# CLT_Assignment

Fan Feng

2021/3/5

## Problem 1

A machine produces rope at a mean rate of mean of 4 feet per minute with standard deviation of 5 inches. Assume that the amounts produced in different minutes are independent and identically distributed, approximate the probability that the machine will produce at least 250 feet in one hour.

```
# 1 feet = 12 inch
mean_len = 48
sd = 5
mark = (250*12)/60
prob_1 = pnorm(mark, mean_len, sd/sqrt(60),lower.tail = FALSE)
print(prob_1)
```

```
## [1] 0.0009728868
```

The probability that the machine will produce at least 250 feet in one hour is 0.097%

#Problem 2

Assume that the distribution of the number of defects on any given bolt of cloth is the Poisson distribution with mean 5, and the number of defects on each bolt is counted for a random sample of 125 bolts. Determine the probability that the average number of defects per bolt in the sample will be less than 5.5 defects per bolt.

```
prob_2 = pnorm(5.5, 5, sqrt(5)/sqrt(125), lower.tail=TRUE)
print(prob_2)
```

```
## [1] 0.9937903
```

The probability that the average number of defects per bolt in the sample will be less than 5.5 defects per bolt is 99.4%

#Problem 3

Suppose that the proportion of defective items in a large manufactured lot is 0.1. What is the smallest random sample of items that must be taken from the lot in order for the probability to be at least 0.99 that the proportion of defective items in the sample will be less than 0.13?

```r
for(num in 1:10000){
  if (pnorm(0.13, 0.1, sqrt(0.1*(1-0.1)/num), lower.tail=TRUE) > 0.99){
    break
    }
}
print(num)
```

```
## [1] 542
```

The smallest size of random sample of items is 542.

#Problem 4

Suppose that 16 digits are chosen at random with replacement from the set 0, . . . , 9. What is the probability that their average will lie between 4 and 6?

```r
e_x = 4.5
e_x_square = 28.5
var = sqrt(28.5 - 4.5*4.5)
prob_4 <- pnorm(6, e_x, var/4) - pnorm(4, e_x, var/4)
print(prob_4)
```

```
## [1] 0.7385259
```

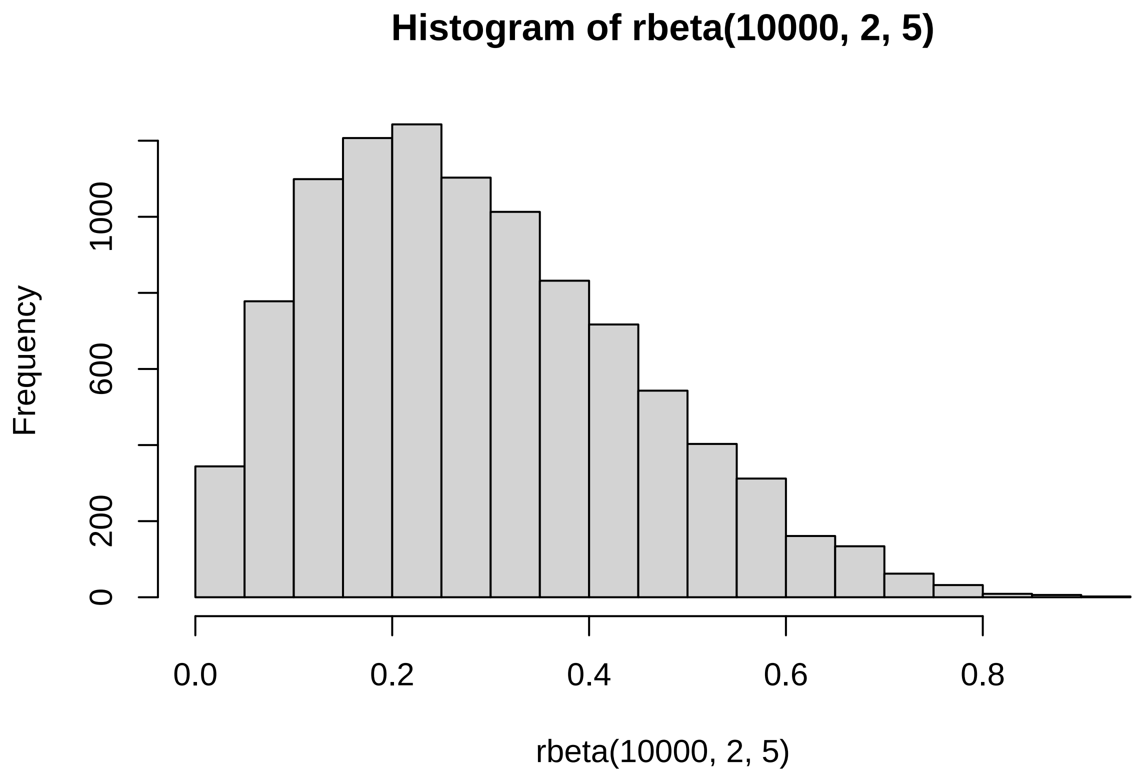
The probability is 73.9%.

#Problem 5

Select a skewed distribution from which to sample. Using R, demonstrate the convergence of the mean value of your samples to the Normal distribution. Assume you are making this demonstration to someone who has little or no statistical training. Your demonstration should take no more than 10 minutes. Along with your code, outline the commentary you would use in your demonstration.

First, we chose a skewed distribution as following:

```r
hist(rbeta(10000,2,5))
```




```r
a = 2
b = 5
ex = a/(a+b)
var = a*b/((a+b)^2 * (a+b+1))
print(ex)
```

```
## [1] 0.2857143
```

```r
print(var)
```

```
## [1] 0.0255102
```

```r
clt = c()
k = 5000
for(i in 1:5000){
  clt[i] = (sum(rbeta(k,2,5))-k*ex)/sqrt(k*var)
}
```

```r
print(mean(clt))
```

```
## [1] 0.004528719
```

```r
print(sd(clt))
```

```
## [1] 0.9873224
```

Shapiro-Wilk's method is widely recommended for normality test. It is based on the correlation between the data and the corresponding normal scores.

```r
shapiro.test(clt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  clt
## W = 0.99943, p-value = 0.1303
```

p value is 0.5893, which is much larger than 0.05. It implies that the distribution of the data are not significantly different from normal distribution. Therefore, the new distribution is a normal distribution.