MAS 6024, Assignment 2
Registration Number 170224545

Part 1

(i). Write a function using the ecdf of a given set of realisations and for general, possibly non-finite, values of c and d. return a numerical approximation. The R code is below:

```
x <- rt(10, 2)
fx <- function(c, d){
  f <-ecdf(x)
  e <- f(d) - f(c)
  return (e)
}
```

(ii). Firstly, we sample 10 elements x from distribution t with 2 degrees of freedom. The R code is below:

```
x <- rt(10, 2)
fx <- function(c, d){
  f <-ecdf(x)
  e <- f(d) - f(c)
  return (e)
}
```

When c = -1, d =1

```
fx(-1,1)

## [1] 0.3
```

When c = 4, d= 6

```
fx(4,6)

## [1] 0
```

Then, we sample 10000 elements x from distribution t with 2 degrees of freedom.

```
x <- rt(10000, 2)
fx <- function(c, d){
  f <-ecdf(x)
  e <- f(d) - f(c)
  return (e)
}
```

When c =-1, d =1

```
fx(-1,1)

## [1] 0.5764

## [1] 0.5764
```

When c =4, d=6

```
fx(4,6)

## [1] 0.0142
```

```
y1 <- pt(1,2) - pt(-1, 2)
y1

## [1] 0.5773503
```

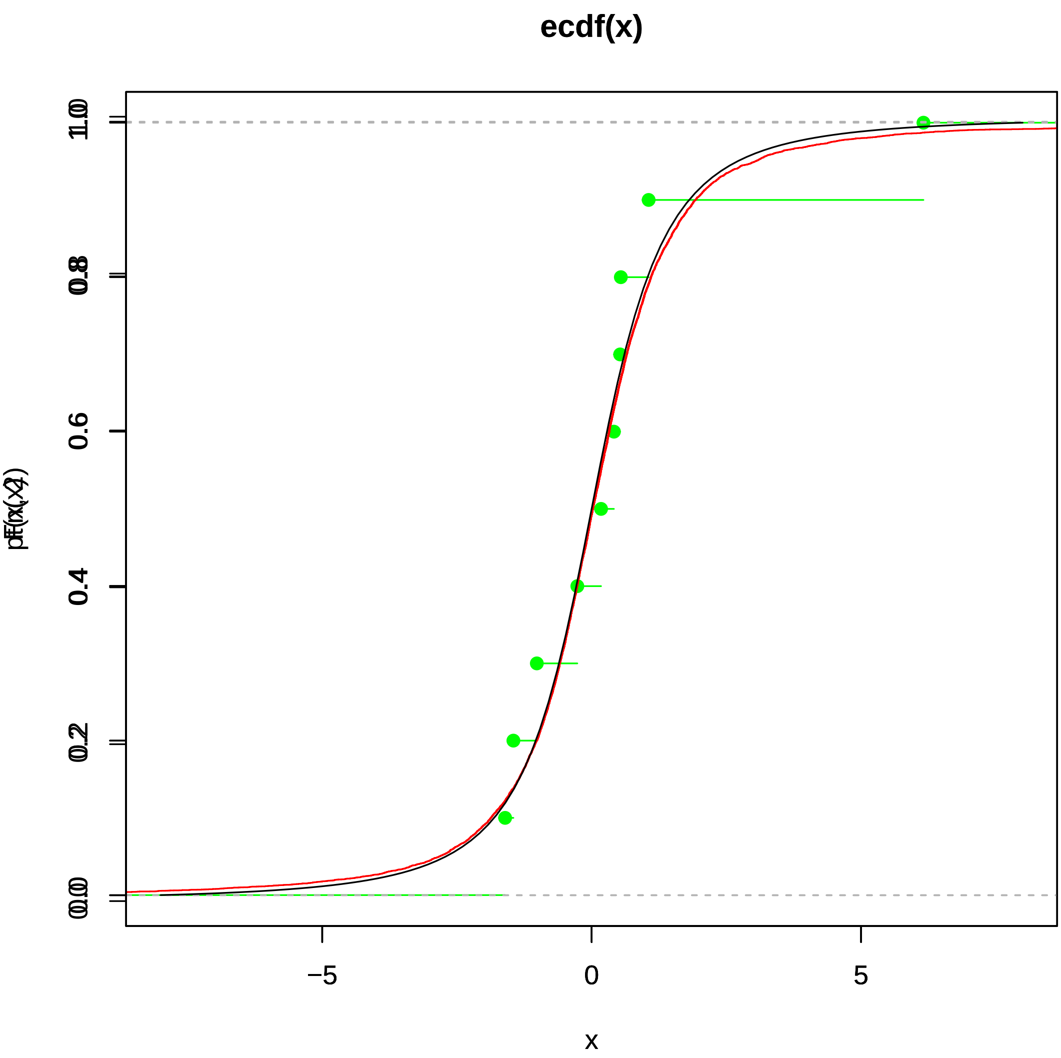
```
y2 <- pt(6,2) - pt(4,2)
y2

## [1] 0.01525974
```

As we can see , when number of realisations simulated increases, The approximations is closer to the exact probability value.

(iii). The black line is the real value of t distribution, the green line is the edcf figures with 10 realisations and the red line is the edcf figures with 10000 realisations.

```
x <- rt(10,2)
plot(ecdf(x),xlim=c(-8,8),col="green")
par(new=TRUE)
x <- (rt(10000,2))
plot(ecdf(x),xlim=c(-8,8),col="red")
par(new=TRUE)
curve(pt(x,2),xlim=c(-8,8))
```



Part 2

(i). Assume the probability of defective lea???et is p,

$$p = \frac{a_1 + a_2}{n}$$

$$a_1 = n_1 * p_1$$

$$a_2 = n_2 * p_2$$

$$n_1 = \frac{N_1}{N_1 + N_2} * n$$

$$n_2 = \frac{N_2}{N_1 + N_2} * n$$

$$P_1 = C_{n_1}^{x_1} p_1^{x_1} (1 - p_1)^{n_1 - x_1}$$

$$P_2 = C_{n_2}^{x_2} p_2^{x_2} (1 - p_2)^{n_2 - x_2}$$

(ii).

$$P = \frac{N_1 p_1 + N_2 p_2}{N_1 + N_2} = \frac{10000 * 0.05 + 20000 * 0.02}{10000 + 20000} = 0.03$$

When $p_1 = 0.05$, $p_2 = 0.02$, $N_1 = 10000$ and $N_2 = 20000$, we can say that the proportion is 0.03.

Because,

$$a = NP = 200 * 0.03 = 6,$$

We use 'a1' refer to defective leaflet prodeced by the first device, 'b1' refer to good leaflet prodeced by the first device, 'a2' refer to defective leaflet prodeced by the second device, 'b1' refer to good leaflet prodeced by the second device, we can have the R code:

```
x =c('a1','b1')
p_1 <- sample(x,size = 10000,TRUE,prob=c(0.05,0.95))
x =c('a2','b2')
p_2 <- sample(x,size = 20000,TRUE,prob=c(0.02,0.98))
x =c(p_1 , p_2)
p_3 <-sample(x ,200)

e = 0
for(i in 1:200){
  if (p_3[i] == 'a1'){
    e = e+1}}

f = 0
for(i in 1:200){
  if (p_3[i] == 'b1'){
    f = f+1}}

g = 0
for(i in 1:200){
  if (p_3[i] == 'a2'){
    g = g+1}}


h = 0
for(i in 1:200){
  if (p_3[i] == 'b2'){
    h = h+1}}
e

## [1] 3

f

## [1] 61

g

## [1] 2

h
```

```
## [1] 134

e+g

## [1] 5
```

The value of 'e+g' is the figure of defective leaflets in 200 sample.
If we run more times , the figure of defective leaflets will be closer to 6.
It means that When the sample size is 200, the will be 6 defective leaflets.