

# Math 1180:Lab4

Kyle Gaffney

January 29th, 2014

## 1 Probability Density Function (PDF)

A function  $f(x)$ ,  $x \in (a, b)$  is a PDF if it satisfies that  $f(x) \geq 0$  for every  $x \in (a, b)$  and if  $\int_a^b f(x)dx = 1$ . In other words the probabilities must be positive and must sum to 1.

### 1.1 Example

Suppose that a molecule is diffusing from a cell according the following PDF,  $f(t) = 0.1 * e^{-0.1t}$ .

First check that this function is a PDF. How would you do this?

What is the probability that the molecule has left after 1 second?

$$P(0 \leq \text{Departuretime} \leq 1) = \int_0^1 f(t)dt = \int_0^1 0.1 * e^{-0.1t} dt = -e^{-0.1} + e^{-0} = .095$$

So if we were to run this experiment 1000 times on average 95 molecules would have left the cell within 1 second.

### 1.2 Simulating in R

This type of PDF is related to the Exponential Distribution which you will learn more about later. To simulate times of these molecules leaving the cell we can select random numbers drawn from the exponential distribution with rate constant  $k$ , where the corresponding PDF is  $f(x) = ke^{-kt}$ . To do this in R use the following command:

```
ltimes=rexp(n=100000,rate=.1)
```

As this is a massive list of randomly simulated leaving times the following command will let us view the top line of the stored list to get a feel for what the rexp command did.

```
head(ltimes)
```

Additionally we can collect further statistics from the list with the command "summary"

```
summary(ltimes)
```

The easiest way to visualize all of this data is with a histogram. To do this in R we can use the following command:

```
hist(ltimes,xlab="Leaving Time (Minutes)",main="Leaving Time",freq=F)
```

In R we can get the corresponding PDF for any Distribution by replacing the r with a d. For example for the exponential distribution command, rexp, we would use the command dexp to get the density function.

```
tlist<-seq(0,100,.1)
lines(tlist,dexp(tlist,rate=.1),col="red")
```

Here we used tlist going from 0 to 80 because the maximum time that we observed should be somewhere near to 80. The lines command placed the PDF on top of the randomly generated data in the Histogram. Were they similar? Should we expect them to be?

What happens if we increase the number of boxes in our histogram? What does this show? Does the PDF still lie over the peaks of the boxes?

```
hist(ltimes,xlab="Leaving Time (Minutes)",main="Leaving Time",freq=F,breaks=500)
lines(tlist,dexp(tlist,rate=.1),col="red")
```

## 2 Cumulative Distribution Function (CDF)

The Cumulative Distribution Function is related to the PDF through integration.  $F(x) = \int_a^x f(y)dy$  is the definition of the CDF. What can we say about the CDF?

1. The CDF is always positive and always increasing. Why?
2. The CDF will approach 1 as x approaches infinity (or b if the domain of the PDF is finite (a,b))

Getting the CDF for our example problem:

$$F(x) = \int_0^x f(y)dy = \int_0^x 0.1e^{-0.1y}dy = -e^{-0.1y} \Big|_{y=0}^{y=x} = -e^{-.1x} + 1$$

Let us generate a histogram of the CDF for the data that we created. We can do this with the following code:

```

accumulation<-function(t){
N=length(ltimes)
length(ltimes[ltimes<t])/N}

tlist2=seq(0,100,length=25)
approxcdf=seq(1,25)
for(i in seq(1,25)){approxcdf[i]<-accumulation(tlist2
[i])}

```

Now let us plot this approximate CDF that we get from the data:

```
plot(tlist2,approxcdf)
```

Finally we can add the actual CDF that we found from integrating the PDF:

```

actualcdf<-function(t){1-exp(-0.1*t)}
t<-seq(0,100,.01)
lines(t,actualcdf(t),col='red')

```

### 3 Assignment for the Week

1. Suppose that we consider a smaller molecule which has a higher rate of leaving the cell,  $k=2$ , such that our PDF is  $f(t) = 2e^{-2t}$ . First prove that this function is in fact a probability density function (check it against the 2 conditions for a PDF). Then by altering the above commands plot a histogram of generated data from the exponential distribution and the PDF line on the same plot. Finally alter the commands above to create an approximate CDF from the data and compare it to the actual CDF (which you can solve for by hand).

HINT: The 80 which appeared in the example problem was selected because it was near the maximal time that it took a molecule to leave the cell when the rate was  $k=0.1$ . This will have to be changed now that the rate is a lot different.

2. Suppose that now the PDF is function which has no anti-derivative.  $f(t) = \frac{1}{2.34}e^{\cos(t)}$ ,  $t = (0, 1)$ . Now find the probability that the molecule has left the cell between  $t=(0.4,0.6)$ .

HINT: As this function has no antiderivative you will need to use numerical integration (see lab 10 from last semester).