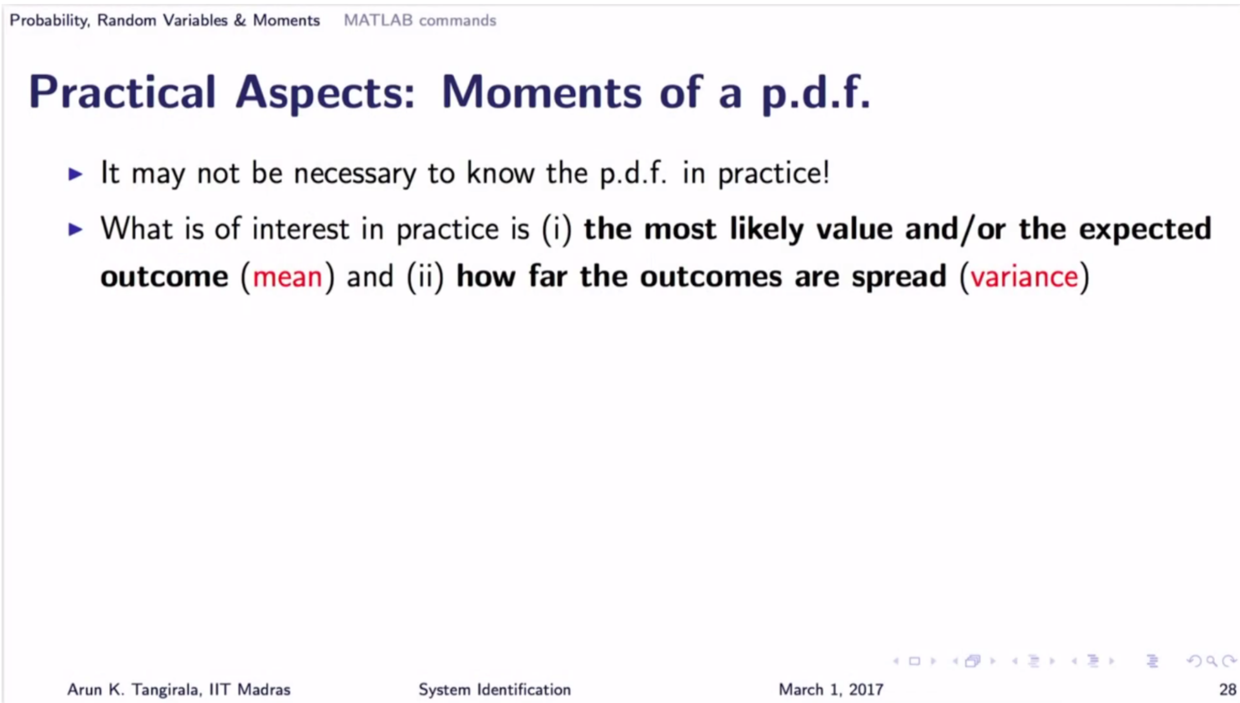


Then you had the Chisquare PDF and then you have the uniform density function. Each of this has a mathematician expression. That's the first thing you should observe. At least the three that I am showing you on the screen. And secondly each of these PDFs is characterized by some parameter. Those are called a parameters of the PDF. For example in Gaussian you have mu and sigma. Those are the two parameters. Whereas in uniform density function I have A and B. and in Chisquare it is a degrees of freedom that which is a parameter. So each PDF if there is a mathematical expression is characterized by one or more parameters. Now not all PDFs necessarily have mathematical expressions. You have to understand that. There are quite a few PDFs that are just tabulated. There is no mathematical formula for it. So it's obviously convenient to work with PDFs that have close formed expressions like this. And fortunately these are the ones that we normally run into. Of course there are others like Poisson distributions and so on but you think of Poisson distributions for discrete valued random variables and so on. As far as continuous valued random variables are concerned the typical ones that we encounter at least in linear modeling or linear random process are the Gaussian density function, the Chisquare and the uniform.

So you should have sufficient familiarity with these three density functions. As a simple homework go to the net or any – consult any resource and find what are the phenomena associated with each of these densities or distribution functions.

For example things – they will tell you Chisquare, sorry Poisson which I am not showing you here. Distributions are associated with number of events happening in an interval. It could be accidents, or it could be number of people passing through in a certain interval or space of time and so on. Those density functions are associated with distributions of averages for example. Anything that is an average and has been constructed from averaging a large number of random variables follows a Gaussian density function. That's what is essence of central limit theorem which we will very quickly touch base on.

So message is that PDFs can have closed form expression I mean have closed form expressions or may not have closed form expressions. And each PDF is associated with some natural or man-made phenomena. So by looking at the phenomenal maybe you can figure out what the appropriate PDF is. It's not always easy to do that.

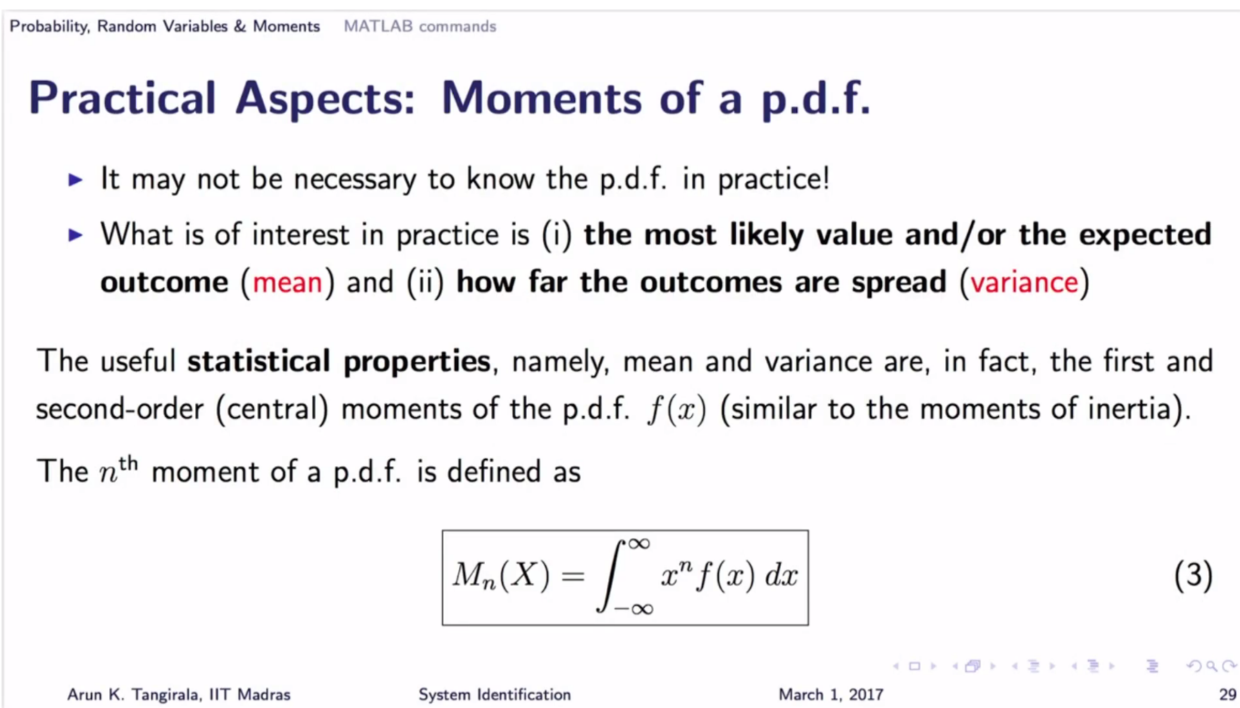


Okay. Now that we have spoken for few minutes on PDFs and CDFs practically when it comes to dealing with random phenomena it is not the PDFs that we generally work with. As a simple example that I always keep giving, [00:23:28] a new city and I have to make a decision on what clothes to pack. It's a completely new city. I don't know. I have never been there and so on. So one of the decisions that I have to make is what clothes to pack, what kind of clothes, it should be woolen clothes, cotton clothes and so on. That depends obviously on the temperature there, the climate there.

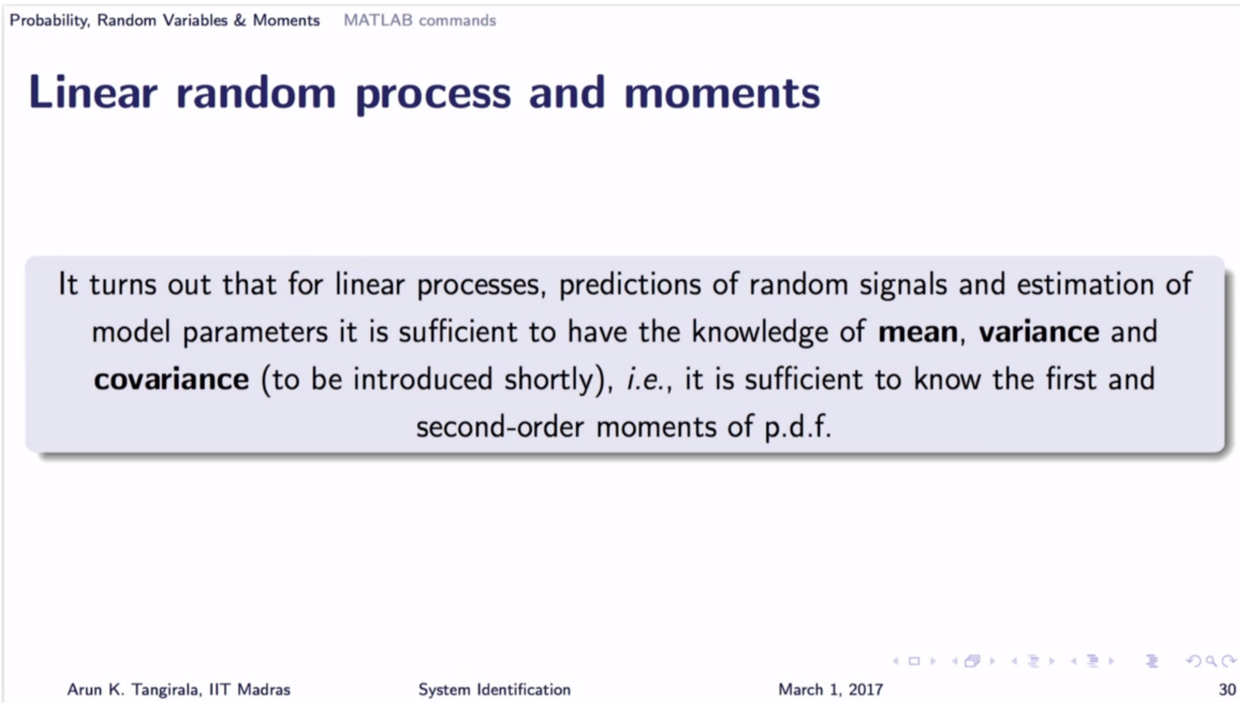
Now obviously the temperature in the new city I cannot predict. So I will treat it as a random variable. I cannot predict accurately and when I am making a decision I go to this website where there is a description of the city. The climatic conditions and so on. Today you have everything documented. And when you look up the climate part the description is never stated in terms of PDF. Even the person who wrote this – who is author of this website knows very well the temperature is not predictable accurately. Maybe also knows that it's a random variable but very rarely or I have never seen the PDF being given. Have you been seen a PDF being given for a temperature in a city? Instead you are given some other pieces of information. And what are those pieces of information? Average, minimum, maximum. They are giving you some idea. They are giving you some idea but not the full idea. Always for a random variable if you want the full information the PDF has to be given or the CDF has to be given. But the PDF is never given, why can't I get the full information because that may not be needed. If I want to get going make some initial decisions and so on it turns out that I don't need the full PDF. And there is another factor which is that estimating the PDF of a random variable from data is a lot more difficult compared to estimating these pieces of information called the minimum, average and so on.

Let me put it the other way. The number of data points for example that are required to reliably estimate PDF maybe much more then the number of data points that are required to reliably estimate the so called moments of a PDF. The moment you think of a density function that moment and this moment come together. So the moment you think of a PDF you can or any density function you also think of moments. In mechanics we have moment of inertia for example. How do you calculate moment of inertia or you talk of other kinds of moments like center of mass, center of gravity and so on. What are they? They are moments. They are giving you some idea. They do not tell you how mass is distributed. So if I look at center of mass it's going to tell me where the center is, around which the mass is distributed but it doesn't tell me how the mass is distributed. But maybe I don't need it. So the reality is that as you will see also in linear random processes you do not need to know the PDF. If you know the PDF excellent. It's great. Then you have the complete information, complete description is available. But if you do not know can you still go ahead and deal or work with random phenomena, the answer is yes. By working with what are know as moments of a PDF.

Now the two moments of interest are the first moment of the PDF which gives you the geometrical center, sorry statistical center of the outcomes. And I have a random variable I know that there are many possible outcomes but to make some initial decisions like your packing the clothes, the two critical pieces of information that I require are the average temperature and how – what is the variability in the temperature that I get to see on any given day.



Given this I can make my decisions and keep moving on in life. That's exactly the situation here. I should be at least given the center around which the outcomes are spread and how far the outcome are spread, some measure. These are respectively mean and variance and they happen to be related to the moments of the PDF. Any nth moment of the PDF is defined as X to the integral X to the n F of x dx strictly speaking there should be a denominator term as well with integral F of x dx but we know already that the integral F of x dx is unity. So we don't include that in the definition.



So the first moment I have already said that for linear random processes is a sufficient known mean variance and something else called covariance we will revisit this point at a later time.