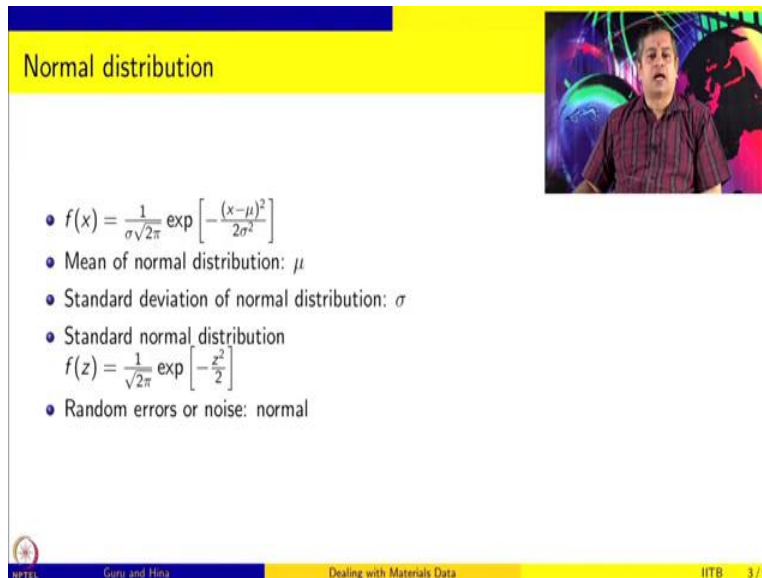**Dealing with Materials Data: Collection, Analysis and Interpretation**
Professor M.P. Guruajan,
Professor Hina A Gokhale,
Department of Metallurgical Engineering and Materials Science
Indian Institute of Technology, Bombay
Lecture 45
Normal distribution and error function

Welcome to Dealing with Materials Data. In this course, we learn about the Collection, Analysis and Interpretation of Data from Material Science and Engineering. We have been looking at dealing with probability distributions using R and we have discussed some district distributions and now we are looking at normal distribution and in this session we are going to learn about error function and how the normal distribution is related to the error function.
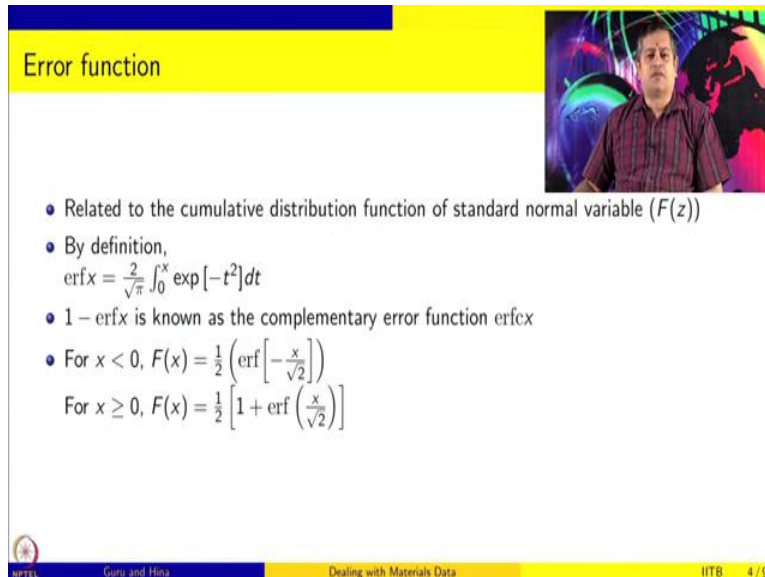
(Refer Slide Time: 00:49)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(x-\mu)^2}{2\sigma^2}]$$

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right]$$

So, we take x minus mu by sigma as z, and then we get the standard normal distribution, which has the normalization factor 1 by root 2 Pi, and it is just exponential minus z square by 2 and we

have also learned that random errors or noise always follows the normal distribution and we saw an example of electrical conductivity in electrolytic tough pitch copper and we saw that repeated measurements if you make because of random errors, you get normal distribution.

(Refer Slide Time: 01:59)

Now we want to learn about the error function and the error function is related to the cumulative distribution function of standard normal variable. So, capital F of z is related to the error function and by definition error function is 1 by root Pi minus x to plus x exponential minus t square Dt.

1 – erfx is known as the complementary error function erfcx

For x < 0, $F(x) = \frac{1}{2}\left(\text{erf}\left[-\frac{x}{\sqrt{2}}\right]\right)$

For x ≥ 0, $F(x) = \frac{1}{2}[1 + \text{erf}(x/\sqrt{(2)})]$

So, in other words, if we can take a standard normal variable and get cumulative distribution function, from that, we can actually calculate their function. For example, 2 times f of x minus 1 will give you the error function of x by 2. So, that is what we are going to look at and we will also learn how the error function is related to the diffusion problem.

(Refer Slide Time: 03:29)

To understand the diffusion problem, diffusion flux and concentration gradient are related by a constitutive law, and it is given as Fick's the first law of diffusion, it says that the diffusion flux is proportional to the concentration gradient and the proportionality constant is called diffusivity and the negative sign indicates that diffusional fluxes are such that concentration gradients will get even down over a long period of time.

$$J = -D\nabla C$$

$$\frac{\partial c}{\partial t} = -\nabla . J = \nabla . D\nabla c$$

$$\frac{\partial T}{\partial t} = \alpha \nabla^2 T$$

(Refer Slide Time: 05:50)

- Solution of Fick's second law for the semi-infinite bar, for the boundary conditions,
  $c(x = 0) = c_s$ and $c(x = \infty) = c_0$, is
  $c = c_s - (c_s - c_0)\mathrm{erf}\left[\frac{x}{2\sqrt{Dt}}\right]$
- Doping, carburisation, ...
- Porter and Easterling, Phase transformations in metals and alloys

$$c(x = 0) = c_s \ and \ c(x = \ \infty) = c_0$$

$$c = c_s - \left(c_s - c_0 \ \mathrm{erf}\left[\frac{x}{2\sqrt{Dt}}\right]\right)$$

And so this is nothing but surface concentration minus the difference between surface concentration and far field concentration error function of x by to root Dt and this solution of Fick's second law is relevant for cases such as doping and carbonization etc.

Because in these cases, you can think of a material which has some given amount of concentration and on the surface we start increasing the concentration. So, that there is diffusion that takes place into the material. So, as the time goes by the concentration keeps changing in the material. So, this is what is described by this equation and all this solutions and the next couple of solutions that I am going to show, they are all described Porter and Easterling Phase transformation in metals and alloys.

So, the next is also a semi infinite bar solution, but in this case, the solution is slightly different, because instead of surface concentration being a specific concentration, it is kept at 0, but the for fill concentration is at the same value c_0 let us say.

So, this is a relevant for things like decarburization, because if you have a material that has some amount of concentration of one of the species and at the surface, beyond the surface, you will see

that there is no concentration of this species. So, there is a concentration gradient and so this will start leaving the material and in that case, the solution is c_0 error function x by 2 root Dt.

(Refer Slide Time: 08:02)

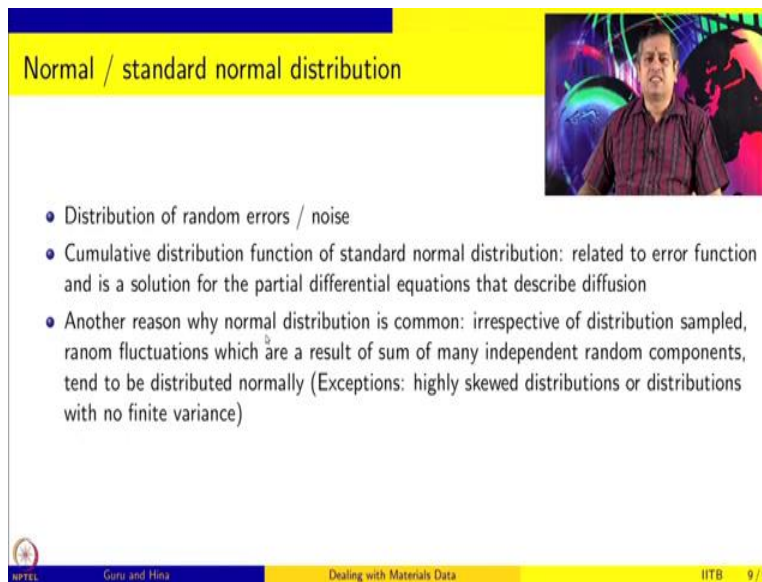

$$c(x = 0) = 0.5 * (c_1 = c_2), c(x = -\infty) = c_1, c(x = \infty) = c_2$$

$$c = \frac{c_1 + c_2}{2} - \frac{c_1 - c_2}{2} \text{erf}[x/2\sqrt{(Dt)}]$$

So, this is relevant for cases where if you take two different compositions of a particular material and weld them together as a function of time how the composition changes for example, you can also think of other cases, you can make c1 and c2 to be 0 and you can imagine putting a small amount of material at the center and allow it to diffuse for example, and that solution is relevant for radioactive tracers, and even for calculating diffusivity for example.

So, all these things are described in Porter and Easterling phase transformation in metals and alloys in brief. There are also other classic textbooks for diffusion, such as Schumann's diffusion textbook, there are also mathematical textbooks, which described the solution of differential equation for different boundary conditions. Because diffusion equation is a partial differential equation, the solution depends on what boundary conditions you assume, so depending on boundary condition, the solution will change.

So, in other words, the complete description of the partial differential equation includes the boundary condition. So, if you give the equation and give the relevant boundary conditions, then you can get solutions and they are different for different boundary conditions. So, here are the 3 cases where error function turns out to be the solution and as we have seen error function is related to the cumulative distribution function of the standard normal variable. So, you can now use R to plot the error function solution for these three scenarios and that is what we are going to do in this session.

(Refer Slide Time: 10:45)

And just to summarize, so we have been looking at normal or standard normal distribution, we have noticed that distribution of random errors or noise follows this distribution. We have also seen that the cumulative distribution function of standard normal distribution is related to error function which happens to be a solution for partial differential equations that describe diffusion and heat conduction and so on.

So, it is relevant from that point of view also and there is another reason why normal distribution is common. In fact, the word normal itself says that this is expected to be the most common distribution, irrespective of distribution from which you sample, if there are random fluctuations, which are result of many independent random components and they also tend to be distributed anomaly, there are exceptions, highly skewed distributions and distributions with no finite variants do not follow this.

But in general, you will find that normal is very common, and that is why in many, many things, we use normal distribution or standard normal distribution as a sort of benchmark and we described everything else with respect to the normal distribution. So, we will see a couple of examples of this and in many times we are actually interested in knowing if a particular data that we have received is now normal or not, or does it follow normal distribution or not?

How to look at those things etc will be things that we will look at in the following sessions. But in this session now, let us go back and look at the diffusion solution and try to use R to get the diffusion solution.

(Refer Slide Time: 12:30)

So, so as usual, we will get R, so this is version 3.6.1 action of the toes is what it is called and we need to look at the working directory, we need to make sure that we are at the right directory and the first thing to do is to actually plot the error function. So, let us do this. So, so we are going to plot error function from between minus 5 and plus 5 and what we are going to plot this. So, you can see that P norms at 01 is nothing but, the cumulative distribution function of the standard normal variable two times that minus 1 is actually going to describe the error function that is what we have seen and that is what we are going to use here.

So, if you do this, you will see that the function as it goes to minus infinity goes to minus 1 as it goes to plus infinity plus 2 plus 1. That is because error function of minus x is nothing but minus of error function x, so it is about 0, it is just a negative of this. So, that is why the function looks like this. So, if you go to Porter and Easterling, for example, you will see there are functional plotted and here is the plot that gives the same thing for us.

(Refer Slide Time: 14:00)


```
cs = 1.4
c0 = 0.1
D  = 4e-11
x <- seq(0,500e-6,1e-6)
t = 10
Z1 <- x/(2*sqrt(D*t))
Z11 <- 2*pnorm(Z1,0,1)-1
plot(x,cs-(cs-c0)*Z11)
t = 50
Z2 <- x/(2*sqrt(D*t))
Z22 <-2*pnorm(Z2,0,1)-1
points(x,cs-(cs-c0)*Z22,col="blue")
t = 100
Z3 <- x/(2*sqrt(D*t))
```



```
points(x,cs-(cs-c0)*Z22,col="blue")
t = 100
Z3 <- x/(2*sqrt(D*t))
```

```
Z33 <-2*pnorm(Z3,0,1)-1
points(x,cs-(cs-c0)*Z33,col="red")
abline(v=0)
abline(h=0.9)
abline(v=sqrt(D*10))
abline(v=sqrt(D*50))
abline(v=sqrt(D*100))
```
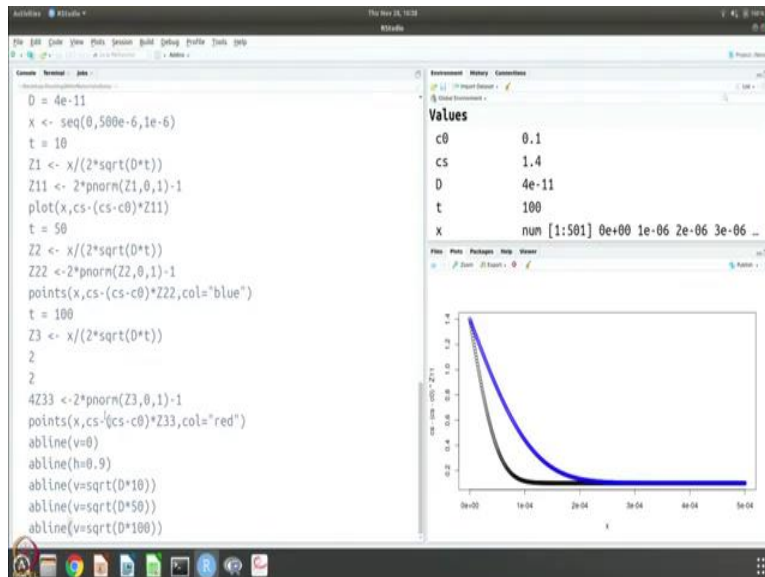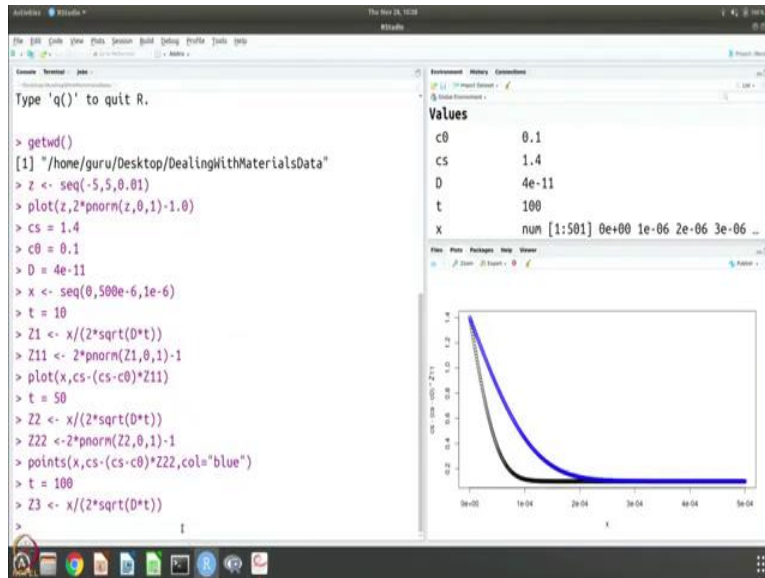
So, the next step is to get the semi infinite bar solution and for that, we have to achieve certain coefficients for diffusion and so on and so let me first get the full code to take this. So, what we are trying to do in this is to actually reproduce some of the figures that are there in Porter and Easterling. So, let us assume that the surface concentration is 1.4.

Let us assume that the surface concentration is 1.4 and far field composition is 0.1 and diffusivity is 4 to power minus 11 meters square per second and we are going to look at some 0.5 millimeter distance to which we are going to starting from surface some 0.5 millimeter is the distance to which we are going to look at diffusion and we are going to consider time of 10 seconds, 50 seconds and 100 seconds.

So, if you look at for these numbers, Porter and Easterling tells you that is about 1000 seconds is what it will take for the competition to penetrate the entire 2 millimeter distance for example, but we are going to plot the early stages at 10, 50 and 100 seconds. So, that z1 is x by 2 square root or Dt and this is the parameter that goes into the error function solution that we saw. So, so the solution itself is given in terms of error function.

So, this is a parameter for error function. So, for that parameter, we have to calculate the cumulative distribution function from the standard normal distribution multiplied by 2, subtract 1. So, that is the error function actually. So, this is a parameter of their function, because the T is changing for every x.

So, we need to calculate that parameter from that parameter for the given x and t, we can find the solution and x a sequence, so it goes from 0 to 0.5 millimeter and then for time t equal to 10 seconds for example, so we are getting the solution and parameter and from the parameter we are getting the solution and we are going to plot and remember the solution had the surface concentration minus the difference between surface and for fill concentration multiplied by the error function solution.

So, this is the first solution for 10 seconds. Similar things for 50 seconds and 100 seconds and so they are 50 seconds and 100 seconds solutions are going to be marked in blue and red. So, after we do that, I also want to plot certain lines. So, this is the 0, that is the surface we are going to plot. I am going to also draw a line at root Dt, different lines and I am going to show you. So, if you look at the figure Porter and Easterling it shows that the distance from the surface into the material is actually proportional to square root Dt. So, so this, this from the surface, the distance to which penetration happens is actually so that is what this line horizontal line is good to show us.

(Refer Slide Time: 17:48)
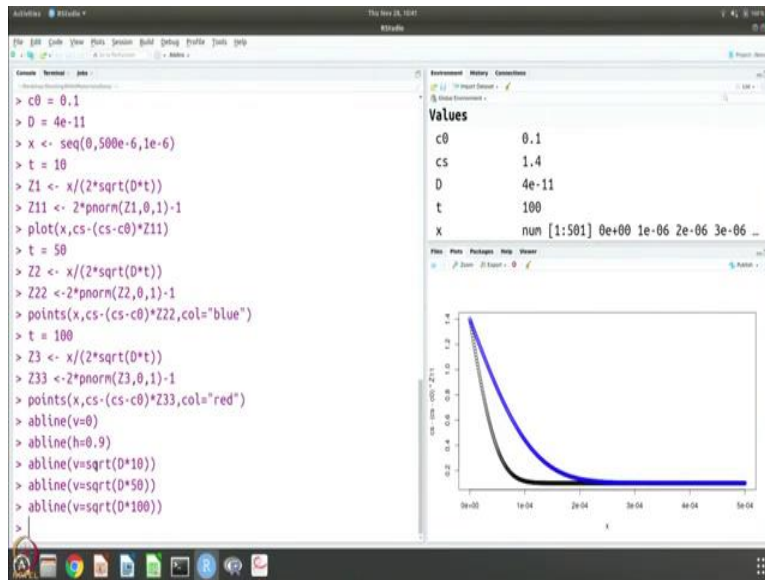
So, here is the solution. So this is the surface and this is the 10 second solution. So, you can see that this is proportional to root Dt because this is the root Dt value that we have plotted for 10 seconds and this is 50, this is for 100 and you can see as the time proceeds the diffusion, the flux increases into the material because there is a concentration gradient this is 1.4 this 0.1, so the material will keep accumulating in the system, so that the concentration gradients are getting even out.

So, that is where we are moving towards and that is what is shown and this is a figure that is there in Porter and Easterling, for example. So, you can generate the same solution using R of course, we can also generate the other solutions and the second solution is of course, we want to keep the surface concentration to be 0 and we want to look at far field composition to be 0.1, and then we want to look at what happens to the solution.

(Refer Slide Time: 18:57)
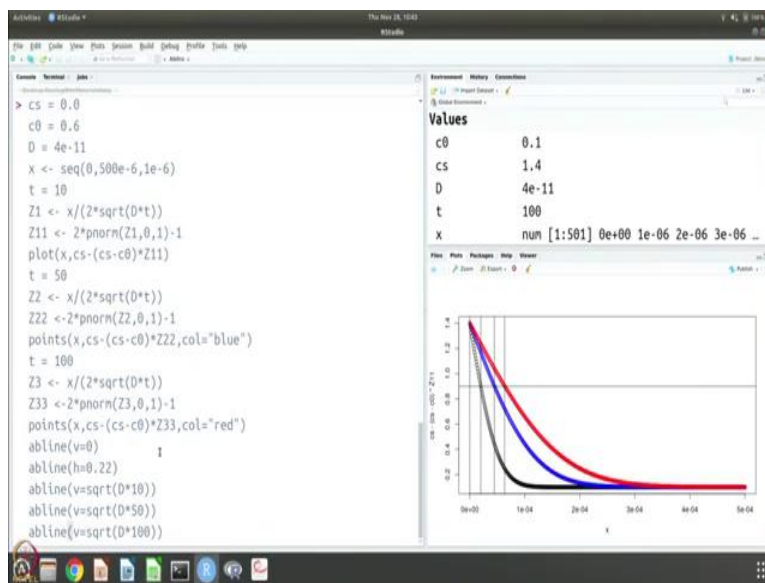
So, let us do that and so, we are doing the same thing. Surface concentration is 0, for fill composition let us say is 0.6 and same diffusivity and we are again looking at some 0.5 millimeter and we take 10 seconds 50 seconds and 100 seconds and we are going to use the same formula because when you substitute cs to be 0, it reduces to c_0 times z1, 1, the error function solution so, that is what we saw.

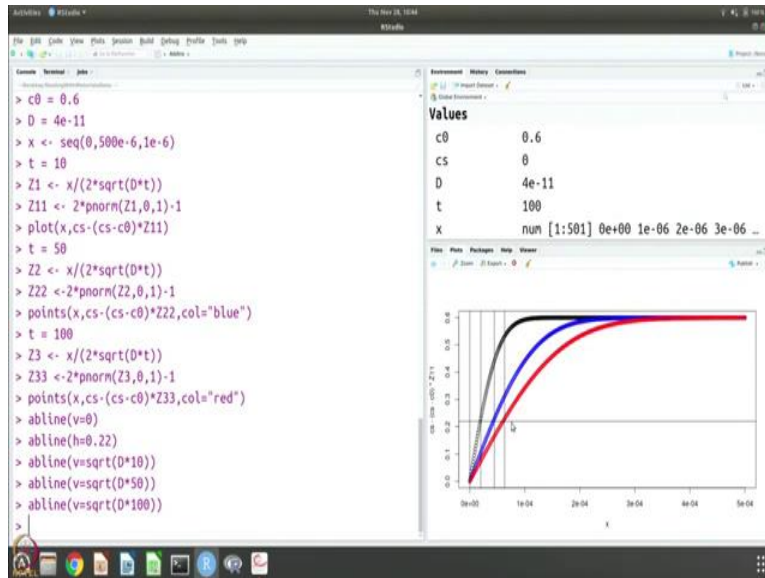So, because this is also a semi infinite bar solution, so, it should be the same solution and again we have drawn certain vertical and horizontal lines to show what is happening to the solution. So, if you plot, so in this case, the material has very high concentration 0.6 at the surface, the concentration is 0.

So, obviously, there is concentration gradient between what exists inside the material and at the surface somatic, the concentration starts changing these carbon atoms start living from the surface this is known as decarburization, and that leads to this kind of profiles and as you can see again here also we see that as time goes on the extent to which decarburization takes place.

The concentration for example, in the bulk is falling and how much does it fall depends on the square root Dt, which is the diffusion distance that is why we call it diffusion distance because, dimension wise it is the distance and it basically tells you for a given time for a given diffusivity how much will be the depth to which considerable diffusion would have taken place. So, that is what is shown by this line to indicate that as time goes by 10, 50, 100 seconds the depth also keeps increasing to which decarburization takes place.

(Refer Slide Time: 21:04)
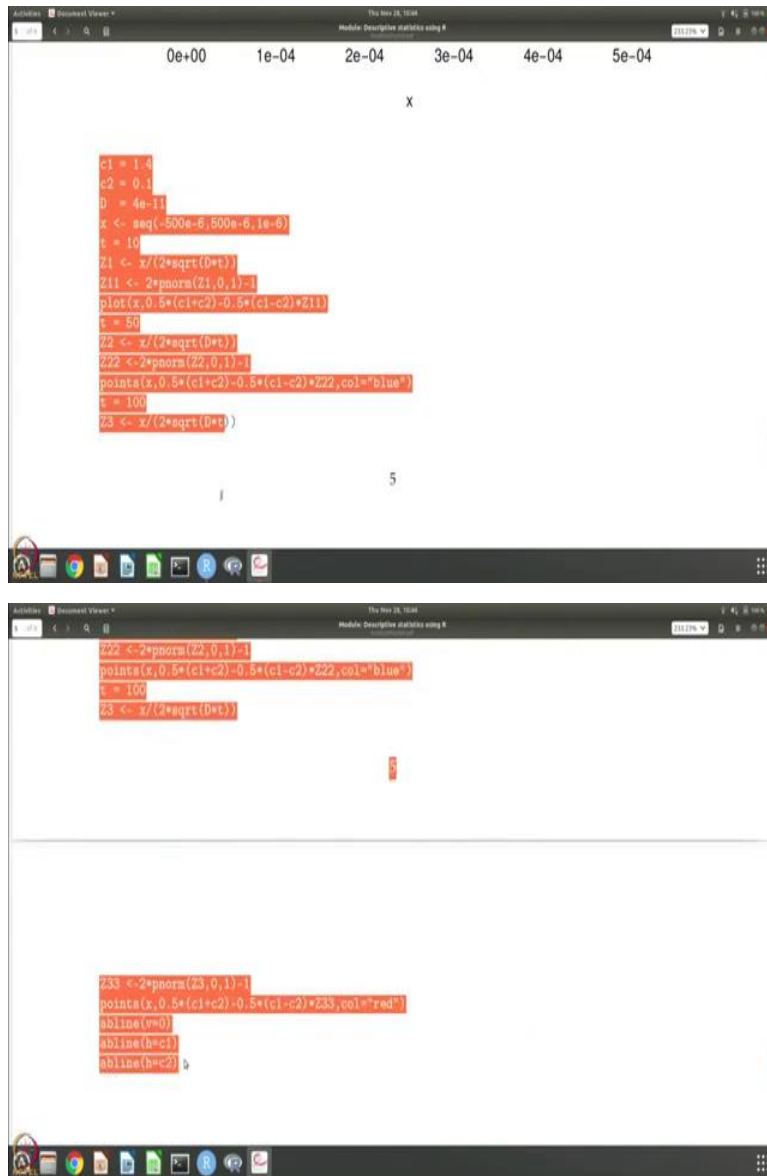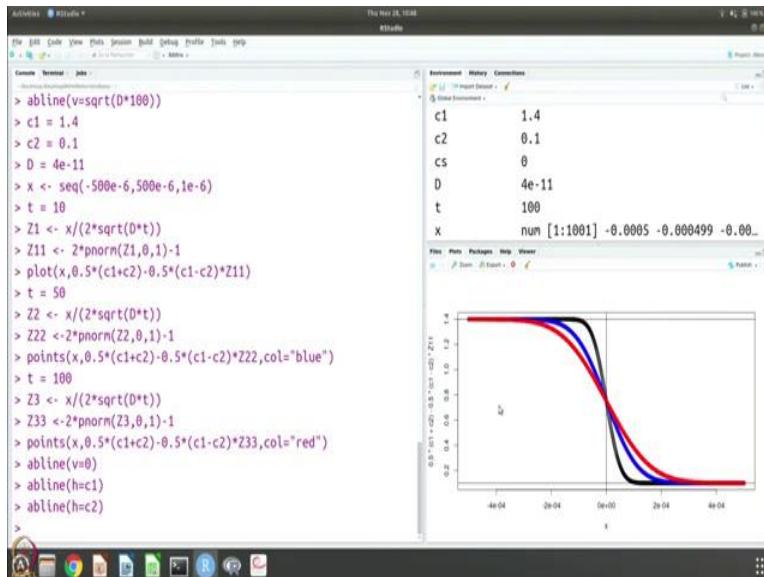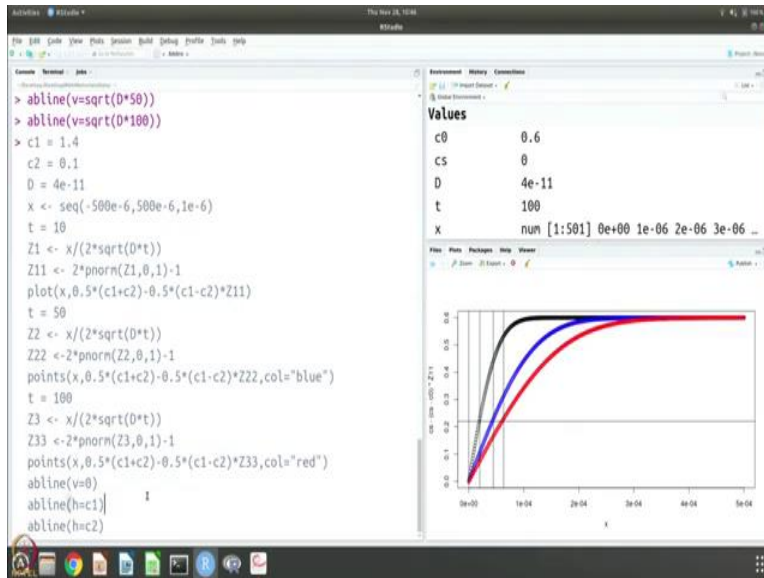
Of course, the last solution that we want to look at is the two semi-infinite bars put together with far fied compositions, which are given by c1and c2. So, that is what we have done here. So, let us assume again that c1 is 1.4, c2 is 0.1. So, we are taking two steels which compositions, with composition 1.4 on 1 side, and 0.1 on other side and we are putting them together and diffusivity is the same for 4e power minus 11 meters square per second and we are going to look at so now we the origin is at where we have put this material together.

So on the minus you will see the material with the composition 1.4 on the positive side you will see composition a material composition 0.1 and at 0, so it will be an average of these two composition.

So, it will be 1.5 divided by 2, which is like 0.75 and again we are going to look at solutions for time 10, 50 and 100 and for solution you just have to evaluate this parameter which is x by 2 square root Dt and for that parameter you have to evaluate the error function error function is given by the cumulative distribution function of the standard normal distribution two times that to minus 1 actually is there a functional solution and remember error function is an integral and the quantity that is inside the integral gets integrated out. So, it is the boundary condition.

So, the solution is 0 to x by 2 square root Dt. So, that is what gets substituted and that is why it is a function of this value and that is what is happening here and then we want to plot the solution. So, it is 0.5 c1 plus c2 minus 0.5 c1 minus c2 times error function. So, that is the solution and this is for 10 seconds 50 seconds 100 seconds. As usual we draw some lines to show for example, where the interfaces, where these two things are put together and what is the far field composition on either side c1and c2 for the left and right respectively.

So, let us do this. So, this figure is also there in Porter and Easterling. So, you see that, we took a material with 1.4 and we took another material with 0.1. So, initially this was the composition profile and as time goes by in this the composition keeps decreasing in this the composition keeps increasing. So, the area under this part should be exactly equal to the area and this part by mass conservation because the carbon that is leaving from here is entering here and that this depth of penetration of the diffusion curves is proportional to root Dt.

So, as you can see 10, 50, 100 etc, you will see that it is penetrating more and more, in this case it is losing carbon in this case it is gaining carbon and of course, the after an infinite amount of time you expect that everything becomes 0.75. So, that will be the composition that the system is trying to reach at the end after a infinitely long time, that is the composition that is going to take place.

But as you can see, as the time proceeds, you will see that for example, in 10 seconds, how much diffusion took place and the 50 and 50 to 100. It is less because the concentration gradients that are coming down as diffusion progresses, which means that it will take longer and longer to achieve the same amount of diffusion, so things also gets slowed down compared to the initial stages. So, it will take really long time for composition to reach the uniform value everywhere.

So, we have looked at three different solutions for diffusion equation and all these solutions are based on the cumulative standard normal distribution function and it is related to the error function

to this formula two times cumulative distribution function minus 1 is basically the error function and that happens to be the solution for the diffusion equation and you can get this solution using R because R can calculate this cumulative distribution function.

Typically, if you look at textbooks, like Robin or Porter and Easterling, you will find that they list out error function solution for different values. So, you have to manually calculate and for values in between you have to either do interpolation or find tables that will calculate it for you or right programs that will do it for you. But of course with R you can just call this pnorm and get the solutions.

So, to summarize we have looked at normal distribution, it is a continuous distribution and random errors give rise to this distribution irrespective of from which distribution is sample if there are random errors, they also follow this distribution. So, it is very, very common to see this.

It also has relevance to material science and engineering specifically, because the solution for the diffusion equation which happens to be one of the important mechanisms by which atomic movement happens in materials especially in solids, which leads to lots of phase transformations and micro structural changes, which are very, very important from an engineering point of view because all of the heat treatment is basically based on phase transformation and how fast or slow this phase transformations take place, namely diffusion.

So it is very, very important to have an idea about the solution for the diffusion and it so happens that this normal distribution function has a relationship to the solution of the diffusion equation for certain boundary conditions and that is what we have explored in this session we will continue looking at normal distribution a little bit more before we move on to the other probability distributions, thank you.