ECON 403A HMWK1

Pujan Thakrar

October 1, 2018

Qustion 1: *The Birthday Problem*

suppose there are n=25 people in a room Assume the following:

1. Ignore leap years and assumer there are only 365 days in a year.
2. Births are uniformly distributed throughout the year.
3. The people in the room are randomly distributed throughout the year.

i. What is the probability that two or more of them have the same birthday? Solve this analytically.

Prob(at least 2 people have the same birthday) = 1 - Prob(no one has same birthday)

Prob(no one has same birthday) $= \frac{365}{365} \times \frac{364}{365} \times \frac{363}{365} \times \ldots \times \frac{341}{365}$

Prob(no one has same birthday) $= \frac{365!}{340!} \times \frac{1}{365^{25}} \approx 0.4313$

Prob(at least 2 people have the same birthday) = 1 - Prob(no one has same birthday)

Prob(at least 2 people have the same birthday) $\approx$ 1-0.4313 $\approx$ 0.5687 $\approx$ 57%

ii. Solive [i] computationally in R using the **prod** function. Next, to better understand the relationship between n and the probability p, plot p against n by looping through rooms of size 1 to 60.

```r
n=25                                                     #set size of room
y=365^n
x=prod((365-(n-1)):365)/y     #probability that no one has same birthday
p=1-x           #probability that at least two people have same birthday
print(p)
```

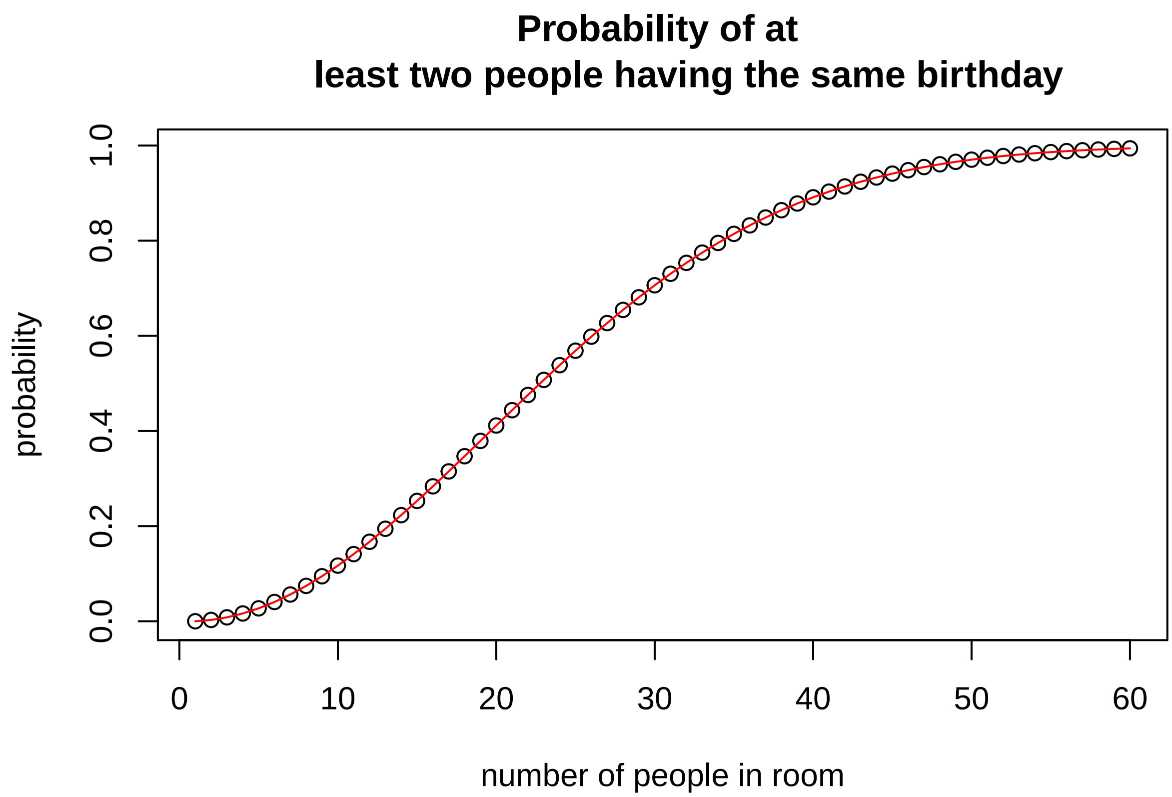
```
## [1] 0.5686997
```

Plot probability against the room size

```r
vec<-0

for(i in 1:60){
  n=i
  y=365^n
  x=prod((365-(n-1)):365)/y
  p=1-x
  vec[i]=p                              #store probabilities in a vector
}

plot(1:60,vec,xlab="number of people in room",ylab="probability", main="Probability of at
     least two people having the same birthday")
lines(1:60,vec,col="red")
```



iii. Based on your results from [ii], what is the minimum number of people in a room such that the probability of a match is greather than or equal to 50%?

```
which(vec>=0.5)
```

```
##  [1] 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45
## [24] 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
```

From this result, we can see that there needs to be at least 23 people in the room for there to be a probability of greater than or equal to 50%.

Qustion 2: *The Birthday Problem -Again*

Based on the same assumptions of Problem 1, write a simulation in R that generates 100,000 simulated rooms of 25 people, and plot a histogram of the density vs. the number of birthday matches. The histogram represents your approximation of the distribution of birthday matches. Do your results agree with Problem 1? *Hint:* Consider usint the R functions **sample** and **unique** to make the calculations easier.

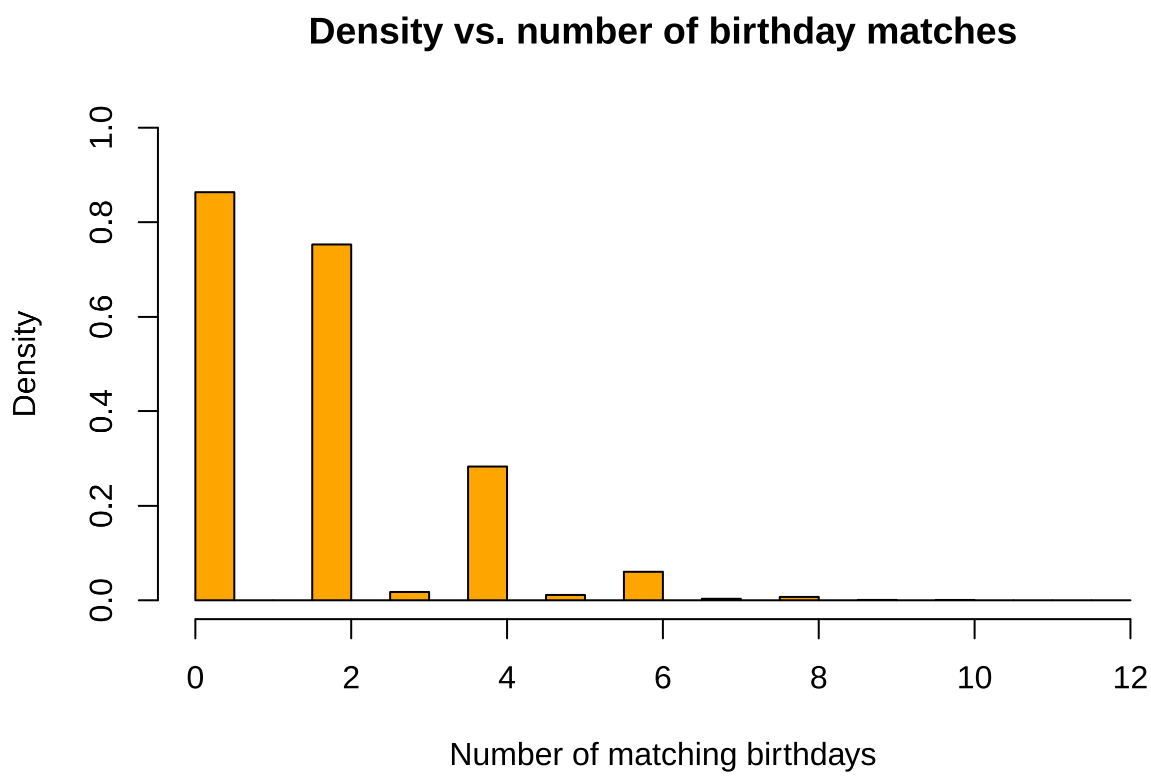
```
match=0                                            #set if it is a match
num_match=0                         #set the number of matches that occur
people=25
for(j in 1:100000){
  room<-sample(1:365,people,replace=TRUE)
  if(length(unique(room))<people){
    match=match+1                               #add 1 is there is a match
    num_match[j]=length(room[room %in% room[duplicated(room)]])
  }else{
    num_match[j]=0                    #indicate that there were no matches
  }
```

```
}

result = (match/100000)        #find the probability over all simulations
print(result)
```

```
## [1] 0.56827
```

```
hist(num_match,freq=FALSE, ylab="Density", xlab="Number of matching birthdays",
     main="Density vs. number of birthday matches", ylim=c(0:1), col="orange")
```



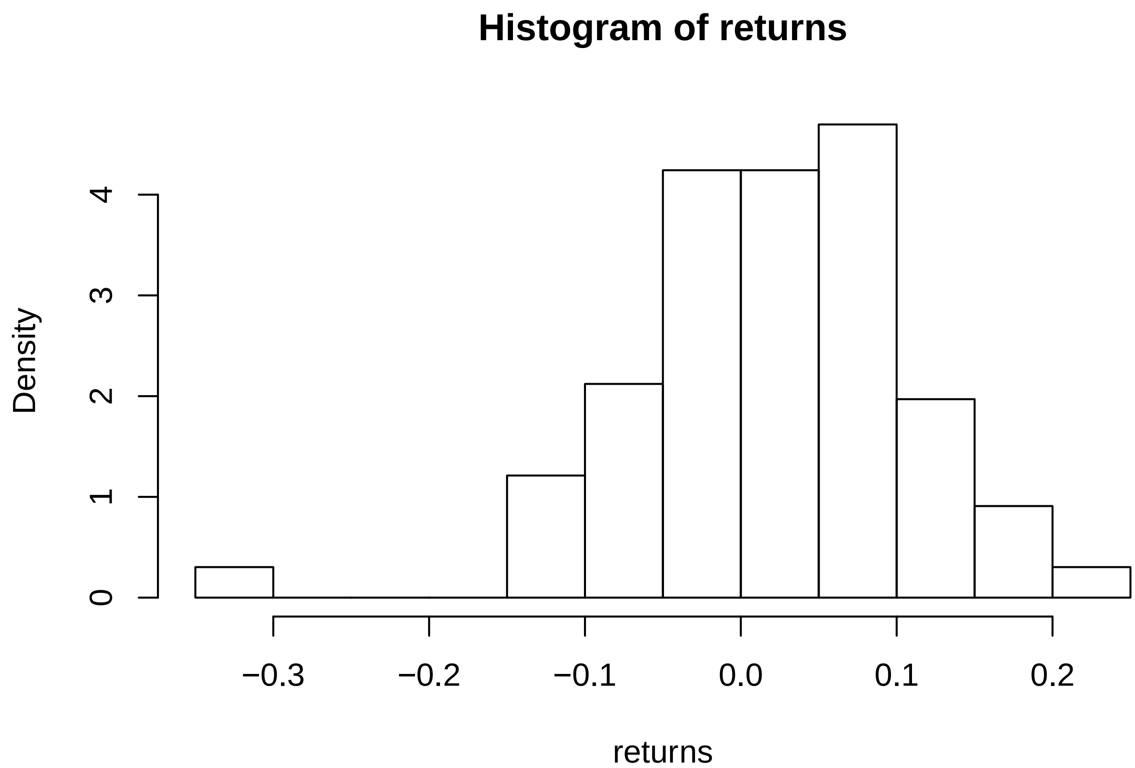
Since my answer came to be approximately 57% again, my results do agree with my results from Question 1.

Question 3

R Exercise: Plot a histogram of monthly returns of your favourite stock (use at leaset 10 years of data), and fit an appropriate distribution. Comment on your fit. Use this distribution to compute P(r>10%), where r=return.

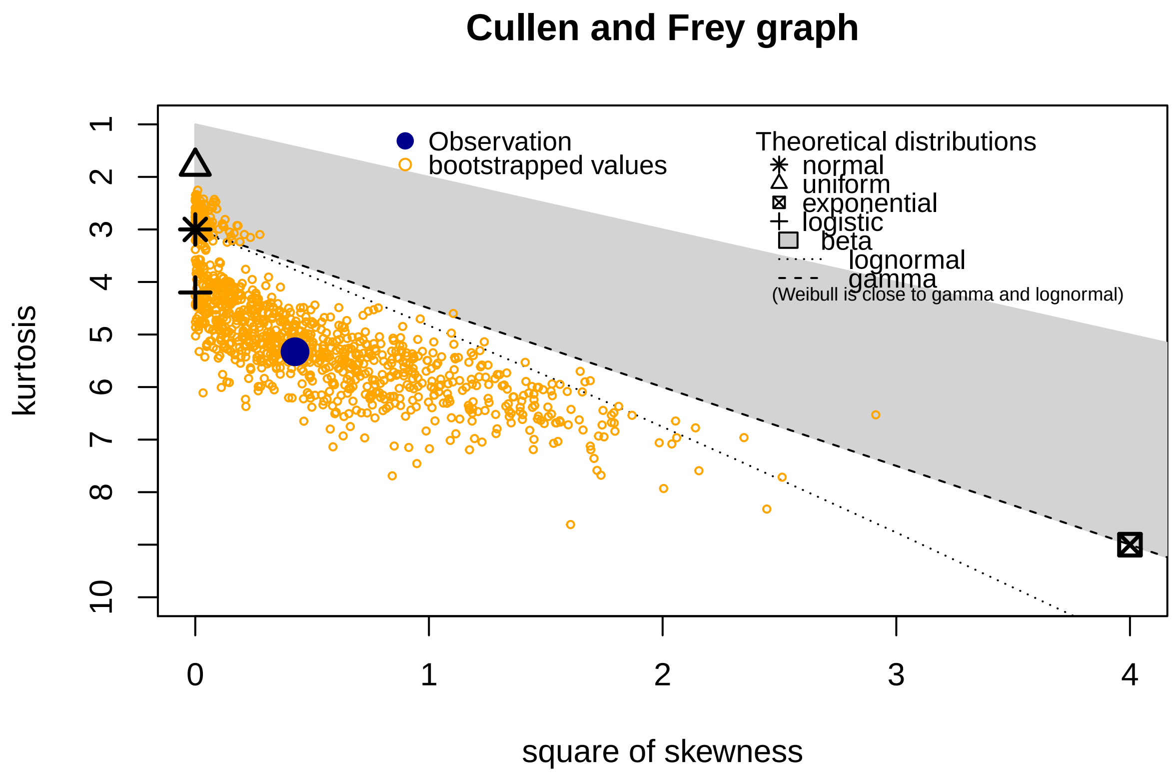
Taking stock data from APPLE between September 2007 to September2018, I get the monthly returns and plot the following histogram.

```
returns<-monthlyReturn(AAPL)                          #get monthly returns
hist(returns, prob=TRUE)                                   #make histogram
```



In order to determine which distribution method will have the best fit, I use the bootstrapping method and achieve the following Cullen Frey graph.

```
returns<-as.vector(returns)
descdist(returns, boot=1000)
```

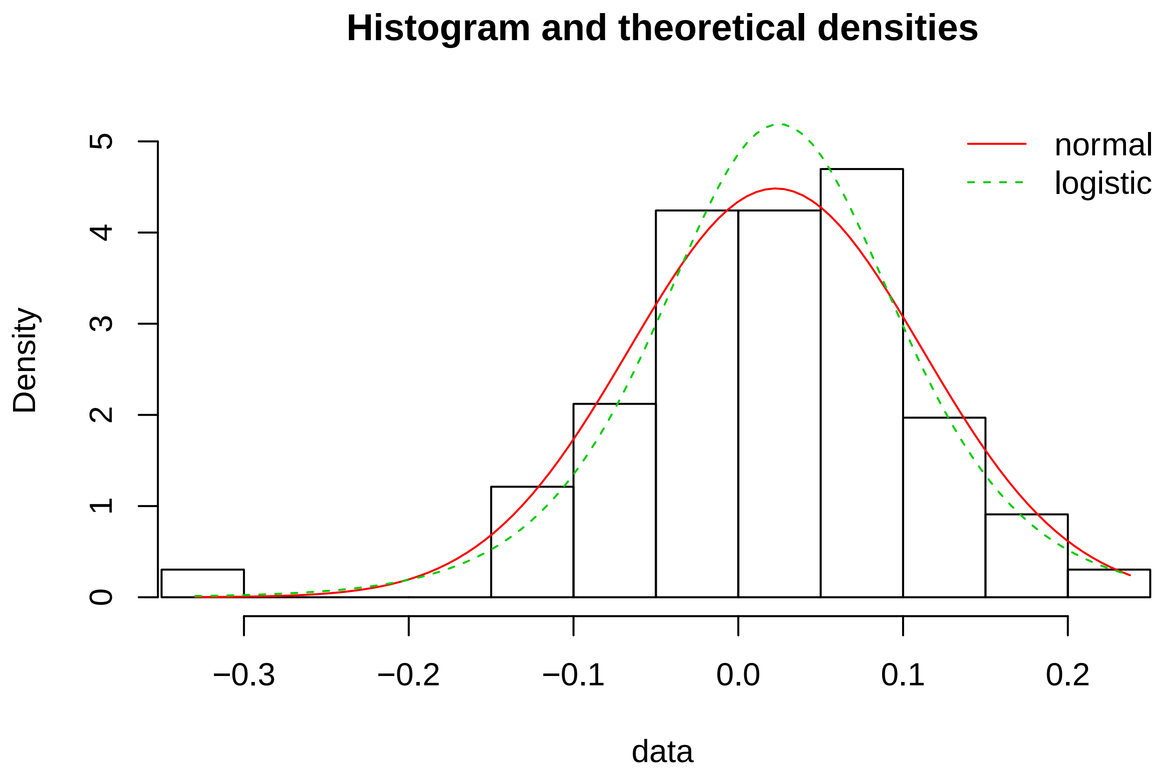



```
## summary statistics
## ------
## min:  -0.3295582   max:  0.2377012
## median:  0.02520135
## mean:  0.02269983
## estimated sd:  0.08931283
## estimated skewness:  -0.6534544
## estimated kurtosis:  5.32779
```

Given this information I believe the best fit in the logistic distribution. As we see below, the fit is much better in comparison to the normal distribution. Its higher peak accommodates more data points.

```
fn <- fitdist(returns, "norm")
fl <- fitdist(returns, "logis")

plot.legend <- c("normal", "logistic")
denscomp(list(fn, fl), legendtext = plot.legend)
```



To compute P(r>10%), I get the following result:

```
plogis(0.1, location=mean(returns),scale=sd(returns), lower.tail = FALSE)
```

```
## [1] 0.2961917
```

Therefore, the probability that r>10% is about 30%.

Question 4:

Let X and Y have the joint density $f(x, y) = cx(y - x)e^{-y}, 0 <= x <= y < \infty$.

(a) Find c

$$\int_0^\infty \int_0^y cx(y-x)e^{-y}dxdy = 1$$

$$\int_0^\infty \int_0^y x(y-x)e^{-y}dxdy = \frac{1}{c}$$

$$\int_0^\infty \int_0^y xye^{-y} - x^2e^{-y}dxdy = \frac{1}{c}$$

$$\int_0^\infty \int_0^y xye^{-y} - x^2e^{-y}dxdy = \frac{1}{c}$$

$$\int_0^\infty \frac{y^3e^{-y}}{2} - \frac{y^3e^{-y}}{3} dy = \frac{1}{c}$$

$$\int_0^\infty \frac{y^3e^{-y}}{6} dy = \frac{1}{c}$$

$$\int_0^\infty y^3e^{-y} dy = \frac{6}{c}$$

$$-y^3e^{-y} + \int 3y^2e^{-y}dy = \frac{6}{c}$$

$$-y^3e^{-y} - 3y^2e^{-y} + \int 6ye^{-y}dy = \frac{6}{c}$$

$$-y^3e^{-y} - 3y^2e^{-y} - 6ye^{-y} + 6\int e^{-y}dy = \frac{6}{c}$$

$$\vdots$$

Solving using the bounds of integration going from 0 to ∞ we get:

$$6 = \frac{6}{c}$$
$$c = 1$$

(b) find $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$

First solve for marginals

$$f_x(x) = \int_x^\infty x(y-x)e^{-y}dy$$

$$= \int_x^\infty xye^{-y} - x^2e^{-y}dy$$

$$= x\int ye^{-y}dy - x^2\int e^{-y}dy$$

$$\vdots$$

$$= x[xe^{-x} + e^{-x}] - x^2e^{-x}$$
$$= xe^{-x}$$

$$f_y(y) = \int_0^y x(y-x)e^{-y}dx$$

$$= \int_0^y xye^{-y} - x^2e^{-y}dx$$

$$= y\int xe^{-y}dx - \int x^2e^{-y}dx$$

$$\vdots$$

$$= \frac{y^3e^{-y}}{2} - \frac{y^3e^{-y}}{3}$$

$$= \frac{y^3e^{-y}}{6}$$

Therefore,

$$f_{X|Y}(x|y) = \frac{6x(y-x)}{y^3}$$

and

$$f_{Y|X}(y|x) = \frac{(y-x)e^{-y}}{e^{-x}}$$

(c) Find $E(X|Y)$ and $E(Y|X)$

$$E[X|Y] = \int_0^y \frac{x6x(y-x)}{y^3}dx$$

$$= \int_0^y \frac{6x^2y - 6x^3}{y^3}dx$$

$$= \frac{1}{y^3}\int_0^y 6x^2y - 6x^3dx$$

$$\vdots$$

$$= \frac{1}{y^3}[2y^4 - \frac{3}{2}y^4]$$

$$= \frac{y}{2}$$

$$
E[Y|X] = \int_x^\infty \frac{y^2e^{-y} - xye^{-y}}{e^{-x}} dy
$$

$$
= \frac{1}{e^{-x}} \int_x^\infty y^2e^{-y} - xye^{-y} dy
$$

$$
= \frac{1}{e^{-x}} \int_x^\infty y^2e^{-y} dy - \int_x^\infty xye^{-y} dy
$$

$$
\vdots
$$

$$
= \frac{1}{e^{-x}} [x^2e^{-x} + 2xe^{-x} + 2e^{-x} - x^2e^{-x} - xe^{-x}]
$$

$$
= x + 2
$$

Question 5

Suppose that X1, X2,. . . , X10 is an i.i.d. sequence from an N(0, 1) distribution.Generate a sample of N = 10^{3} values from the distribution of max(X1, X2,. . . ,X10). Calculate the mean and standard deviation of this sample.

```
x<-0                                                        #make x empty

for (i in 1:1000){                                          #loop 10^3 times
  samp<-rnorm(10,mean=0, sd=1)                #take a sample of 10 normal RVs
  x[i] <- max(samp)                                         #store the maximums
}

#calculate the mean and standard deviation
mean(x)
```

```
## [1] 1.52893
```

```
sd(x)
```

```
## [1] 0.5576135
```

Question 6

Generate i.i.d. X1,. . . , Xn distributed Exponential(5) and compute Mn when n = 20. Repeat this N times, where N is large (if possible, take N = 10^5, otherwise as large as is feasible), and compute the proportion of values of Mn that lie between 0.19 and 0.21. Repeat this with n = 50. What property of convergence in probability do your results illustrate?

```
x<-0

for( i in 1:10^5){
  samp<-rexp(20, rate = 5)
  x[i]=mean(samp)
}

sub <- subset(x, x>=0.19 & x<=0.21)
length(sub)/length(x)
```

```
## [1] 0.17545
```

```
x<-0

for( i in 1:10^5){
  samp<-rexp(50, rate = 5)
  x[i]=mean(samp)
}

sub <- subset(x, x>=0.19 & x<=0.21)
length(sub)/length(x)
```

```
## [1] 0.27627
```

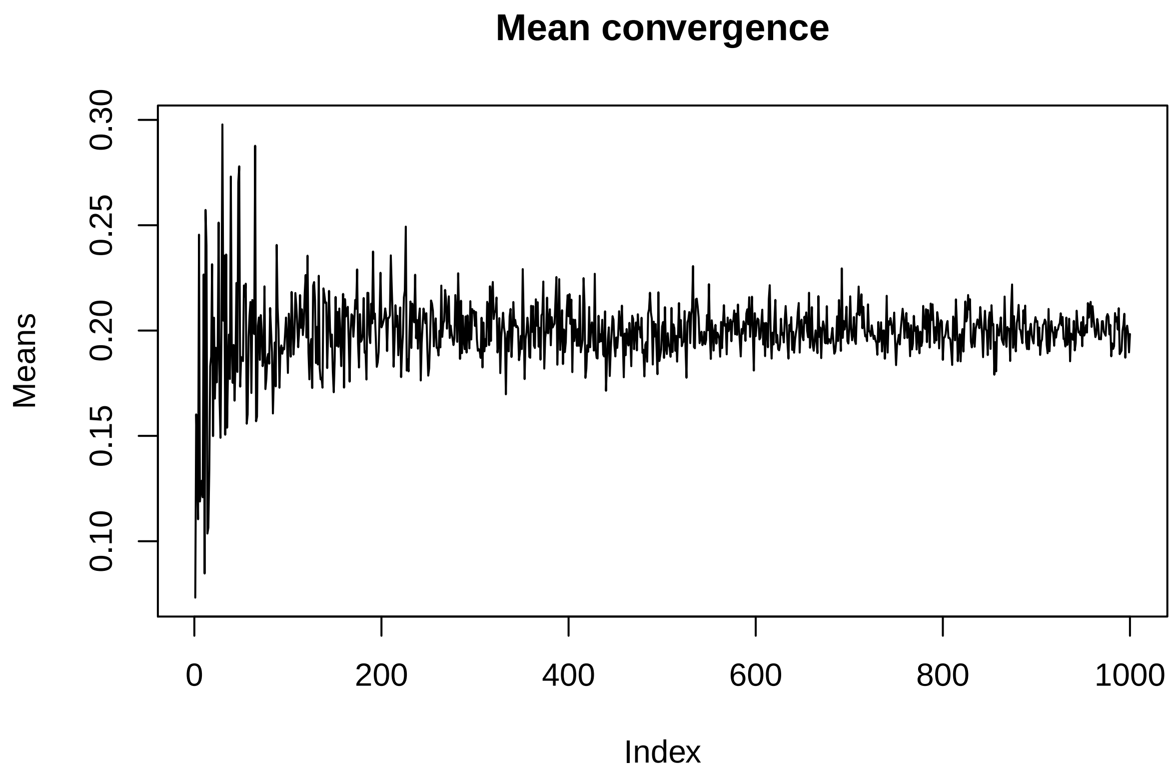
For n=20 the proportion is approximately 18% and approximately 27% for n=50. This illustrates the Weak Law of Large Numbers.

Question 7

Generate i.i.d. X1,. . . , Xn distributed Exponential(5) with n large (take n =10ˆ5 if possible). Plot the values M1, M2,. . . , Mn. To what value are they converging? How quickly?

```
sz<- 1
x<-0
for(i in 1:10^3){
  samp<-rexp(sz,rate=5)
  x[i]=mean(samp)
  sz<-sz+1
}

plot(x, type = "l", ylab="Means", main="Mean convergence")
```



From the plot we can see that the sample means converge to 0.2 and begins converging very quickly (almost right away).

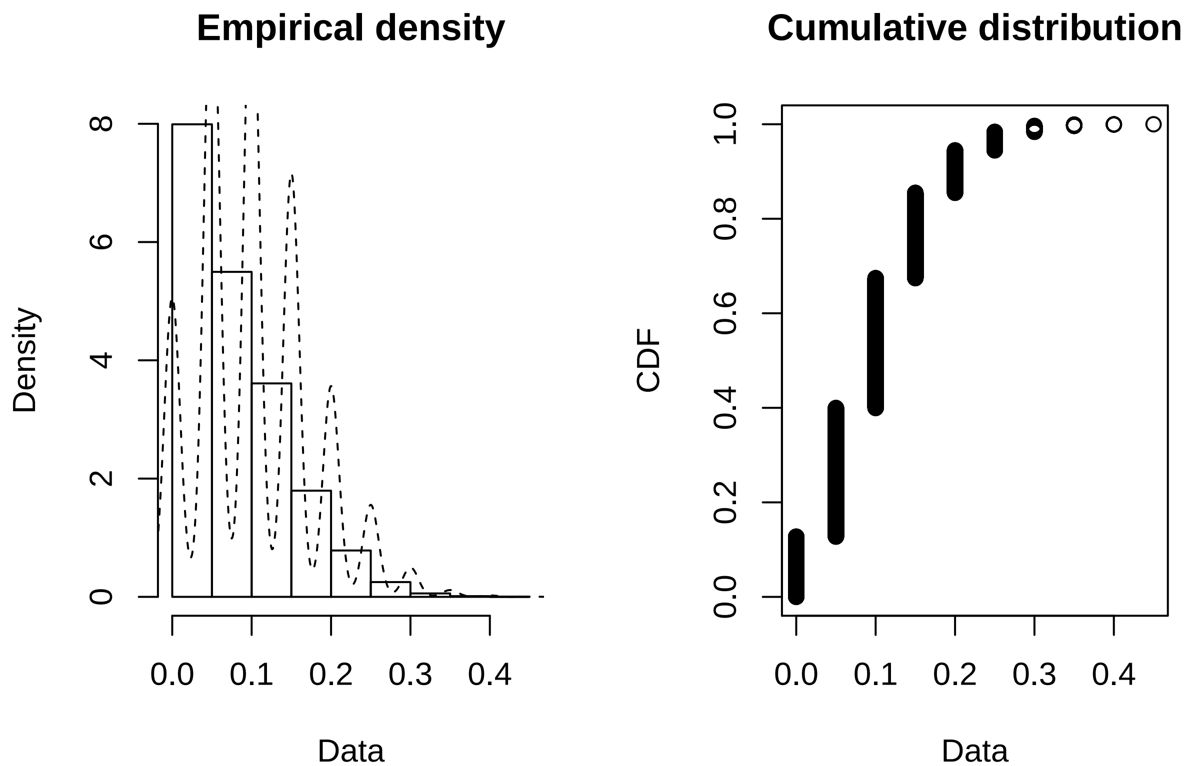
Question 8

Generate N samples X1, X2,. . . , X20 from the Binomial(10, 0.01) distribution for N large (N = 10ˆ4, if possible). Use these samples to construct a density histogram of the values of M20. Comment on the shape of this graph.

```
x<-0

for( i in 1:10^4){
  samp<-rbinom(20, size=10, prob=0.01)
  x[i]=mean(samp)
}

plotdist(x, histo=TRUE, demp=TRUE)
```



We can see that the graph is skewed and the density is converging to zero.