Statistical Mechanics
Prof. Ashwin Joy
Department of Physics
Indian Institute of Technology, Madras

Lecture – 02
Continuous Probability

Good morning to all of you. So, today we will talk about Continuous Probability distributions and to begin the discussion I would like to summarize what we learnt in the last class. We discussed about discrete random variables and in the context of discrete random variables you always talk about a set of some outcomes.

(Refer Slide Time: 00:41)

So, I will talk about for example, in the case of a dies which has 6 faces, then the set of outcomes is basically 1, 2, 3, 4, 5, 6 you have no more no less than these 6 outcomes when you toss a die that has 6 faces. You could take coin and write down the set of outcomes as head and tail. So, it typically what this means is that the probability of having an outcome as 3 in the case of a die is basically 1 upon 6, because there are 6 faces and the die is unbiased.

So, each face has a probability of 1 by 6 and the case for coin the probability of getting a head in a single toss is 1 by 2, because the coin is also unbiased ok. Now before we proceed to the discussion of continuous probabilities; let me tell you a few important axioms related to or axioms in the context of probability. So, these are like facts stated

without proof and any notion of probability must satisfy this ok. So, the first axiom of probability is that the probability of any outcome.

Let us say E is the outcome, E could be the face of a die or the face of a coin has to be always positive. So, probabilities are always positive it cannot be negative. So, this is called as a positivity of probability, you cannot have probability of getting a head of a coin as negative, it has to be always positive. Second important axiom of probability is connected with events that are disconnected or independent suppose I ask you what is the probability of getting a event E 1 or E 2 ok.

So, this is basically meaning that I am asking the question; that what is the probability of having either E 1 or E 2 provided E 1 and E 2 are independent or disconnected. What this means is that suppose you take, you ask me the simple question you take E 1 as head and you take E 2 as tail and you ask me in a coin toss what is the probability of getting a head or a tail, the answer is 1, because you have only head or tails.

(Refer Slide Time: 04:19)

So, you will say that probability of getting a head or a tail is nothing but the sum of these 2 probabilities and that is 1 which means you will always get either a head or a tail. And of course, as I said head and tails have to be disconnected events ok which is definitely the case here. The third important axiom of probability is that it is always normalized which means if I have a set of events E 1, E 2 all the way to E n then these probabilities sum up to 1.

So this is a very important condition on probability that your probability is normalized so this is called as a normalization condition on probabilities. So, if you have information to all the events then the sum of probabilities of all these events; in the case of a coin tosses I know probability of a head plus probability of a tail is equal to 1. So, for the coins you know this and for the die you know the probability of 1 plus probability to 2 all the way to probability of 6 this will had a two 1 for the.

So, we have 3 different 3 important conditions here, one is the positivity of probability, probability of any event has to be always positive. And if we have 2 events which are mutually disconnected or independent then the question what is the probability of getting either E 1 or E 2 is nothing, but the sum of these 2 probabilities. And the final question is: if you have a collectively exhaustive set of events that you have no more than E 1 E 2 all the way to E 1 E n events you have all the events when a sum of probabilities of each of these adds up to 1 ok.

So, 3 important axioms connected with probability, discrete probabilities are also seen to carry forward into the domain of continuous probability distributions as we shall see right. So now, I just want to tell you few important concepts regarding measurement of probabilities.

(Refer Slide Time: 07:05)

So, how does one measure probabilities of any event? So, this is the question how do you measure.

Well, there are 2 ways to measure probability of any event and just by the difference in the nature in which they are measured I call them objective method and the subjective method, I will give an example what is this meaning of objective and subjective ok. So, suppose you have, you take the example of the dies 6 face die, now suppose you ask me what is the probability of getting an outcome let us say 4.

In a die which is 6 faces well I would say you toss the die n number of times and count how many times you get 4. So, N 4 being the number of times you get the face 4 and N being the total number of times you have tossed your die. Now, if N is large that is if I take the N here to infinity then the probability of the face 4 coming up in a single toss becomes the right hand side ok.

So, if you N to be 1000 or 10000 ok so, this would give you the probability of getting the probability of getting a particular face 4 out of N measurements, but you would also say that in the case of a die toss since each face is unbiased you can argue that no experiment is needed. And, I can simply say that the probability of getting you know the face 4 is simply 1 by the total number of faces ok.

So, in this case for the particular case for die toss you could have answered the question subjectively without carrying out any experiment because, you knew exactly the number of faces and you knew the fact that the die is unbiased. So, the objective and subjective methods will lead to the same answer, you do not need to do it objectively the subjective method suffices. So, I can simply say that the limit N tending to infinity number of times you got the face 4 divided by N basically tends to 1 by 6 ok.

So, both this method we will give you the same answer; in the particular case of a die toss subjective method suffices, but you shall also run across methods were a situations where you may not know how to compute the probability subjectively. In that case your only resort is to do it objectively, do a simulation perhaps or do an experiment perhaps and find out the probabilities objectively fine. So now, let us proceed to the discussion on continuous random variables.

(Refer Slide Time: 11:17)

So, we will talk about continuous random variables now and they are of great importance in statistical mechanics, I am going to spend significant amount of time building the mathematics of continuous random variables. So, you can think of you know set of random variables going from some a to b. So, your random variable x is in the set a to b ok. So, you can think of your x to be between a and b and S of x is nothing, but any outcome of a measurement of x ok. So, S of x basically refers to an outcome of measurement on x.

So, if you measure x you will always get S of x between a and b, that is a meaning of S of x ok, now I can take a to b minus infinity and b to b as plus infinity. So, if the same S of x will become minus infinity to plus infinity. So, it is like here there was a bounded interval, but I am not respected to these bounded intervals, I can take S of x to be unbounded. It is possible to have an outcome of measurement of x unbounded; for example, the velocity of gas particles in this zoom has a velocity distribution that is unbounded.

You can have in theoretical probability you know to have a it is possible to have a finite probability of finding a gas particle with very large velocities, its only becoming 0 in the limit velocity is getting infinite ok. So, velocity distribution is one such example, what needs to be understood, here is that notice in the case of discrete probabilities the normalization imply that your sum of all probabilities x i being in one of the N outcomes

on x. This was 1 this was your normalization condition that you had all the outcomes N outcomes and the probability of each outcome if added gives you 1.

For continuous probabilities or probability of a continuous random variable this normalization condition reads as of course I cannot have a summation now, because I have infinite number of measurements in any interval whether finite or bounded interval. I have to write down my normalization as some minus a to plus a p of x dx as 1 ok. Because, whether the interval is bounded in this case I have a bounded interval ok. So, my interval is basically minus a to plus a S x being a measurement of the random variable. Even though the interval is bounded the random variable in principle can be anywhere in this interval.

So, there are infinite real numbers that I can conceive of in this bounded interval minus a to plus a which means I cannot work with this summation I worked in the discrete probability case; I have to work with an integral. Now a direct consequence of this integral is that the probabilities in the second case have become dimensional. This has a very important point that p, and I am going to use a different color just to alert the reader. The very first qualitative difference that our probabilities have picked that they are now dimensional.

(Refer Slide Time: 16:07)

So, when you go from discrete to continuous the first thing that happens to probabilities that they become dimensional and the dimensions here are 1 upon x, because right hand

side is 1 which is dimensionless. So, probability is a now dimensional which means they are not really probabilities, they are not really probabilities. In fact, one should call them as probability densities; they are now called as probability density. So, this is an important difference that you must keep track of when you are discussing continuous variables.

So, continuous random variables are described by probability density functions, not probability themselves. Only discrete random variables are specified by probabilities which are dimensionless. So, here I will say that p of x are dimensionless they are just numbers your numbers in the case of continuous probabilities p of x are probability densities. So, how does one construct probability densities? So, how do you that is the next question that is come to my mind ok. So, if I am going from discrete probabilities which are purely numbers how do I compute Probability Density Functions or you know commonly refer to as PDF's. So, that is the next topic.

(Refer Slide Time: 18:27)

So, to construct PDF's, you need notion of cumulative probability function. So, this needs a notion of Cumulative Probability Function also called as CPF and I shall define what is the CPF. So, you take off this capital P of x as a cumulative probability function and this P of x is nothing, but it is nothing but is the probability of getting you know a measurement x in the range minus infinity to x ok.

Suppose my x varies between minus infinity to infinity and P of x, this capital P of x simply tells me if x belongs to minus infinity to plus infinity. So, P of x is simply probability of getting a measurement on x between minus infinity to x. So, let us sketch this schematically. So, this will go to plus infinity and this will go to minus infinity. And if I am plotting the cumulative probability function, then I know at minus infinity the probability of getting an x is 0. Because that is just one point and as you increase x is value is slowly cross the y axis at which point you will say the cumulative probability function is half and then as you go towards the positive infinity.

So, let me just do this properly it will saturate to 1 ok, it will saturate to 1. So, this point would be 1 and this point will be half, this is the origin 0. Now, it has to be half at origin, because if x is a random variable between minus infinity to plus infinity, the probability of getting an x between minus infinity and 0 should be half. Because that is precisely the half range of x, if x is between minus infinity and plus infinity; the question what is the probability of getting an x between minus infinity and 0 is half because, precisely where half of your measurements are.

And naturally the question the answer and the question what is the probability to find in x between minus infinity to plus infinity. The answer is 1 because, you always find an x on the x axis. And so the CPF is 1 as x goes to infinity, because that is the definition of CPF.

(Refer Slide Time: 22:17)
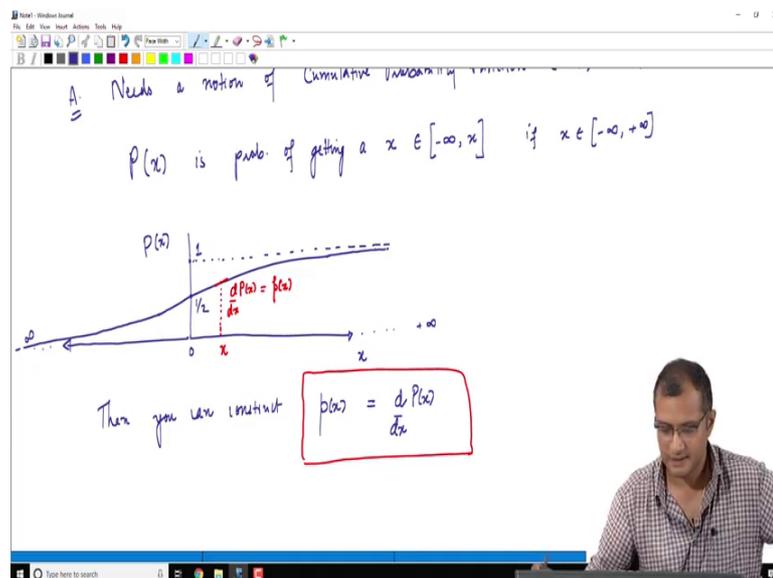
But then, you can construct the probability density function p of x as simply the derivative of the cumulative probability function. This is the definition of probability density function. So, if you want to know what is the probability density around some point, let us say this point. This is the x, you simply compute the derivative here called d by dx of P of x and this is nothing but the probability density at point x.

So, that is basically the measurement that is basically the way you construct a probability density function knowing the cumulative probability function. So, this is one way to construct PDF's; let us now proceed to some examples where you will encounter you know in statistical mechanic.

(Refer Slide Time: 23:43)
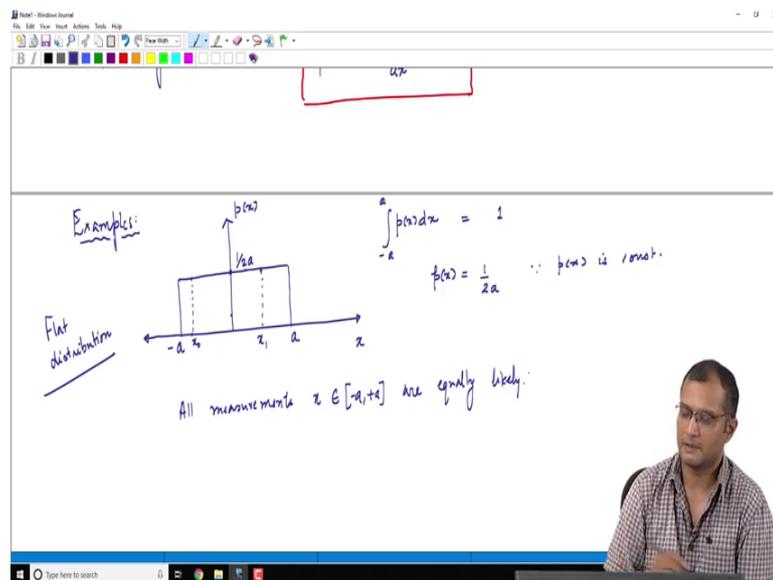
So, some examples of PDF's that you see in physics I will just bring up 2 simple examples and you can refer to the upcoming tutorial for more important example. So, I will just give you 2 examples of 2 commonly found examples of PDF's. So, one is a step distribution step probability density function.

So, this is like saying that the probability of funding an observation between minus a to plus a is constant is uniform. So, it is a flat distribution function if I ask what is the probability of getting a you know at this point, well the probability is this which is 1 upon 2 a. This point on the x axis 1 upon 2 a, it has to be 1 upon 2 a, why? Because if I normalize my PDF between minus a to plus a as the range of my PDF I must get 1 ok.

And since p of x is constant; so, it can come out of the integral now integral dx between a and minus a is simply twice a the right hand side is 1. So, this will give you one upon 2 a. So, this is you know an example of a PDF that any measurement that you take on the x axis between minus a and plus a the probability is probably density is 1 upon 2 a. So, this is also called as a flat distribution ok. So, all measurements between minus a and plus a are equally likely ok. So, all measurements you know x between minus a and plus a are equally likely.

(Refer Slide Time: 26:21)
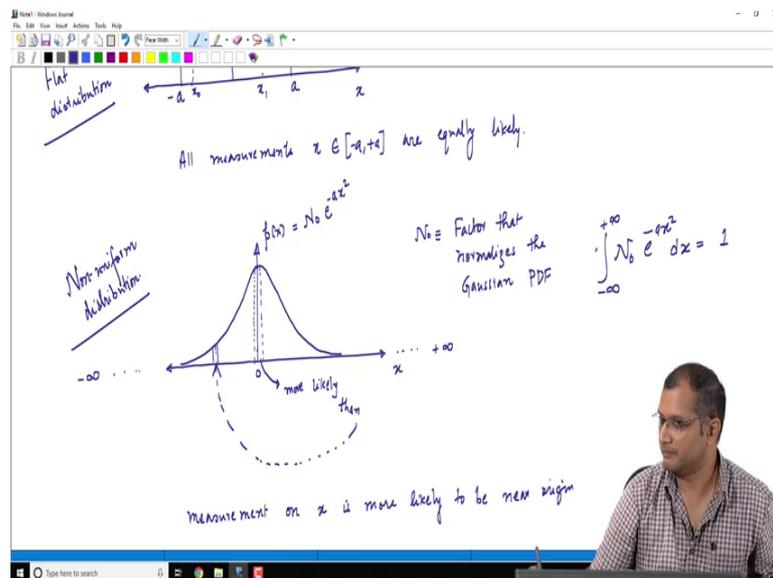
You can also take some non uniform distributions which is I given the example of a uniform distribution; I am also giving you some important non uniform distributions ok. So, one distribution that you have probably heard already in statistical in your prior encounter with kinetic theory is a velocity distribution which is Gaussian. So, I will take a one dimensional Gaussian distribution; now this distribution is in does not the domain of the independent variable is unbounded, also called as a bell curve ok.

So, x will go from minus infinity to plus infinity and this PDF here is maximum at x equals to 0 and this PDF has a functional form of some normalization constant that will be made clear later into E raise to minus x square. So, this distribution function falls on both sides of the y axis and normalization constant is basically you know this is basically the factor that normalizes the Gaussian PDF's, the Gaussian PDF. Meaning that I can

calculate what is my N 0 by simply integrating my PDF and demanding that this b equal to 1.

This will give me N naught ok, we shall do this later on and find out what is precisely the value of this n naught, but for the purpose of this discussion it is not necessary to computed. Now just compare this distribution with the previous distribution, here it is more likely to find an observation of x close to the origin. So, this region is more likely then maybe this region and the simple reason is there probability densities are much smaller here compare to the probability density around origin.

So, most of your measurements will be sittings near the origin, where p of x is maximum and as you go far away on either side of the x axis from the origin you see that the PDFs become smaller and smaller, which only means that if you take a measurement on x it is likely to be situated near origin ok. For example, if x was nothing, but the x component of velocity of articles in this room then if you take a measurement on the x component of velocity is of particles in this room you will find that it is more likely to find a particle with 0 x component of velocity, which means its just says that the average x component of velocity in his room would be 0. And so, we will be the y component and so will be the z component and that is because the particles in this room do not have a drifting centre of mass movement.

So we have 2 different types of probabilities here, flat the probability distributions and non uniform probability distributions and we will discuss several such distributions in the upcoming lectures and in the tutorial we will discuss how to go from a flat distribution into a non uniform distribution my TA's will discuss through the aid of some problems and it shall become more clear how do we do this ok. So now, I am going to basically describe one very important method of mapping.

(Refer Slide Time: 31:11)

One distribution to another, so I am going to partly answer the previous question leaving some room for discussions for the to the assignments that is coming up ok; but I will discuss some theoretical steps and one example I will discuss in this class just to aid the understanding. Now suppose you I have a continuous random variable x and the PDF of this random variable is p of x ok.

So, p of x is the PDF x is a random variable p of x is the PDF. Now suppose, let me just define let us take some example of this PDF ok, before we complete the question. So, let us say this PDF is exponential. So, say p of x is e raise to minus x and e raise to minus x is a PDF, valid PDF that goes as an exponential. This point is 1, this is 0 you have infinity and if you want to check that p of x is a valid PDF you can compute its norm ok. So, you take the range 0 to infinity and this will come out to be minus e raise to minus x in the range 0 to infinity, which will be 1.

So, this is the properly well behave probability density function. So, the norm is 1. So, norm a question is if p of x is e raise to minus x and that is given to you, you consider another random variable now this is my question so please pay attention here. So, consider a new random variable continuous random variable let us say it is y and this y has the relationship with x, let us say this relationship is quadratic y is x square hm.

(Refer Slide Time: 34:21)

What is the PDF of y? Meaning let us say the PDF of y is p of y and this is the function of y what is this p of y this is the function of y that is my question. So, the answer to this question is basically relies on conservation of norm or conservation of the normalization conditions. So, for each measurement of y each measurement of x if there is a unique measurement of y by the way this measurement need not be unique the function y could be non monotonous, but as long as you account for all the measurements on y appropriately.

Then you can write down this condition, but p of x dx which is nothing but the probability of getting measurement x in the interval x 2 x plus d x ok, is the same as PDF p y of y dy. So, this PDF of y is called as p of y, this is subscript here is only to indicate that they are looking at the PDF of y that is what the subscript here indicates. So, this is basically this is basically saying that it is same as the probability of getting y in the neighborhood of y plus dy.

(Refer Slide Time: 36:35)

So, basically what we want to say is that whether you measure x and compute px dx its the same as p of y dy ok. And we want to chase this probability density, knowing the probability density of x and the relationship y equals to x square. So, this is what we have in our hand. So, you can actually write down this equation 1 has p of y equals to p of x dx by dy, where this quantity dx by dy is called as the Jacobean of the transformation. So, I am transforming the probability on x to the probabilities on y and this transformation is affected by a Jacobian, please note that I have kept mode sign for the Jacobian because I cannot have a negative Jacobian.

And, so one can compute the left hand side by computing the derivative dx by dy. So, you can take this I will call this as maybe 1 a and you can take the derivative d by dy of equation 1 a and what you get? You get one as twice x dx by dy ok, which basically means that dx by dy is 1 upon 2 x ok. So, if you plug dx by dy an equation 2 what do we get?

(Refer Slide Time: 39:05)
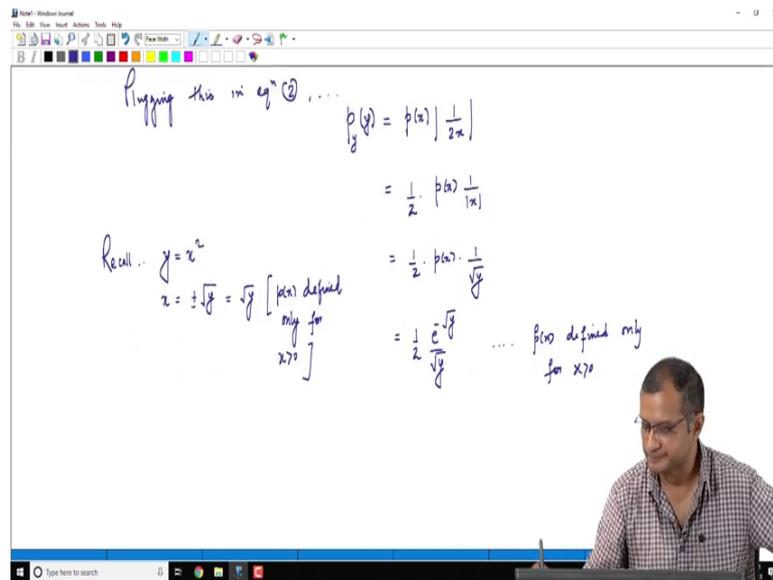
We get the PDF of y this is the function of y as PDF of x into mod of 1 upon 2 x, because that is what is the dx by dy, but it is nothing but 1 upon 2 into p of x into 1 upon mod x ok. Now recall that our relationship between y and x was the simple quadratic relationship. So, y was x square which means that I can write x as plus or minus square root of y.

But here I need mod x ok. So, I can write this as 1 upon 2 p of x into 1 upon square root of y because it does not matter whether you have take plus or minus mod of plus or minus is always plus and our original PDF was exponential. So, I am simply writing for p of x as e raise to minus of x and x is square root of y divided by square root of y. Now I cannot take minus x here for the simple reason that my PDF is not defined; since x is if you look our original PDF it is defined only for positive x ok. So, I will not take the negative values here as p x defined only for positive x.

So, I cannot take the negative square root of y here in the numerator you can write down here that p of x defined only for x greater than 0. So, you cannot take p of you cannot take a negative x in the definition of p of x ok. So, with that this definition of p of y emerges.

(Refer Slide Time: 41:47)
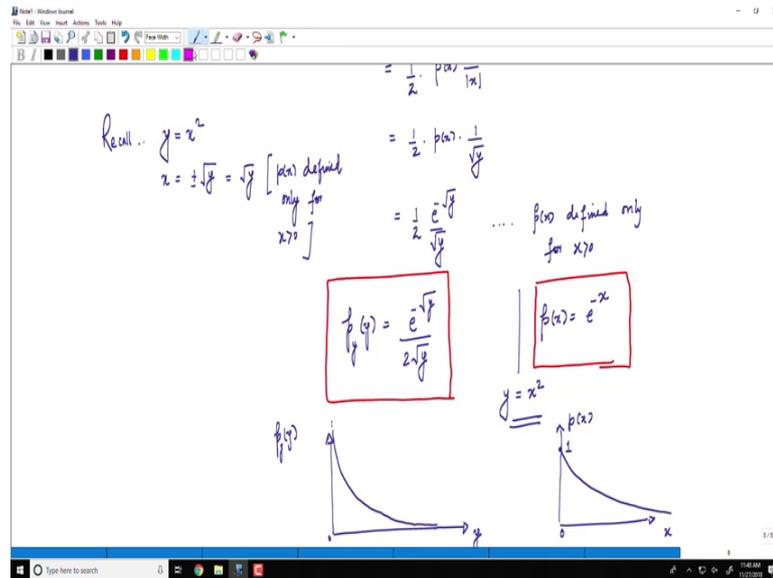
So, I can write down my the distribution of y that I was chasing as e raise to minus square root of y upon twice of square root of y, recall that our original distribution on x was e raise to minus x. So, if you are original distribution of e raise to minus x and your x and y are related by this expression, then this is the final form for the distribution of y and we have used the method of Jacobian's to map one distribution into another.

So let us plot this distribution also. So, our new distribution is basically this is my p of y that the function of y it has to be like this at y equals to 0 this will blow up, because there is a square root of y singularity here. But that is not to be that does not pose any concern to me because this singularity is integrable I will explain what is the meaning of this integral singularity let me first sketch this function here ok. And our p of x is the old function this is 1 0. So, this PDF that we obtain for y is also normalized and this is because this singularity at y equal to 0 is integrable.

So, you check if, these are the checks you must perform is normalized otherwise it is not a proper PDF ok. So, this check can be done by simply integrating this function in the domain of y. So, I will take this function and integrate in the domain of y domain of y is 0 to infinity and you can write this as 0 to infinity, full the form of p of y is E raise to minus square root y upon twice of square root y d y ok.

The singularity at y equal to 0 poses no issue and that can be seen by simply making a substitution of square root of y as the sum variable u thereby giving me one upon twice square root of y dy as d of u and the limits y going from 0 to infinity are now transform to my limits on u as 0 to infinity, limits do not change.

So, my integral becomes 0 to infinity e raise to minus u du which is the nothing but minus E raise to minus u in the limits 0 to infinity which is nothing but 1. So, as we can see the singularity was integrable and my PDF on y is also well behaved. So, you can take home conclusion is that the p of y is well behaved ok. So, that is how you map one distribution into another through the method of Jacobian's. And once you do this mapping, please check that your distribution is well behaved it has only axioms; you know it is obeying all the axiom that we discussed in the context of probabilities, that it is positive, it is norm conserving, these are the 2 properties that you should always check.

So, we come to an end here, and in the next class we may be discuss 1 or 2 more examples, and proceed towards the computation of characteristic functions which is very important method of computing moments or averages in any distribution. So, we will take it up from there in the next class.