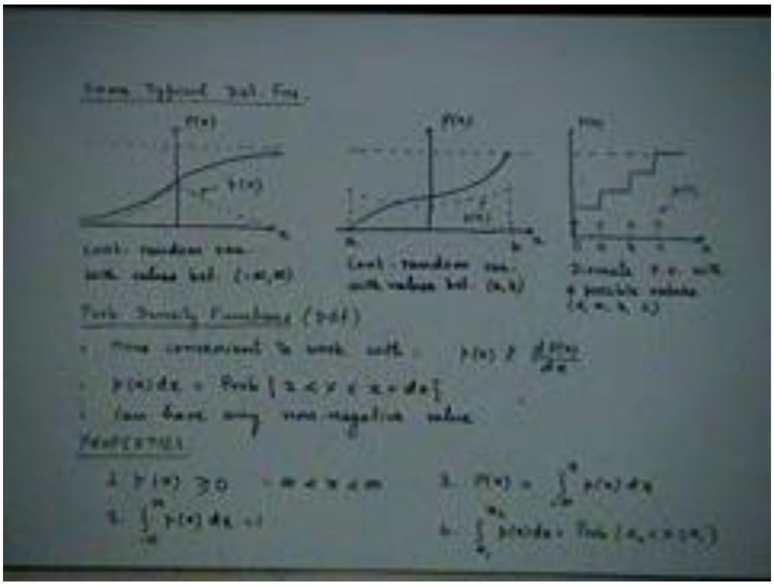
Communication Engineering
Prof. Surendra Prasad
Department of Electrical Engineering
Indian Institute of Technology, Delhi

Lecture - 28
Review of Probability
Theory and Random
Variables – II

We continue then with our review of probability theory and random variables, which we started last time. And, if you recollect, may I have your attention please, if you recollect, we had discussed the concept of the probability, distribution function of a random variable.

(Refer Slide Time: 01:23)



I am just introduce, the concept of the density function of random variable, which was defined to be the derivative of the distribution function. Typical notation is for a distribution function, use a capital letter like p capital P or capital F, for a probability density function, we use a small letter either small p or small f or something like that. So, it is typically no convenient to work with and the definition implies, this definition implies that if you were to look at the probability of an event. That the random variable x, takes a value between small x and small x plus d x, then the value of that probability in terms of the density function would be p x d x.

That is the area under vertical strip, located around the point x having the width of d x and it can have only non negative values that follows from the definition. Because, we earlier noticed that, p of x is a non decreasing function, the distribution function is a non decreasing function and therefore, the derivative will never be negative, it can be 0 some time it can never be negative. So, therefore, can have only non negative values, the density function can have a, only non negative values.

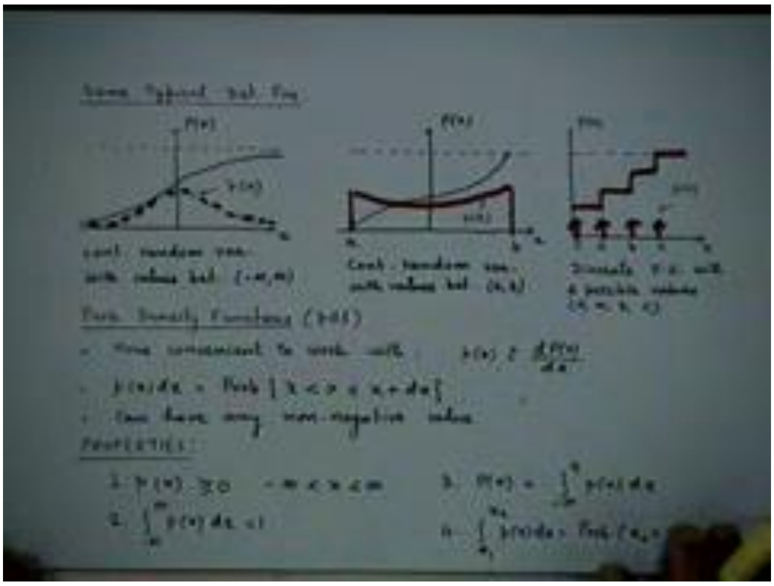
Some of the other property, which immediately follow, from the basic properties of the distribution function or this is one, that is the it has to be positive, the one which I have just mentioned, for all values of x. Secondly, the area under the density function, will specify the probability, that x lie between, minus infinity to plus infinity and that is the probability of a, sure event, therefore, it must be equal to 1.

Similarly, if you look at the area under the density function, on the ((Refer Time: 03:23)) under the density function over the segment of the real line, from minus infinity to x, that will be the probability, that x takes the value between, minus infinity to small x, which by definition is the distribution function. And of course, it follows some of the fact that, the integral of p x, would give you the distribution function, so this is the distribution function.

And finally, if you want to integrate the density function, between two limits x 1 and x 2, over the interval x 1 and x 2, then the probability that will give you the probability that the x, lie between the interval x 1 to x 2. So, these are some of the obvious properties, which follow from the definition of the density function, in terms of the distribution function. So, given the density function or the distribution function, we have a complete description of the random variable or we have a complete description of the probabilities of various events, if we may have to work with.

Because ultimately this probabilities will now be, in terms of events, that x lies in a certain interval and I can find the probabilities of such events, by this integrating the density function, over a set of suitable intervals.

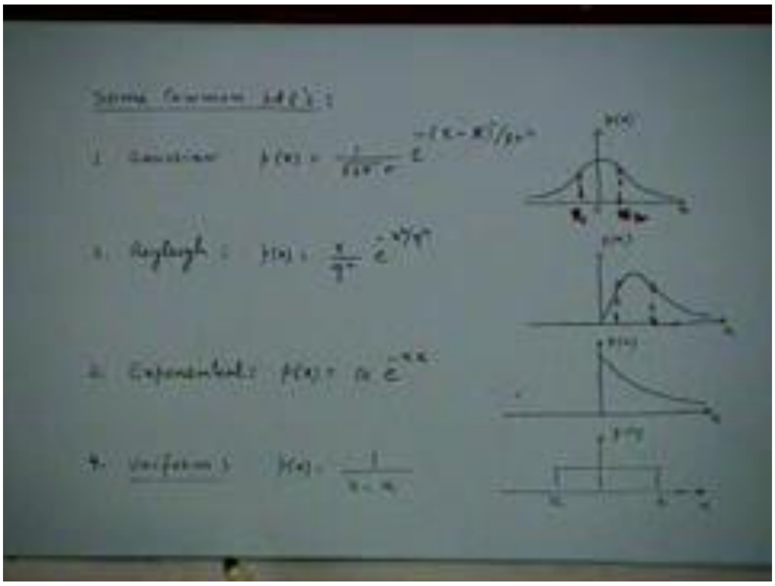
(Refer Slide Time: 04:50)



So, if you want to look at the same example, which we discussed last time for various distribution functions, for such a distribution function, if you take the derivative, you may get something like this, a bell shaped curve. So, that is the derivative is small here, is small here, but is maximum here, so that is one possible form of a density function, for a curve like this, which exist or a finite interval, you can have a density function like that, it could be, it could be something else.

For a discrete random variable, which we discussed last time, the density function would be the derivative of this would be essentially impulses located at, the points 0 a b c etcetera, at which points the distribution function takes a jump. So, the density function in this case is, contains a number of impulses located at those points, which specify the values of random variable x can take. Obviously, x can here taking only values 0 a b c, it cannot take any value between 0 to a, between a to b and so on and so forth. For the distribution function is a continuous function, the density function, here is a, is continuous function, it is defined in terms of, impulse function only.

(Refer Slide Time: 06:30)



Some of the most commonly encountered are used density functions, another notation, when I write, because unfortunately they have a, same initials, the probability distribution function and the probability density function. So, if I have to describe them in terms of initials, they will have the same initials pdf, the convention is, we use capital P, capital D, capital F for the for specifying or talking about, a distribution function and use small letters for talking about a density function.

So, when I write small pdf in small letters, I imply, I am talking about a density function, if I write capital PDF, I am writing talking about a distribution function and I hope by now the difference between these two is clear to all of you. So, some of the most commonly used density functions are these, this one this, bell shaped density function is this was so called very famous and the most commonly encountered distribution function or density function is the Gaussian density function, named after the mathematician Gauss.

Mathematically, it is given by given by this expression p of x is 1 by root 2 pi sigma, sigma is a parameter, e to the power minus x, minus x bar, x bar is another parameter, sigma whole square divided by 2 sigma square. So, this density function is 2 parameters, namely the parameter sigma and the parameter x bar, these are constants of the density function and it has the shape. We will talk about, why this is very commonly used density function or commonly occurring density function in nature, there is a reason for it.

The second example where is, one call it which is called, which is known by the name of a Rayleigh density function. It can be derived from the Gaussian density function that is if you perform a certain non-linear transformation, on the random variable x here, you get a random variable, which has a Rayleigh distribution. They are non-linear transformation being y is equal to square root of x square or something like this, basically y is equal to x square kind of that.

So here, the density function exists only for positive values of x, as you can see, for negative values of x, the density function is 0 whereas, the Gaussian density function was a symmetrical density function. It had probabilities in this case, a random variables x could take, either negative values or positive values with more or less a symmetrical distribution, in this case the random variable can never take negative values, another examples in an exponential distribution function, which looks like this.

This is often used to model queues, the waiting times in queues that you encounter in many problems, like in computer networks or when you go to a bank and you have to be serviced. So, whenever you have a queue, random the people are raving and leaving at random, the appropriate model for the amount of waiting time that you will have to, stand in the queue, is typically described by an exponential density function, the Gaussian model is extensively used to model noise in communication systems.

The Rayleigh density function, also has a set can be derived from the Gaussian density function and therefore, is extensively associated with wherever, you have to deal with Gaussian density function, then it is non-linear transformations. And, we can see, all this density functions in a way imply some a priori knowledge that is associated, by assuming that the density function is like this. For example you are saying that, there is large probability, that the random variable will take values in this interval; because the probability of x lies in; let us just specify some arbitrary interval, suppose I look at this.

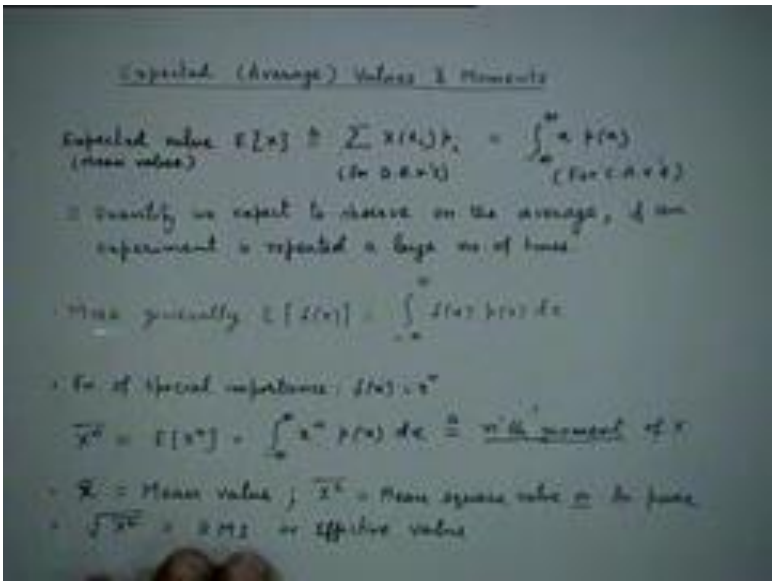
So, the area under this curve, in this interval is specified the probability that x lie between, this x 1 and x 2. So, by the shape of this density function, the kind of assumption that is implied, the kind of model that is implied is, that x will mostly take values from the neighborhood of 0, around the neighborhood of 0. Similarly, you could say something similar here it will take mostly values around the neighborhood of this location of this peak.

So, the probability of x lying in an interval like this is very large, so there is some kind of a priori information, associated with these assumptions. That you have some knowledge, that x lies is likely to lie in this region, somewhere in the, for each of these density functions. However, those are density function, which is known by the name of the uniform density function.

Can use the probability can be anywhere between a to b, which kind of implies, practically no knowledge, except to say that, x lies in somewhere between a and b. But I have no idea, whether there is any preference of intervals or which x may lie, whereas in this case, these preferences are obvious. So, when we have no a priori knowledge, and we have, we cannot say anything about the density function, the safest assumption would be to assume a uniform model, over some interval.

So, this also, very commonly used model to reflect the fact, that we have no a priori information, a priori information is very, very little or smaller. So, these are some of the commonly encountered density functions, now given the density function or the distribution function, rather said it is a complete description of a behavior of the random variable. However, many at times you do not require to work with the complete description, it is too cumbersome to work with the complete description. If you have it and you can work with it, it is nice, but many at times we need to work with, more let us say, gross parameterization of the behavior of a random variable and a complete description.

(Refer Slide Time: 12:57)



When you want to do that, you make use of what are called expected or average values of the, of a random variable, so and also come times called moments of a random variable. So, let me first define, what is the concept of an expected value, it has, you can call it by various names, which is called expected value of the random variable x, remember capital X is the notation for the it, is a name for a random variable. It could be capital X, capital Y, capital Z anything it is a name, and small x denotes the values that it can take, that is the notation we are following.

So, expected value, mean value, average value they all mean the same thing. Let us in general, they all have this and the notation is this, when I say expected value of x, this is how I will write it e of x, the definition of this notation is as follows. Basically, what we are saying is, suppose it is a discrete random variable, then it, can it the sample point can contain value discrete set of points, finite number of finite or infinite number of discrete points.

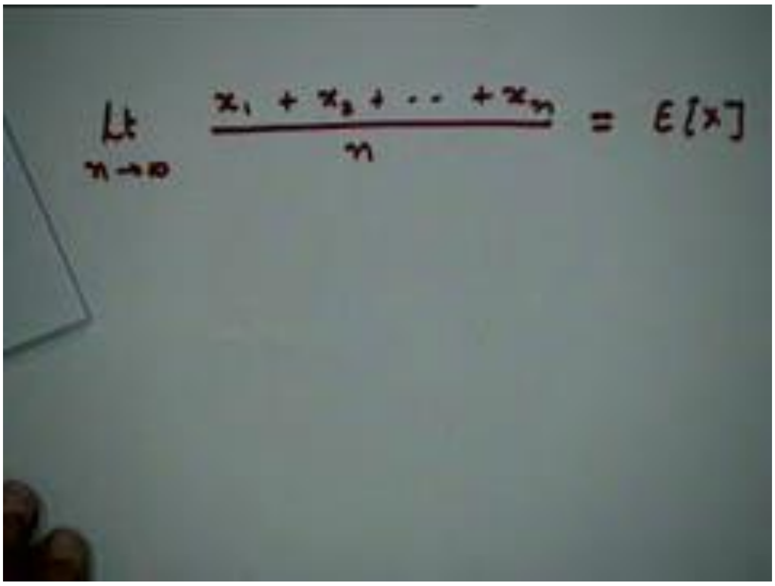
The i th point s sub i has certain a value x sub x of s sub i associated with it, which has a probability pi, so we multiply the i th value of the random variable with the probability of occurrence of that random variable. And sum it over all possible, sample points and that is the average value of the random variable x. If it is a continuous random variable, you do the same thing, you multiply the value x, with the probability, that probability density function of random variable at that point and into create the thing, over minus infinity to infinity.

So, this is for continuous random variables that, is the definition for discrete random variables, this is the definition for continuous random variable. And, the implication the meaning of this is, when I say, this is the expected value of x or this is the average value of x. What it really implies is, it is this value that you will compute form either this or this or this formula, that you are likely to see most often, if you conduct the experiment in very large number of times.

If you perform the experiment large number of times and keep a record of all the values that you get, of the random variable x. You will find that the mean value is the 1, which occurs most often, so on an average, therefore, it we said that it has the maximum value. So, typically for example, this is obvious from some of the density functions, the probability that, if I can just particular case of mean value will turn out to be 0, because as the random variable peaks are 0.

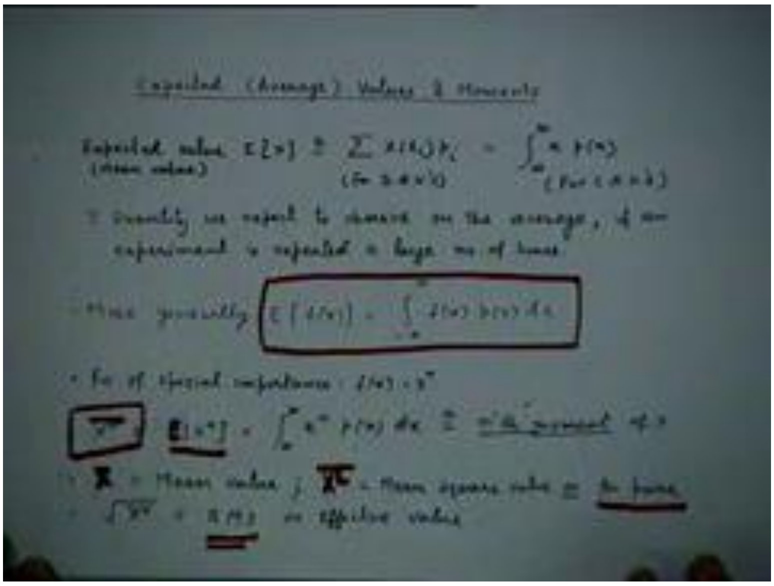
So, you will find, if you were to perform these experiment in large number of times that, this value occurs most often. Alternatively, if you were to do the averaging in the standard automatic sense of averaging, that is what you do, you just absorb the various values x 1, x 2, x 3 in the sense sum them up and divide by the number of observations that value will turn out to be 0. So these two definitions are mutually compatible consistent. This definition and the as of even normally, when you talk about averaging, the concept that you have is that of a sample average, only say, what I mean by the sample average.

(Refer Slide Time: 16:55)



What you mean by the sample average is, if you have random variable x and it takes a, you observe the values x 1 and then x 2 and then x 3, and suppose you make n observations, the sample average is simply this divided by n. And this, if you take the limit as n tends to infinity, can be shown to be equal to the expected value of x, very easy. I am sure you already know this, so I will not go into the proof of this, for that something that we need to understand. That the sample average, will also converse to the expected value, which is defined in terms of the probability density function, so we will leave it that.

(Refer Slide Time: 17:35)



Now, we can realize this notion of an average, to define the average value not of the random variable x itself, but the average value of any function of x. It will have a same definition, except that, the variable x here gets replaced with the function f of x, so that is the average value or expected value of f of x. Now, this is a very general definition of the expected value of any function of x, this becomes a special case of this, were f of x is equal to x.

In general, when f of x is equal to x to the power n, that is the very, that is the case of very special important, where n is any arbitrary integer. We come up with, what is called expected value of x of n, is not it, f of x becomes expected value of x of n and that the median, and that is defined to be the nth moment of random variable x. So, in this sense, the expected value that are defined here, becomes the first moment, because here n is equal to 1.

If I then take n equal to 2 I get the second moment, if I take n equal to 3 I get the third moment and so on and so forth. And also, there is another notation, where that you will see, you will use these two notations interchangeably, either write e expected value of x of n x to the power like this, e of x to the power n or simply put, a bar over x to the power n. So, this, this bar operation this bar notation and this e notation, they both would imply the averaging operation, the expectation operation the implication of both of this is same.

So, this is sometimes were little for short, so for example, the mean value as of defined here would be x bar, the x square bar would be the so called mean square value, these are special names for special moments. So, when you take n equal to two, the corresponding expected value of x square is called the mean square value, because it is the average of the square, square of the random variable. So, we call it the mean square value, sometimes also called average power of the random variable, just the matter of terminology.

The square root of square root of the mean square, square root of the mean square value is called the obviously, RMS value or the effective value, sometimes it is also called as standard deviation, but that is for the case of central moments, not for non central moments. So, this is a no notion of moments of a random variable, so what we have discussed here is the fact, that I can defined a, I can characterize a random variable also in terms of it is moments.

Now, what is this, what is the implication of all this, if you remember the contents in, which I have discussed this was the following. Sometimes we do not require, a complete description of the random variable, a complete description would require to, know the density function or the distribution functions. Many times, it is sufficient for me to characterize for example, the behavior of a random quantity, by just asking, what is average value.

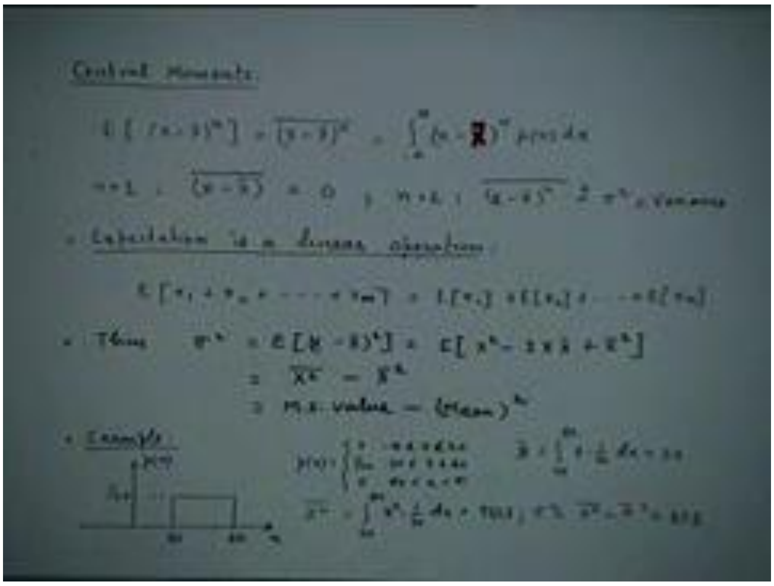
In some, what I may not be interested in, what is the detailed distribution of the marks of this class, if I just want to characterize this class. I will just ask what is, average marks of the class, that some kind of a characterization, it is a very gross characterization, but it is some kind of a characterization, it is not a complete characterization. Similarly, I could ask for the, mean square value of the marks or whatever, so that is another characterization.

So, these are all gross characterizations of the behavior of the random variable, as against the precise characterization, which would give you any kind of detailed information that you want, which is present in the density function or the distributed function, is that clear. So, that is the advantage of working with random variables moments, though as again moments of the type that I described just now, we have another type of moments, where the function f of x, is taken to be x minus x bar to the power n.

That is from the random variable x, you first subtract x bar and what is x bar, the average value. From x you are subtracting out it is average value, then raising it to the power n and finding the average value of that, this is called the nth central moment, as against expected value of x to the power n, which therefore, becomes a non-central moment. Even though this is the moment about the mean, that is the moment about the origin, the non central moment is the moment about the origin, irrespective of what kind of distribution function you have.

Whereas, the central moment is the moment about the mean, is something like the central moment of inertial and the non-central moment of inertia that you might have read in mechanics, very similar to that.

(Refer Slide Time: 23:20)



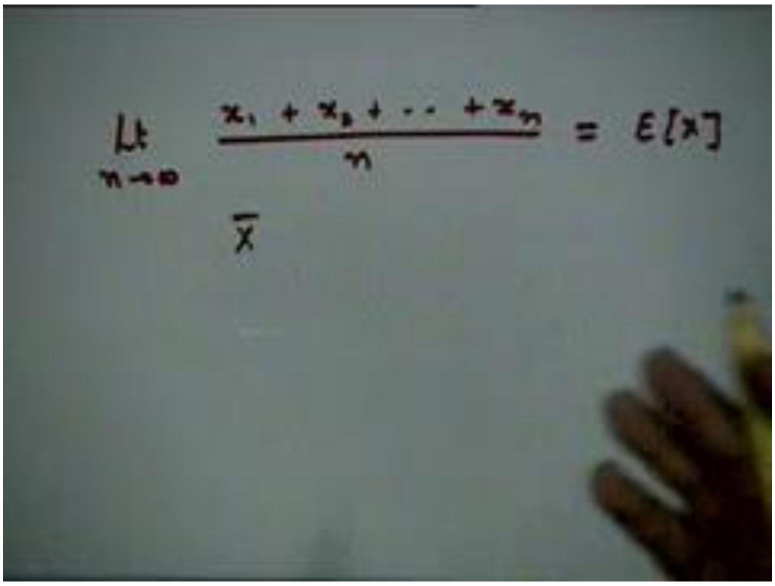
So, that is by definition, integral of x minus x bar this should be capital X bar to the power n p of x d x. You understand that, this when I say when capital I, say x bar the implication, that you write capital X, we are talking about the average value of the random variable, whose name is x. Whereas, the variable of integration is small x, please remember that, so capital X is the name, capital X bar is the value, it is a mean value, so please try to understand this notation.

Now, immediately it follows, that the first central moment will obviously, be equal to 0 is not it, because it will be expected value of x, minus expected value of a constant, which will be equal to the constant itself. So, this will turn out to be x bar minus x bar, which is equal to 0, so the first interval will always be the same. For n equal to 2, the resulting central moment is denoted, usually by the rotation sigma square and it is a measure of the average spread of the random variable around the mean value, and it is it has a name it is called the variance of the random here.

So, I introduce a few names now, the average value, the mean square value, the variance etcetera, these are special moments. The general moment is, a moment of the type expected value of x to the power n or expected value of x minus x bar to the power n. But, these three special moments they maybe the average value for n equal to 1, the mean square value for n equal to 2, and the variance also for n equal to 2, mean square value is a non central moment and the variance is a central moment.

These are all special significance and are very extensively used into description of the, in the gross description of the random variables. So, when I want to have a gross description, sometimes I am quit, satisfied to know, what is average value and what is the variance, or what is the average value and what is the mean square value, any of these two. We usually enough for a gross description, if I want a detailed description, I need to know the density function, remember these gross descriptions can be generated without knowing the density function.

(Refer Slide Time: 25:41)



If you remember that the remember this result, for example, this describe this is the gross description of x bar, because I know that, if I conduct the experiment in a very large number of times, the value of this sample average will be precisely x bar. So, I can simply perform the experiment in large number of times and compute this, and that will give me an estimate of the average value, so I do not need to know, the density function to find this.

To find the density function experimentally is a much more cumbersome job, much more difficult job, than to just find the average value. That is why, one often works with average descriptions, rather than complete descriptions, but of course, in some instances, you must have the complete description, either from physical considerations or based on experimental measurements. The next point that, I would like to make is the nature of this operator a e, that I have introduced just now.

This operator which we defined like this, you can think of this as an operator, which operates on the random variable x and this is how it operates. To perform the operation, in need to know the density function of the random variable, and carry out this operation, so you can think of e of x as an operator. It is obvious, that since this operator is defined in terms of an integral like this, it will be a linear operator. In the sense, if I want to take the expected value of x 1 plus x 2 plus x n, where all this are different random variables, then this will be obviously, equal to expected value of x 1 plus expected value of x 2 plus expected value of x n, this is obvious from here.

If you to write this as, it can be reversely proved, I think I will not go into the details, if you want to write that as x 1 plus x 2 plus x n and then, you have to multiply not with p of x, but what is called the joint density function of all this random variables. It can be shown it will be it will reduce to equivalently this operation, so at the moment I will just mention it, it will become more and more clear, as you solve some problems.

Also, it is very easy to check that, the variance sigma square can be expressed in terms of the mean square value and the average value, very simple by using, the linearity property of this operator. For example, the variance by definition is expected value of x minus x bar square, so you can write this as, expand this as x square minus 2 x into x bar plus x bar square. The expected value of x square is the mean square value the first term, the second term will become equal to minus 2 x bar square and the third terms is already x bar square.

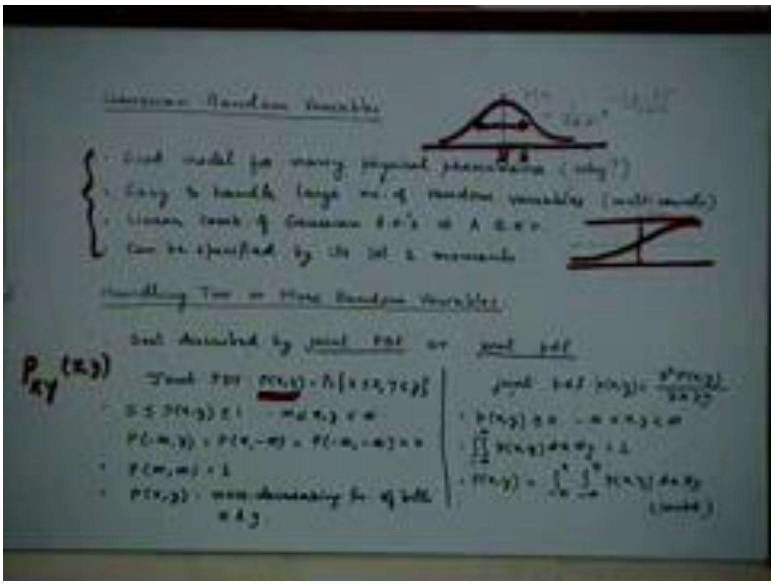
So, the second terms and the third term, will together produce minus x bar square, so if I know the mean square value and if I know the average value, I know the variance or if I know the variance and the mean value, I know the mean square value etcetera. There is a very simple example, of a uniform distribution a random variable having the uniform distribution, between 20 and 40 that is x lies between, twenty and forty were equal probability.

So, what will be the height of this density function, you have to find the area and normalize it with the area, I mean basically we have to make the area equal to 1, so obviously, the height will be 1 by 20. So, that is the description of the density function, which is 0 for x lying between minus infinity and 20, it is 1 by 20 between 20 and 40, and 0 from 40 onwards. If I want to compute x bar based on this, this is what you have, integral of this from 20 to 40, the value will turn out to be 30, which is obvious it has to be center of this distribution function.

The value of the mean square of the random variable would be similarly, 933.3 and the variance will turn out to be 33.3. And, the square root of variance is sometimes also called the standard deviation that will be of the order of five point, something 5.6, 0.5, 0.7. Any questions at this point, these are the concepts which you know, that I am quickly reviewing for your convenience, as I mentioned earlier sometime ago, a few minutes ago, one of the most commonly used models for randomness is gaussianity.

Is the assumption that the random variable has a Gaussian distribution, when I say Gaussian distribution, would basically it implies is the density function has a Gaussian shape, the typical statement would be random variable x is a Gaussian density function.

(Refer Slide Time: 31:18)



And we already seen what a Gaussian density function looks like, it is a bell shaped function and we also know it is mathematical characterization. As you remember, it is described completely by two parameters, which would, which we would called x bar and sigma. Now it can be shown, that if you take the mean value of this density function, it will indeed it will turn out to be the parameter x bar, which can be used in this description.

So this x bar is nothing but, the mean value of this density function, the mean for example, this may not be say always peaking are 0, it may peak at some value, which we are calling x bar. Also, the variance the sigma value of sigma, the standard deviation essentially is the description of the how wide this distribution density function is, whether it is narrow or whether it is wide, the sigma is large that means, the spread around the mean value is large.

Sigma is small, it is a narrow function and they spread along with, along around the mean value is small, so this is the measure of the spread. Unless, this is a good model for many physical phenomena, what is the reason for that, the reason for that, is some a reason, which I am not going to discussion of, which those of you who did require some probability theory. Probably now, is resultant statistics which is known by the name of central limit theorem, have you heard of it.

Student: ((Refer Time: 32:50))

The central limit theorem states that if you add up a large number of random variables, then the resulting random variable would tend to have a normal distribution that is a ((Refer Time: 33:11)) incidentally another name for the distribution is a normal distribution. And the reason why it is, so called is because it is widely seen in practice widely encountered in practice.

So, the result says, if you add up the large number of random variables, irrespective of what the distribution originally is the resultant random variable, which is the sum of all these would tend to have a Gaussian distribution, this can be mathematically proved, without any difficulty. So, I am not going to proof of that, it is very easy, but we these are the scope of this review here, so this is called the central limit theorem or the strong law of large numbers, these are the various names, by which it is proved.

Now, therefore, what is the how does that imply that, this would be the most commonly seen model in physical many physical phenomena. It tells that the many, many physical phenomena have this nature, that what you observe is a gross observation, if you go into the details of that observation it is composed of the outcomes of a large number of experiments, as if they were performed independently of each other.

So, irrespective of how this experiments individually have the characterization, the how do they statistically characterized, what you finally, observe on the macro scale, on the large scale is the sum total of a large amount of minute effects. For example take the noise that we talk about, what is the noise due to, it is due to random motion of electrons somewhere, either in a conductor or in an electronic device and these electrons are flowing very randomly there motion is random.

So, the sum total of all this flow of electrons, is what you absorb as the average current, as a current. So, if you have to look at the current value, which is, so many electrons flowing per second, really it is not one electron or two, the effect of one electron that you are seeing the effect of, two electrons that you are seeing. You are looking at the sum total of large numbers, so irrespective of the distribution of actual flow of electrons, what you really observe, your Gaussian distribution.

So, there is an average value and if you have to look at around you, sometimes find that it is not constant, the current is not constant really, you may assume it is constant, the average value is constant, there is an average drift of electrons in one direction, average rate of flow of electrons. But, if you look at the minute effects you will find that the actual current varies, because sometimes the larger number of electrons is following, sometimes the smaller number of electrons is flowing and therefore, there is a variation and that variation is what we called noise.

Noise is nothing but, this random perturbations, about the mean in this particular phase, so most noise that, you really observe in electronic devices, which is the source of trouble for us in communication arises from this kind of a principle and therefore, tends to have a Gaussian distribution. So, that is the reason and this is true for not only this particular phenomenon, but many other physical phenomenons in nature and that is one of the reasons, why Gaussian model is so popularly used, in many in modeling many physical phenomena.

Also, it turns out besides a physical reason of the kind that I just given, mathematically also its very convenient model to use, you can handle it mathematically very easily, as compared to many other models and particularly, when you have to handle a large number of random variables, rather than a single random variable. A very important property of a Gaussian random variable or Gaussian random variables is that, if you some adopt a few Gaussian random variables 1, 2 or 3, the result always will be a Gaussian random variables and in a combination of Gaussian random variables would produce a Gaussian random variables.

Students: ((Refer Time: 37:11))

That you will have to find, for example the mean value distribute this, Gaussian random variable and a variance would be different, you have to find that, out you have known the mean values of the individual, whether you can directly add or not, will depend whether they are independent or dependent. The mean values will add, but the variance is variance is may not only add,

Student: ((Refer Time: 37:36))

Sorry.

Student: ((Refer Time: 37:38))

No, not necessarily, that is when they are independent, if they are not independent, the variance will not add, because there will be cross terms, which may have positive or negative values. So, we will look at the maybe few problems, to take care of these issues, now the most interesting fact, about the Gaussian random variable is, it is parameters are x bar and sigma. If you remember sigma square and x bar is the first moment, and sigma square is the second moment, it can be completely therefore, described if you know, it is first two limits.

So, whereas for a, arbitrary random variable for an arbitrary random variable, the first two moments will be just precisely that, to gross descriptions, the first order description and the second order description. For the case of Gaussian random variable, these two moments will also be complete descriptions, because if I know this, I know precisely the mathematical form of a, distribution function in terms of these two. In general, to derive a complete description namely the probability density function, from the moments is difficult it is possible, but it is difficult.

And, in general it will require infinite number of moments, first order moment, second order moment, I need to know the moments of all orders first, second, third, fourth, fifth, because then there is a serious expression, that I can use find the density function. Those of, who done the course will know, but I require infinite number of moments to do that, for the case of the Gaussian density function, two moments are sufficient to describe it completely, namely the first and the second moments.

So, it can be, these are some of the very useful and important properties of a Gaussian random variable, which you should keep in mind, in which we shall use extensively. Typically, the distribution function of a Gaussian random variable looks like that, which you already seen that, this that is the density function and that is the distribution function, the integral of this.

Students: ((Refer Time: 39:50))

No, that is not a reason that is the convenient thing that we have, but that is not the reason, the reason is I already described to you. That, physical phenomena would tend to have this distribution, actually tend to have this distribution because of the applicability of the central limit theorem that is the reason. Now, the next thing we, will quickly look at is, the fact that sometimes, in fact, many times, we need to handle more than, one random variable at a time. Then, how do we characterize things, how do we go about, so how do we handle two or more random variables.

Basically what we are saying is, these two random variables, they maybe they may be defined on the same probability space. Remember the probability concept of probability space, there is an experiment, there is a sample outcome and there is a space of events, and there is a probability measure defines on the space of events. But I can define random variables in many ways, it is a mapping from the sample space to the real line, I can define this mapping in many different ways.

So, suppose I have two or more random variables, that is two or more different ways of defining this mapping, and I want to characterize this different random variables that I have together, many at times we need to do that. So, when you want to handle two are more random variables, we can more or less extend all the notation that we discussed for. A single random variable to handle the situation, by little use of I defined that yesterday, instead of joint probabilities. We talked about joint probability distribution functions and the joint probability density functions, and this is how they are defined.

The joint the now the proper notation for this would be for example, for p x y proper notation should be p x y, because I am talking about two random variables x and y comma x y, and what are the definition of p x x, p sub x of x, ((Refer Time: 42:24)) to the x is less than or equal to small x or the x lies between minus infinity to small x. So, the joint probability distribution function of x and y, is the probability that x is less than or equal to small x and y is less than or equal to y small y.

So the, it is a joint description of the behavior of x and y, so x lies between, minus infinity to x y lies between minus infinity to y. The corresponding density function is the second order derivative of this distribution function, with respect to x and then with respect to y or with respect to y and then with respect to x. Now, some of the properties should more or less obviously, follow from here let us, look at them, one look at them one by one.

Obviously, this will lie between 0 and 1, because what is the value in, that x is less than minus infinity and y will also less than minus infinity that will correspond to the probability 0, probability of n possible event. What is a sure event x is less than infinity less than or equal infinity and y is less than or equal to infinity, that is a sure event obviously, x and y will always be less than or equal to infinity. So, that will be the maximum probability that you can have.

For any other values of x and y the probability will lie, somewhere in between these two, so the probability of x and y the joint probability of x and y will always be in between 0 to 1, for x y going between minus infinity to plus infinity. If I put, one of them as equal to minus infinity either x or y, what will it give will give,

Student: ((Refer Time: 44:28))

The impossible event again, so therefore, the joint probability of all these three kinds, probability of minus infinity y, x minus infinity or putting both of them to minus infinity would also, that is the possibility of being possible event. If I make, both of them as infinity is this one I already said this, in general, it will be non decreasing function of both x and y. So, now if we have to visualize this, we have to think of these three dimensional picture, p x y is the plot of a surface or ((Refer Time: 45:09)) a plane.

So, you have a plane and you have a surface or the plane, you have to visualize it like that and which has these properties, I cannot draw these pictures you have to imagine it. The corresponding density functions properties would be, same properties it will be always positive, the volume under the surface, under the plot of the density function surface, should always be 1.

And, if I integrate this the joint the surface representing the joint density function between minus infinity to x and minus infinity to y, that will be the probability of the joint probability of x and y, ((Refer Time: 46:01)) distribution function of x and y. So, these are some obvious, extensions of what we discussed for the case of a single random variable. Any questions here, I am going little too fast here, I know that, but then that is deliberate, because I know that you have already know these things. So, you see, but if there is any questions I can slow, slow down just a little, is it ok.

Students: ((Refer Time: 46:24))

Good, one a couple of other important properties, this is very important, if I make ((Refer Time: 46:37)) variable equal to infinity not minus infinity, then what else that equivalent is saying, that equivalent is saying that x is less than or equal to infinity it, does not matter. Basically, we are saying that x value is of no consequence obviously, x will be somewhere less than or equal to infinity, so the probability of this event, will be essentially dictated ((Refer Time: 46:56)) by the probability of the event y is less than or equal to y, so it will be nothing but, equal to p y y.

So, essentially this is a very important point, the distribution of function of y can be obtained from the joint distribution function, by putting x equal to infinity. Similarly, the distribution function of x can be obtained from the joint distribution function, by putting y equal to infinity. So, we say that, in relation to p x y, p y, is a marginal distribution obtained by marginalizing the joint distribution function with respect to x, similarly p x is marginalized distribution function, obtained from the joint distribution function of x and y.

So, in relation to these joint distribution functions p x x and p y y are called the marginal distribution functions, that is basically inter relationship between these two than anything else, we can talk about these independent of the joint distribution function. But, we have talk of these, in relation to the joint distribution function, the relationship is expressed by using the word marginal, and this is the marginalization process in the density function domain.

If I want to derive the density function of x, from the joint density function of x and y, basically I have to allow y to take, all values from minus infinity to plus infinity. So, integrate the density function from minus infinity to plus infinity, the resultant will be the density function of x alone, and similarly for the density function of y. So, these are the marginal pdf's marginal density functions, obtained from the joint density functions, so joint density function would imply a marginal density function that is the density function of x alone or of y alone and similarly, the distribution functions.

Now, here is some example is small simple picture, which I could draw for a density function of two random variables, which is like this. It is what kind of a density function is this.

Student: ((Refer Time: 49:26))

It is a uniform density function, where x lies anywhere between x 1 and x 2, y lies anywhere between y 1 and y 2, but it can lie anywhere in this rectangle for the same probability, the volume under this surface that you have must be equal to 1. The density function here is given by this, 1 by x 2 minus x 1 into y 2 minus y 1, so now whatever, so that is the simple picture that a, in normal density function for example, would be very similar to the bell shaped that I have discussed, but it will be a, just you think of that bell shape rotating around the x y plane.

So, that becomes the two dimensional normal density function or two dimensional Gaussian density function. I am not drawn the picture here, but you can imagine it I am sure. You have seen many bell it will probably have the shape of that bell, now these concepts, which we have discussed can be now extended to any number of random variables, I have talked about two, but you can talk about three, four any number and we will in fact use them, when we go to random processes. Because, we will find the random processes required, this rules that we are talking about here, I think I will stop here and continue from here next time.

Thank you very much.