Practical 1 Sampling Distribution of Statistics/Estimater

Ratnesh D Patil

2024-05-03

Q1.Conduct a simulation study from N($\mu = 10, \sigma^2 = 1$).For the following statistics i)Sample Mean ($\overline{X}$) ii) Sample Variance (S^2)

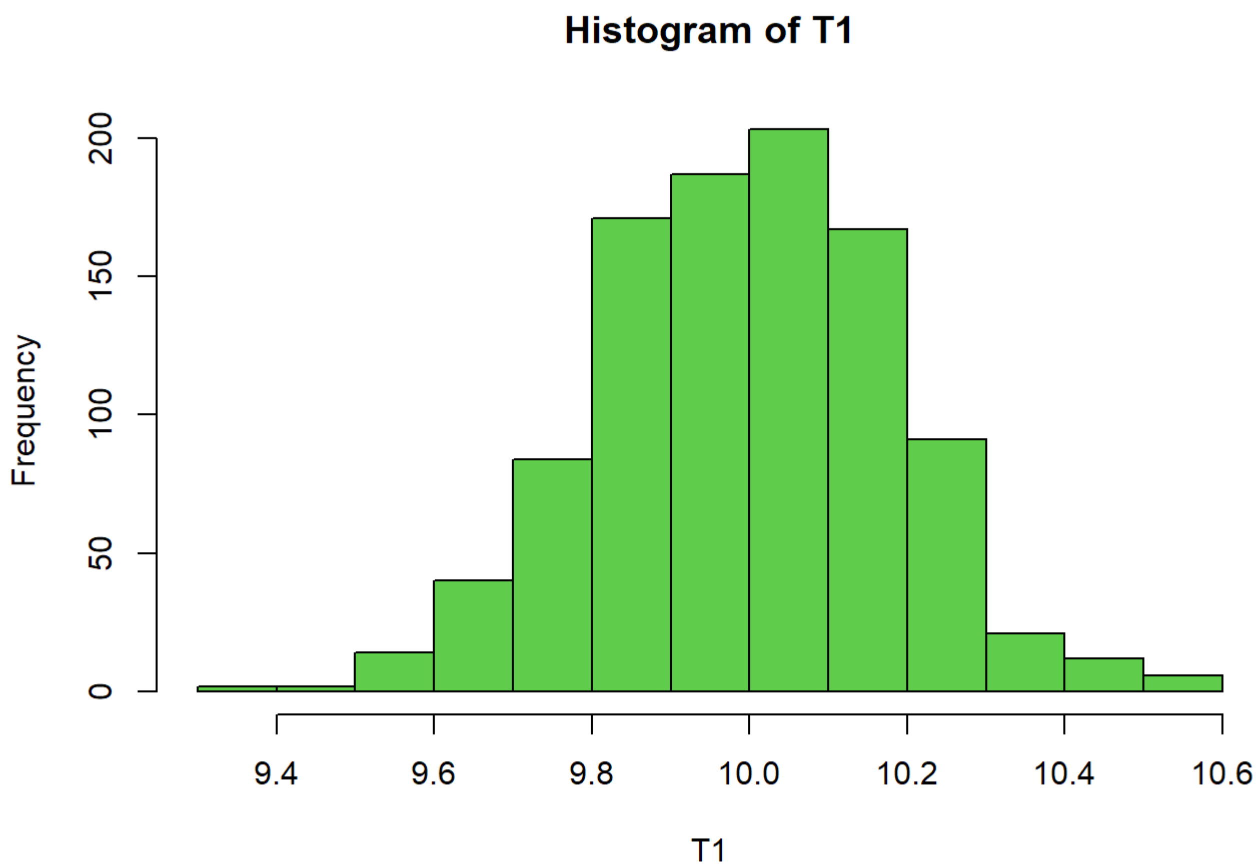
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

iii)Sample Median $\tilde{X}$.

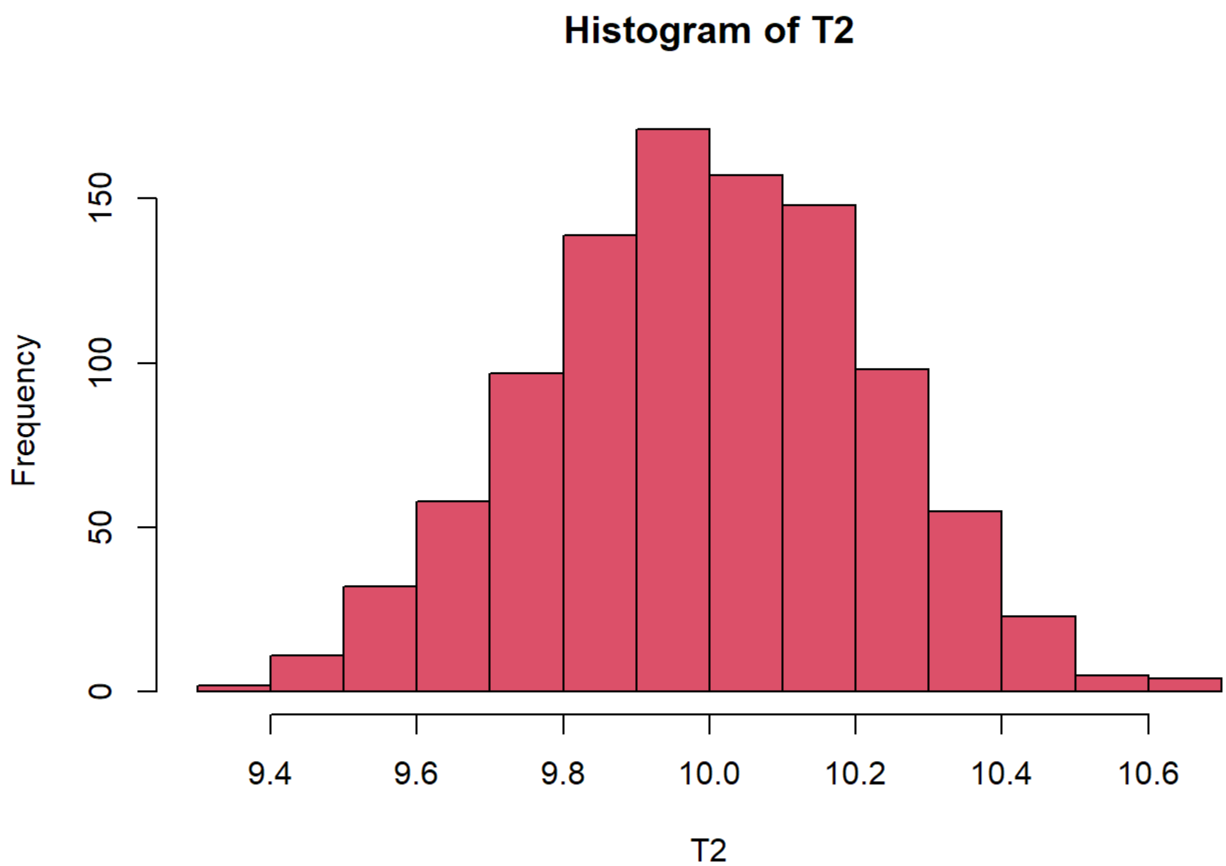
```
n <- c(5,15,30);
mu=10
sigma=1
Mean_T1=Mean_T2=Mean_T3=Mean_T4=var_T1=var_T2=var_T3=var_T4=0
for (i in 1:length(n))
  {x <- matrix(rnorm(1000*n[i],mu,sigma),1000)
  X_bar <- apply(x,1,mean)
  T1 <- X_bar
  T2 <- apply(x,1,median)
  T3 <-apply(x,1,var)
  T4 <-((n[i]-1)/n[i])*apply(x,1,var)
  Mean_T1[i] <- mean(T1)
  Mean_T2[i] <- mean(T2)
  Mean_T3[i] <- mean(T3)
  Mean_T4[i] <- mean(T4)
  var_T1[i] <- var(T1)
  var_T2[i] <- var(T2)
  var_T3[i] <- var(T3)
  var_T4[i] <- var(T4)
}
cbind(n,Mean_T1,Mean_T2,Mean_T3,Mean_T4,var_T1,var_T2,var_T3,var_T4)
```

```
##       n   Mean_T1   Mean_T2   Mean_T3   Mean_T4     var_T1     var_T2
## [1,]  5 10.012995 10.018839 1.0210520 0.8168416 0.20816706 0.28292887
## [2,] 15  9.985013  9.976233 0.9915917 0.9254856 0.06926256 0.10263397
## [3,] 30  9.994302  9.993947 0.9975263 0.9642754 0.03405358 0.05025161
##          var_T3    var_T4
## [1,] 0.52216552 0.3341859
## [2,] 0.13129232 0.1143702
## [3,] 0.06587358 0.0615552
```

```
hist(T1,col=3)
```

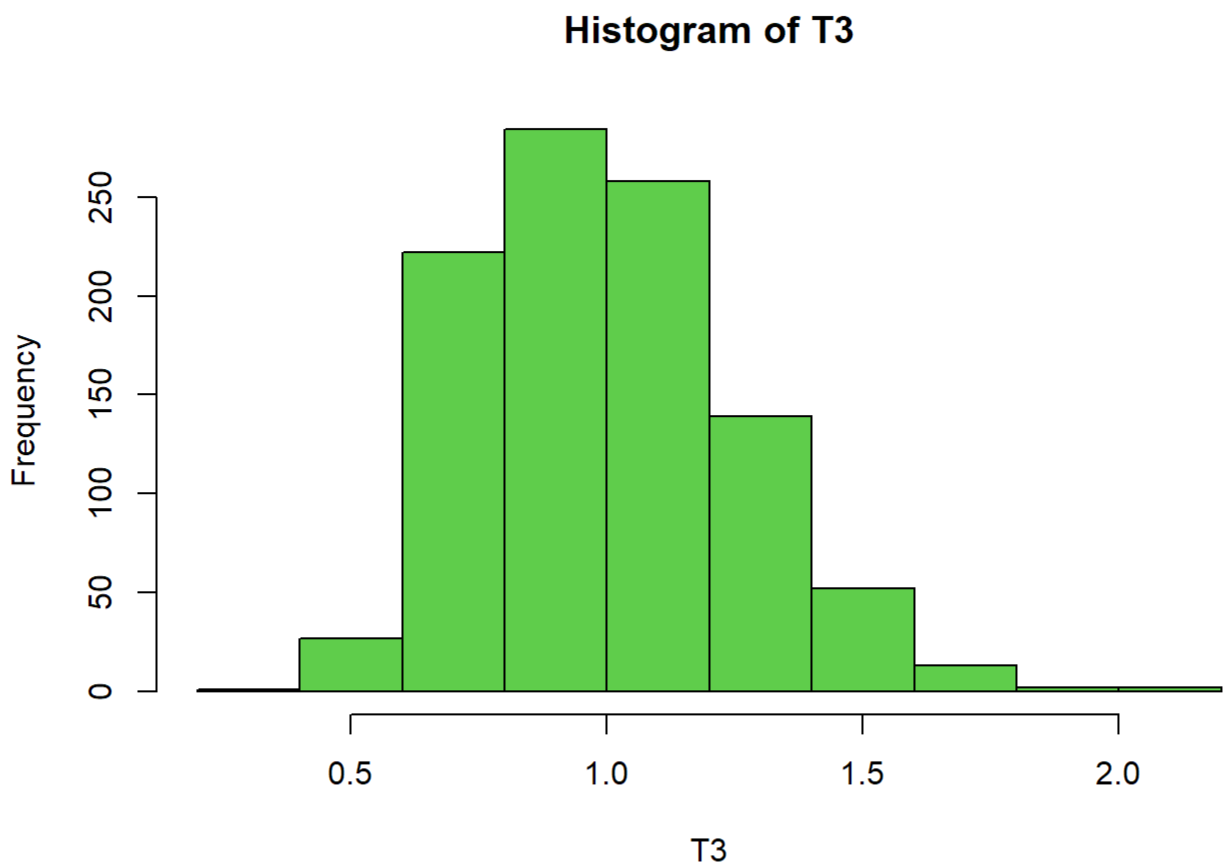

Histogram of T1
Frequency
0
50
100
150
200
9.4
9.6
9.8
10.0
10.2
10.4
10.6
T1

```
hist(T2,col=2)
```

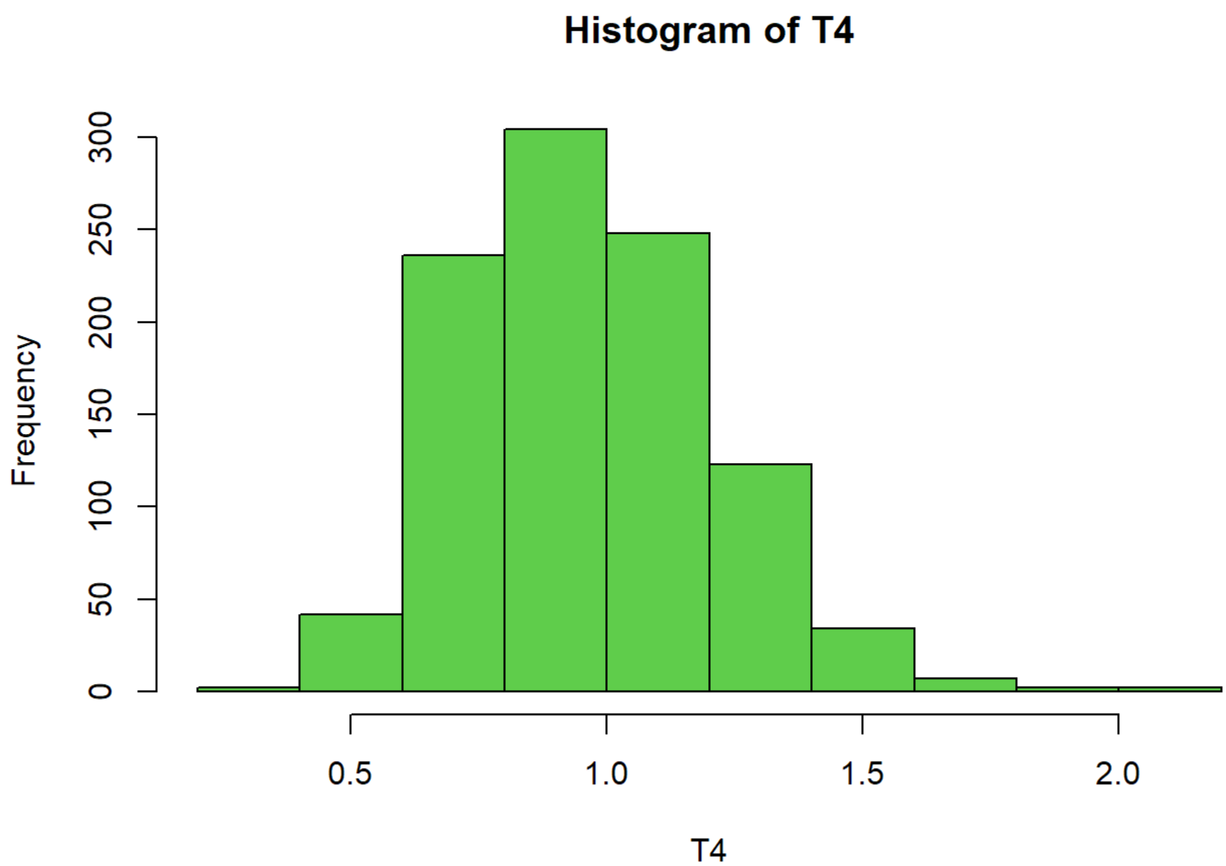


Histogram of T2
Frequency
0
50
100
150
9.4
9.6
9.8
10.0
10.2
10.4
10.6
T2


```
hist(T3,col=3)
```




```
hist(T4,col=3)
```

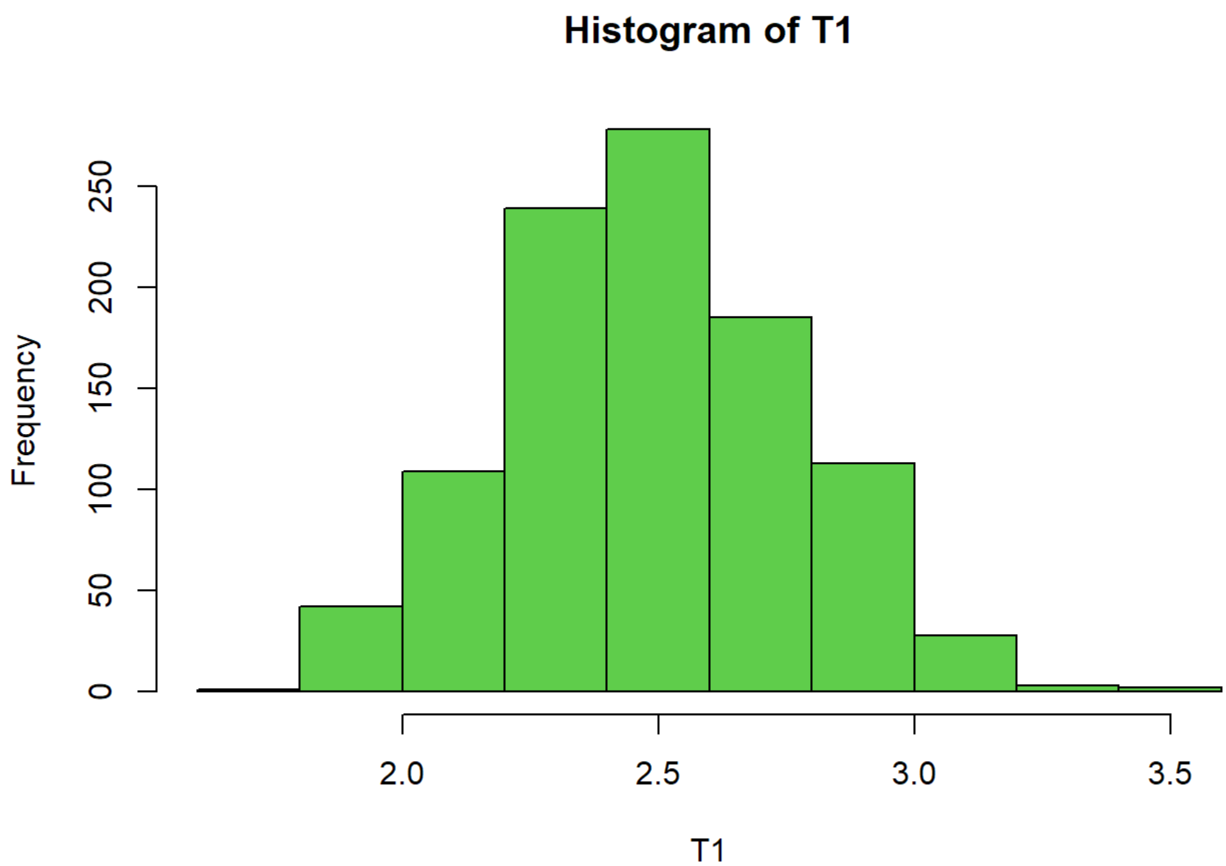


Q2.Conduct a simulation study from Poisson $\lambda = 2.5$ distribution. For the following statistics: i) sample mean $(\bar{\lambda})$ ii) sample variance λ^2

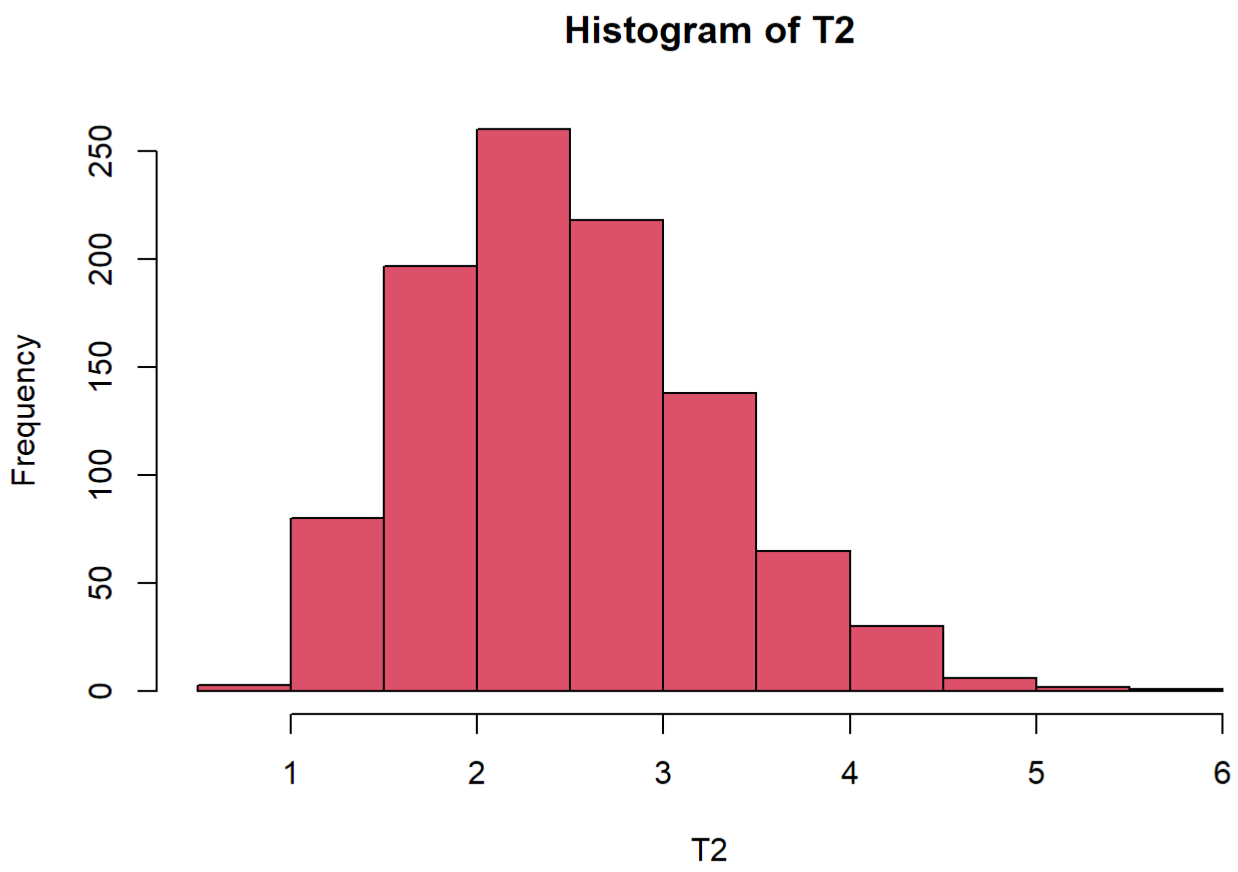
```
n <- c(5,15,30);
Mean_T1=Mean_T2=var_T1=var_T2=0
for (i in 1:length(n))
{x <- matrix(rpois(1000*n[i],2.5),1000)
T1 <- apply(x,1, mean)
T2 <- apply(x,1, var)
Mean_T1[i] <- mean(T1)
var_T1[i] <- var(T1)
Mean_T2[i] <- mean(T2)
var_T2[i] <- var(T2)
}
cbind(n,Mean_T1,Mean_T2,var_T1,var_T2)
```

```
##       n  Mean_T1  Mean_T2     var_T1    var_T2
## [1,]  5 2.544800 2.610000 0.47847143 4.4896697
## [2,] 15 2.519133 2.459333 0.15307588 0.9465746
## [3,] 30 2.498000 2.496520 0.07850562 0.5592842
```

```
hist(T1,col=3)
```




```
hist(T2,col=2)
```

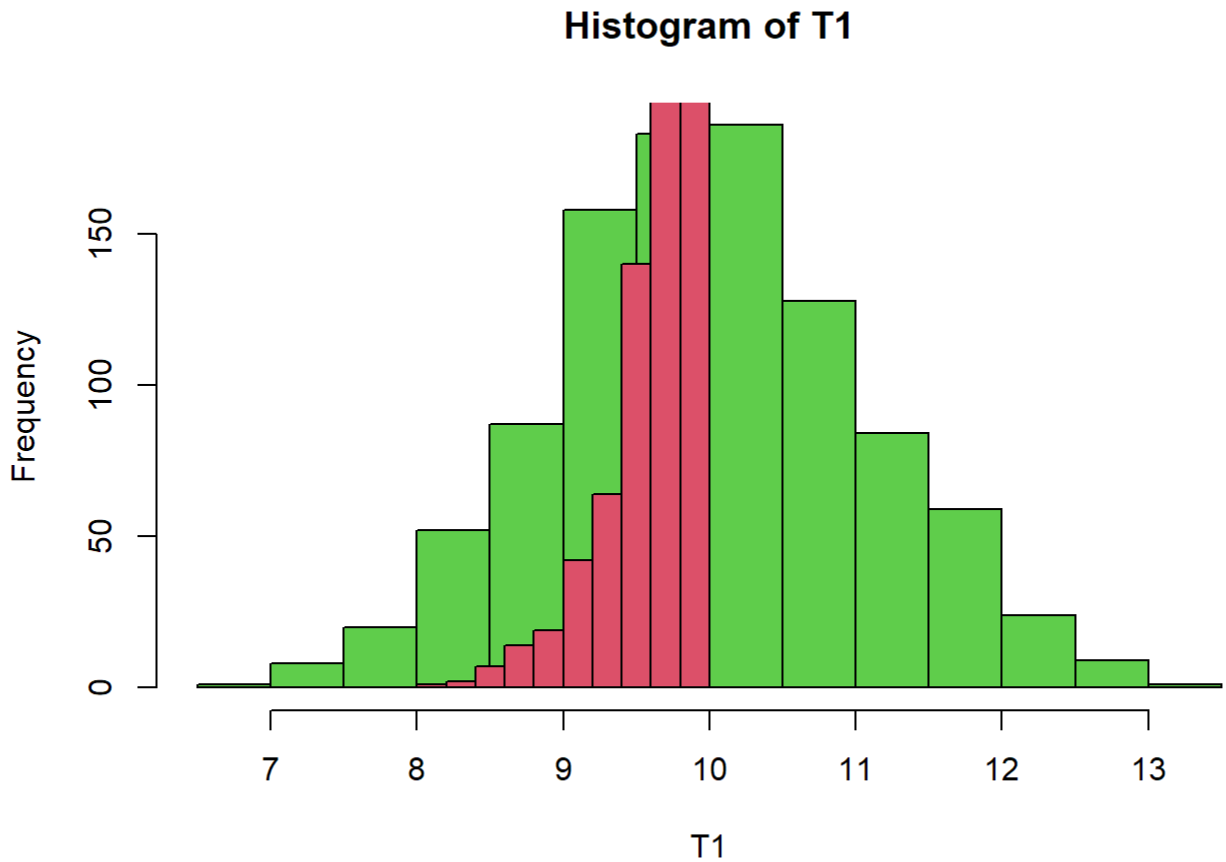


Q3.Conduct a simulation study from $U(0, \theta), \theta > 0$ for the following statistics: i)$2\overline{X}$ ii) $X_{(n)}$

```
n <- c(5,15,30);
Mean_T1=Mean_T2=var_T1=var_T2=0
for (i in 1:length(n))
  {x <- matrix(runif(1000*n[i],0,10),1000)
  X_bar <- apply(x,1,mean)
  T1 <- 2*X_bar
  T2 <- apply(x,1, max)
  Mean_T1[i] <- mean(T1)
  var_T1[i] <- var(T1)
  Mean_T2[i] <- mean(T2)
  var_T2[i] <- var(T2)
}
cbind(n,Mean_T1,Mean_T2,var_T1,var_T2)
```

```
##       n  Mean_T1  Mean_T2   var_T1     var_T2
## [1,]  5 10.14313 8.349766 7.257142 1.97104445
## [2,] 15  9.98314 9.349730 2.200623 0.39958141
## [3,] 30 10.00362 9.680979 1.124983 0.09345632
```

```
hist(T1,col=3)
hist(T2,col=2,add=T)
```



Q4. Conduct a simulation study from $\gamma(\alpha = 8, \beta = 10)$ for the following statistics: $T_1 = \tilde{a} = \frac{m_1^2}{m_2 - m_1^2}$, $T_{21} = \tilde{\beta} = \frac{m_2 - m_1^2}{m_1}$ where,

$$m_1 = E(X) = \overline{X}$$

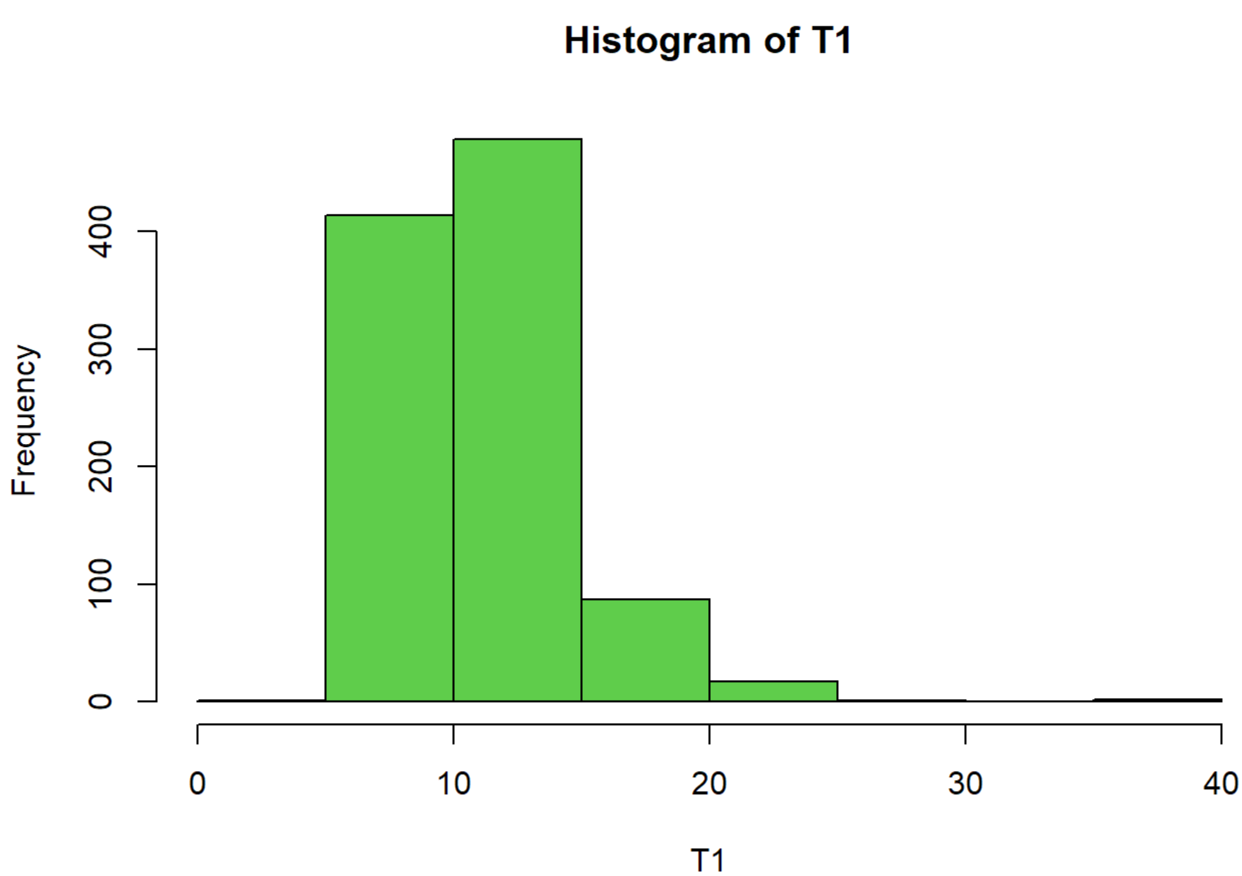
,

$$m_2 = E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

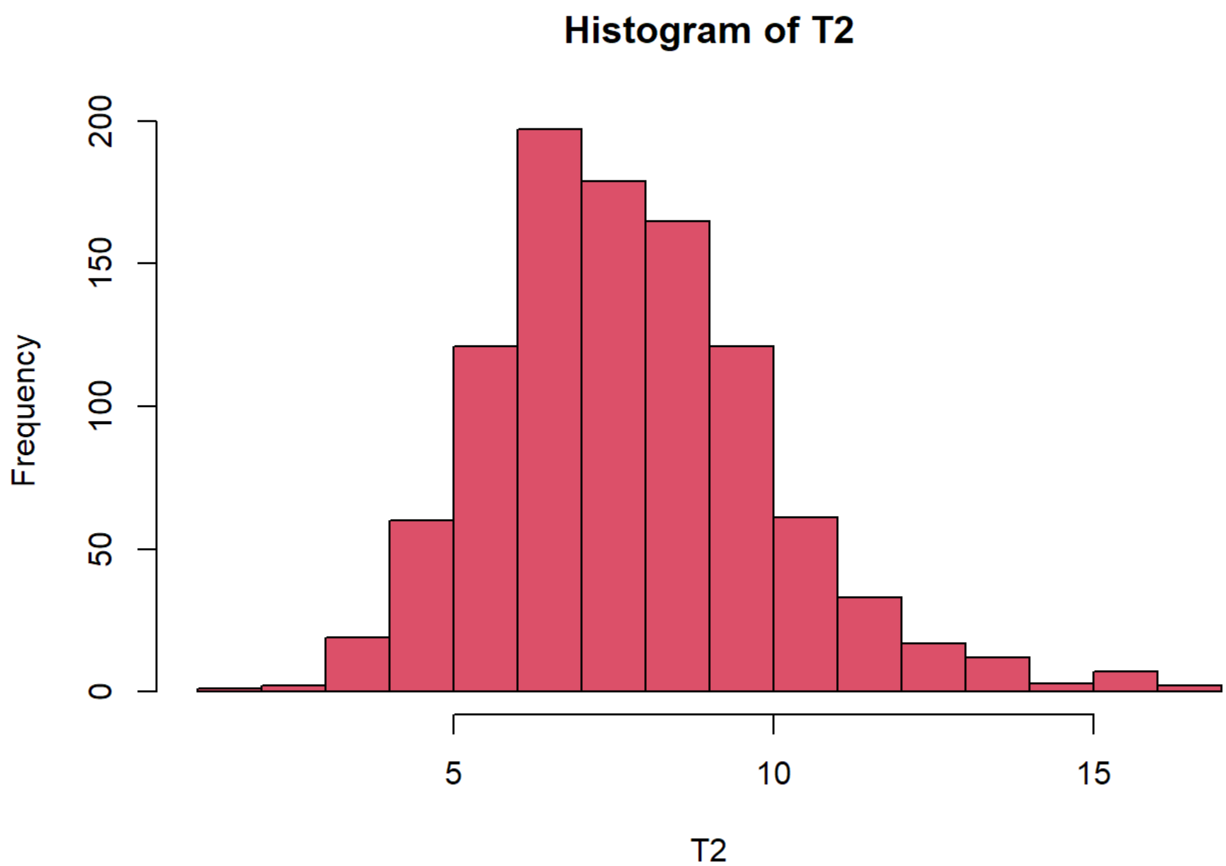
```r
n <- c(5,15,30);
Mean_T1=Mean_T2=var_T1=var_T2=0
for (i in 1:length(n))
{x <- matrix(rgamma(1000*n[i],10,1/8),1000) #rgamma(beta,1/alpha)
m1 <- apply(x,1,mean)
m2 <- apply(x^2,1, mean)
T1 <- m1^2/(m2-m1^2)
T2 <- (m2-m1^2)/m1
Mean_T1[i] <- mean(T1)
var_T1[i] <- var(T1)
Mean_T2[i] <- mean(T2)
var_T2[i] <- var(T2)
}
cbind(n,Mean_T1,Mean_T2,var_T1,var_T2)
```

```
##       n  Mean_T1  Mean_T2     var_T1    var_T2
## [1,]  5 24.72462 6.350857 1244.32589 19.610014
## [2,] 15 12.53613 7.607819   36.55114  9.951160
## [3,] 30 11.18358 7.752012   11.53637  4.789475
```

```r
hist(T1,col=3)
```




```r
hist(T2,col=2)
```



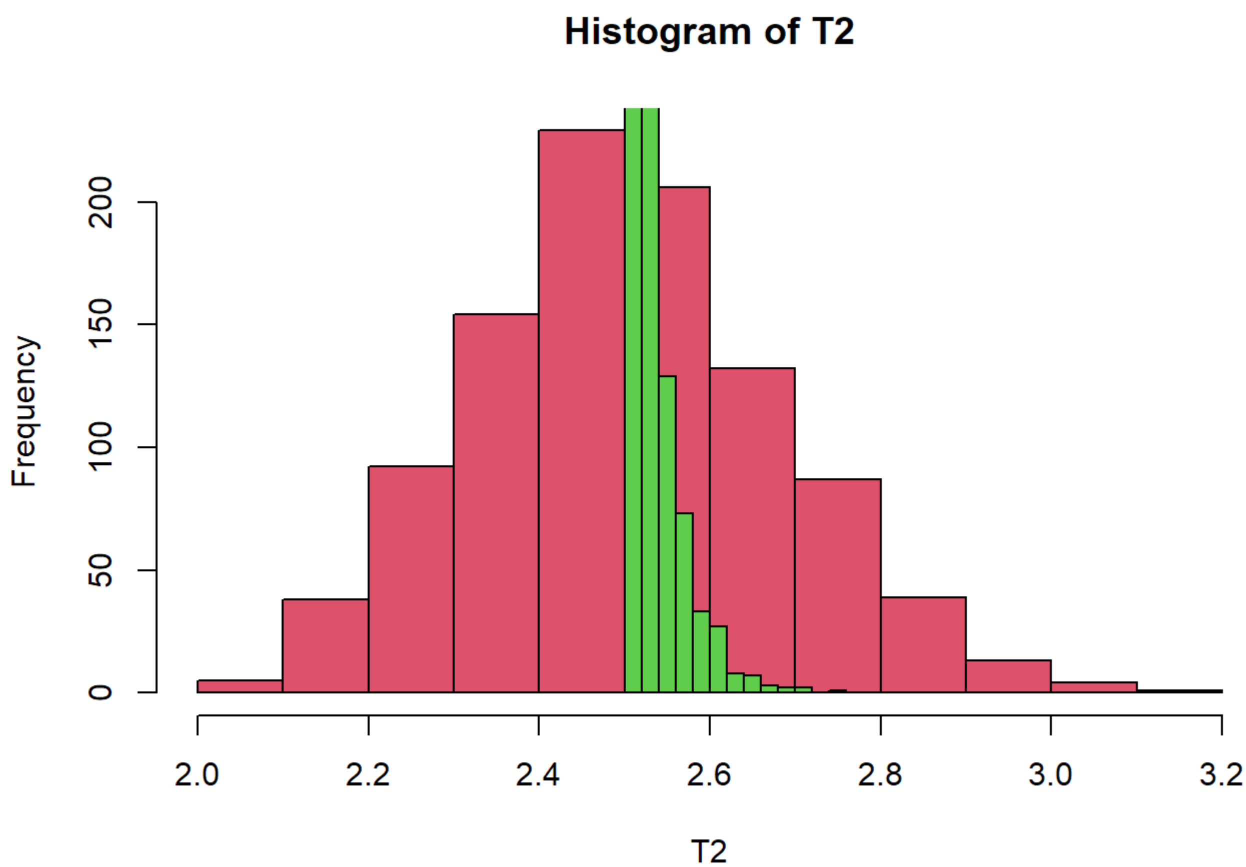
Q5. Conduct a simulation study from Exponential distribution with location parameter θ,for the following statistics: i) $X_{(1)}$ ii) $\overline{X} - 1$

Solution: Here X~ Exponential distribution with location parameter θ for drawing random sample there is no direct command for this. So we need to use Probability Transformation Theorem i.e. X is continuous then $F(X) \sim U(0,1) = y$ then $X = \theta - log(1-y) \sim exp(location = \theta)$

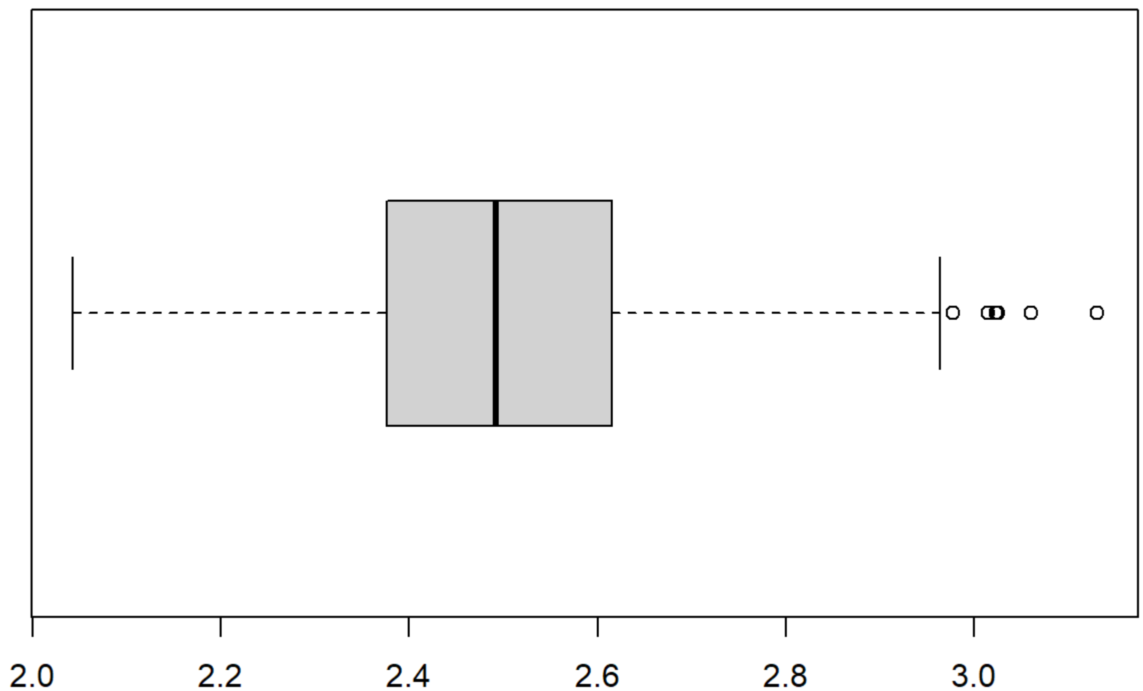
```
theta <- 2.5;
n <- c(5,15,30);
Mean_T1=Mean_T2=var_T1=var_T2=0
for (i in 1:length(n)){
y <- matrix(runif(1000*n[i],0,1),1000)
x <- theta-log(1-y)
T1 <- apply(x,1,min)
T2 <- apply(x,1,mean)-1
Mean_T1[i] <- mean(T1)
var_T1[i] <- var(T1)
Mean_T2[i] <- mean(T2)
var_T2[i] <- var(T2)}
cbind(n,Mean_T1,Mean_T2,var_T1,var_T2)
```

```
##        n  Mean_T1  Mean_T2      var_T1     var_T2
## [1,]  5 2.694785 2.505153 0.037962926 0.20022186
## [2,] 15 2.567210 2.501206 0.004022401 0.06442386
## [3,] 30 2.532209 2.501480 0.001083658 0.03240791
```

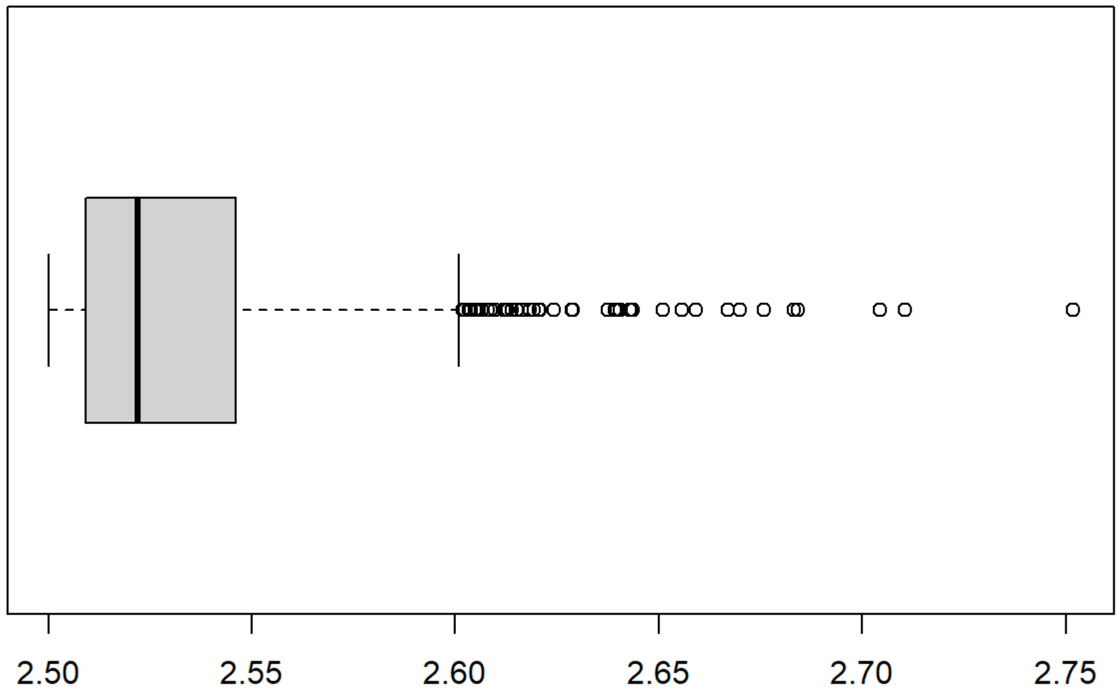
```
hist(T2,col=2)
hist(T1,col=3,add=T)
```




```
boxplot(T2,horizontal=T)
```




```
boxplot(T1,horizontal=T)
```



Q6. Conduct a simulation study from Exponential distribution with location parameter μ and scale parameter σ,for the following statistics:

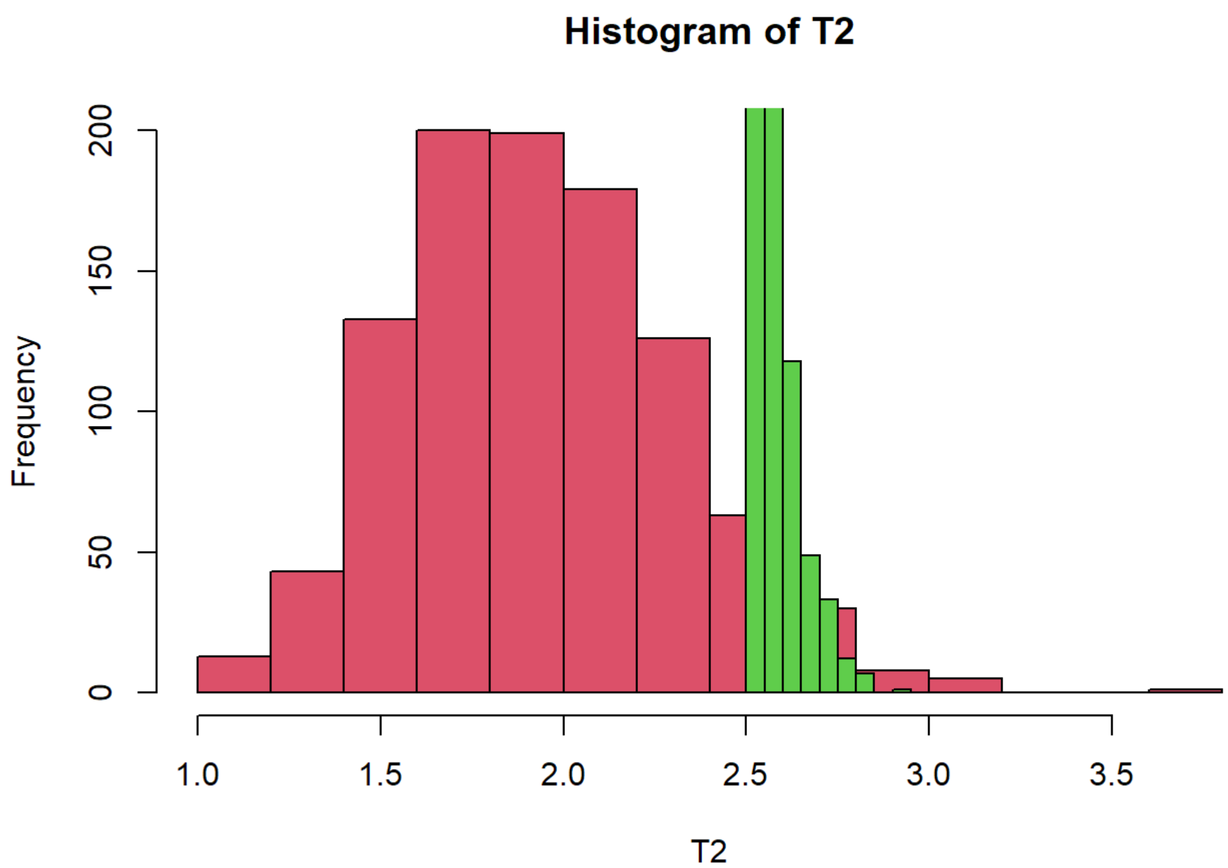
i. $X_{(1)}$
ii. $\overline{X} - X_{(1)}$

solution: for drawing random sample there is no direct command for this. So we need to use Probability Transformation Theorem i.e. X is continuous then $F(X) \sim U(0,1) = y$ then
$X = \mu - \sigma log(1-y) \sim exp(location = \mu, scale = \sigma)$

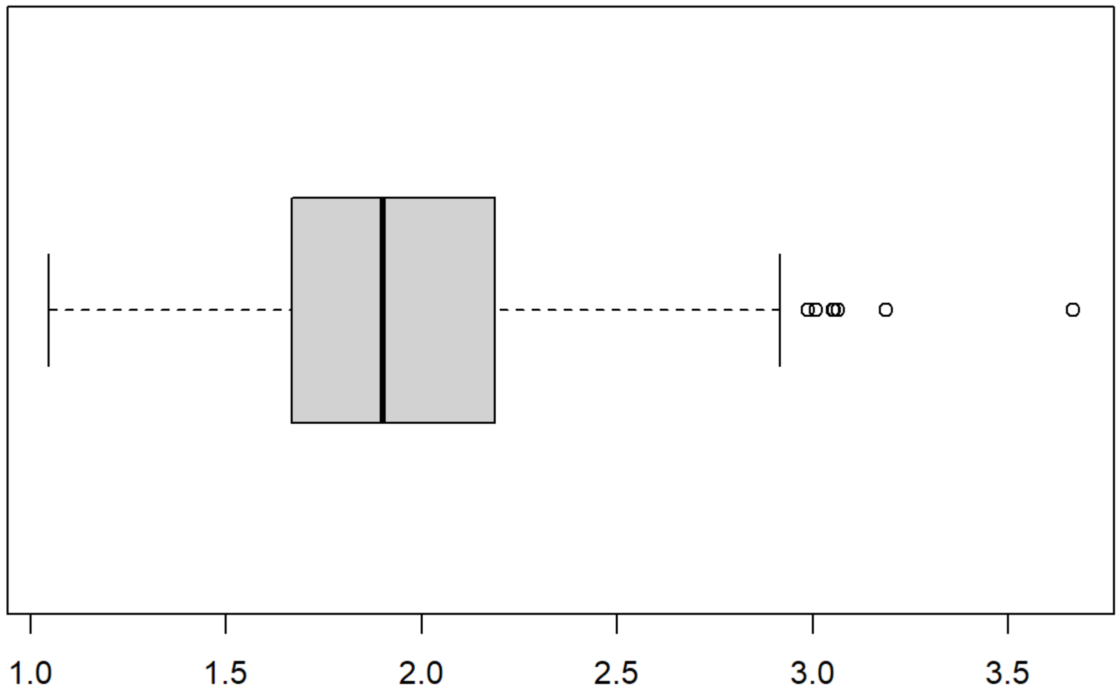
```
mu <- 2.5
sigma <- 2
n <- c(5,15,30);
Mean_T1=Mean_T2=var_T1=var_T2=0
for (i in 1:length(n)){
y <- matrix(runif(1000*n[i],0,1),1000)
x <- mu-sigma*log(1-y)
T1 <- apply(x,1,min)
T2 <- apply(x,1,mean)-T1
Mean_T1[i] <- mean(T1)
var_T1[i] <- var(T1)
Mean_T2[i] <- mean(T2)
var_T2[i] <- var(T2)}
cbind(n,Mean_T1,Mean_T2,var_T1,var_T2)
```

```
##       n  Mean_T1  Mean_T2     var_T1    var_T2
## [1,]  5 2.912890 1.607752 0.15181369 0.6452151
## [2,] 15 2.629474 1.877940 0.01934264 0.2572104
## [3,] 30 2.565559 1.932599 0.00401458 0.1363575
```

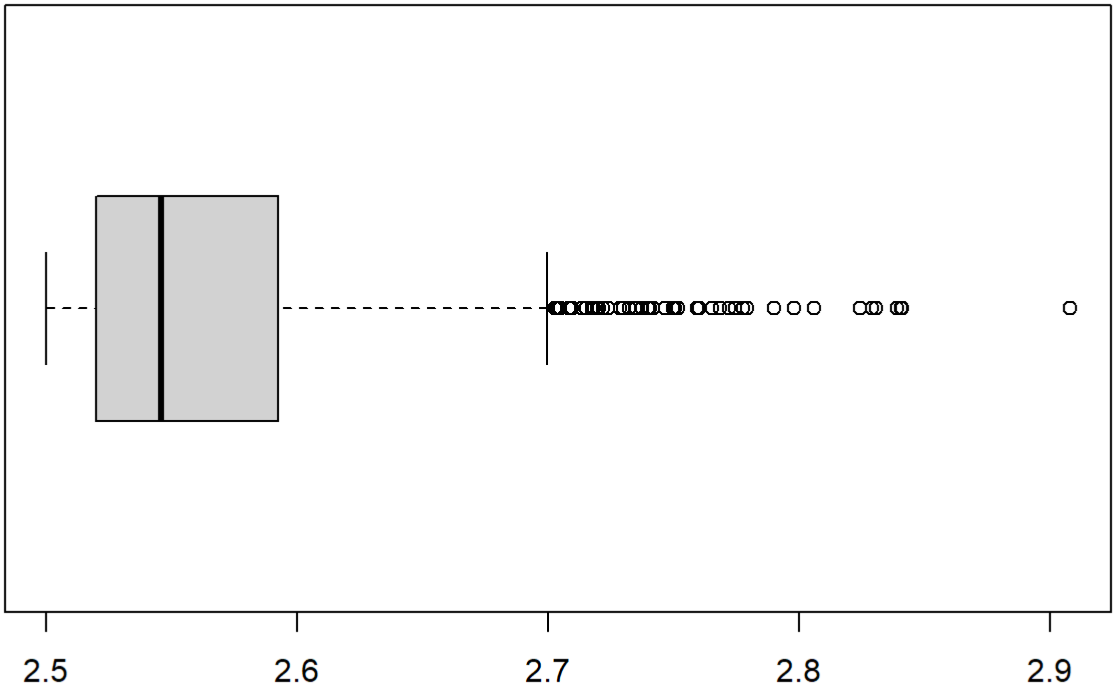
```
hist(T2,col=2)
hist(T1,col=3,add=T)
```




```
boxplot(T2,horizontal=T)
```




```
boxplot(T1,horizontal=T)
```



Q7. Conduct a simulation study from pareto distribution. The PDF of the Pareto distribution:

$$f(x) = \frac{\theta}{x^{\theta+1}} \quad \text{for } x > 1, \theta > 0$$

for the following statistics:

i. $$\frac{\bar{X}}{\bar{X} - 1}$$

ii. $$\frac{n}{\sum_{i=1}^{n} \log(x_i)}$$

Solution: for drawing random sample there is no direct command for this. So we need to use Probability Transformation Theorem i.e. X is continuous then $F(X) \sim U(0,1) = y$ then $X = (1-y)^(- 1/\theta) \sim$ Pareto distribution.

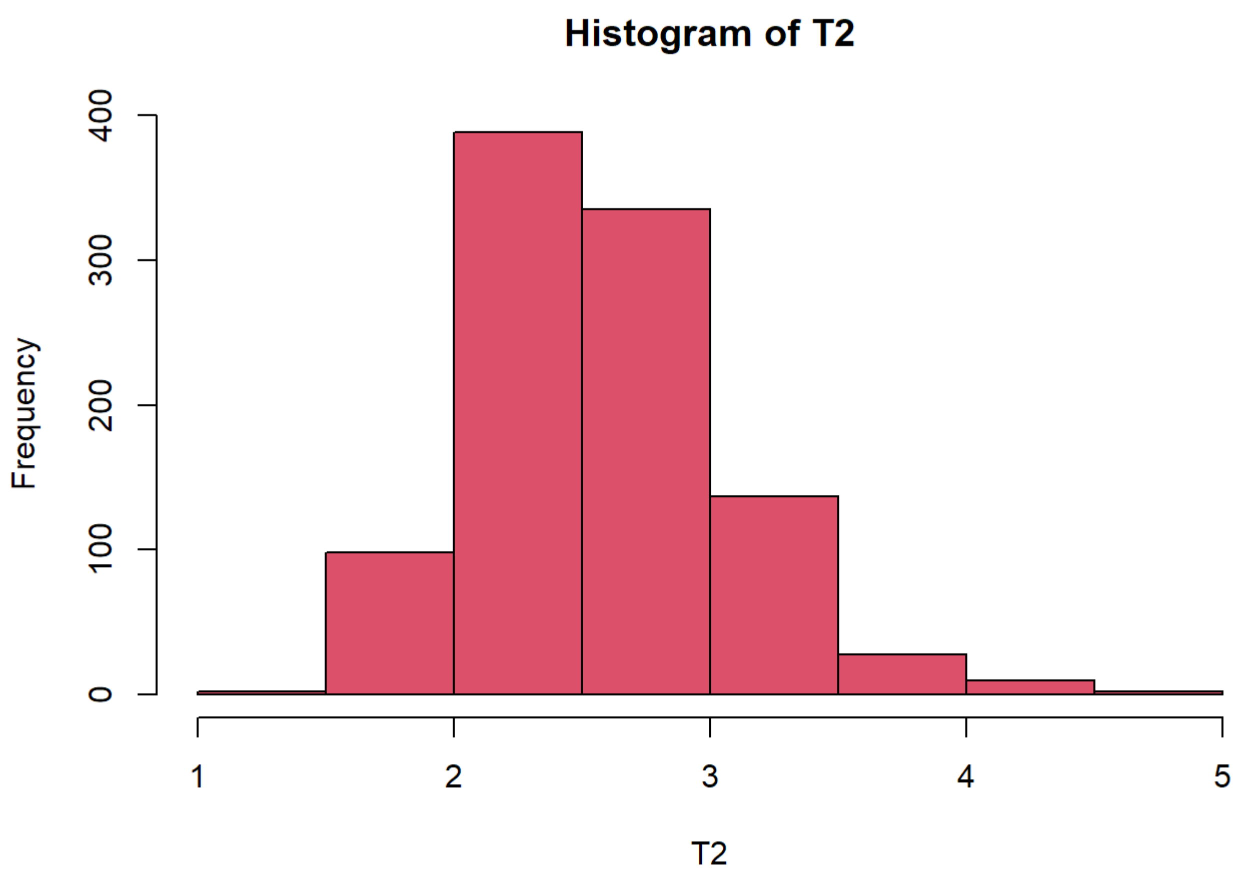
Note:

$$\frac{n}{\sum_{i=1}^{n} \log(x_i)} = \frac{n}{\sum_{i=1}^{n} Z_i} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} Z_i} = \frac{1}{\overline{Z}}$$

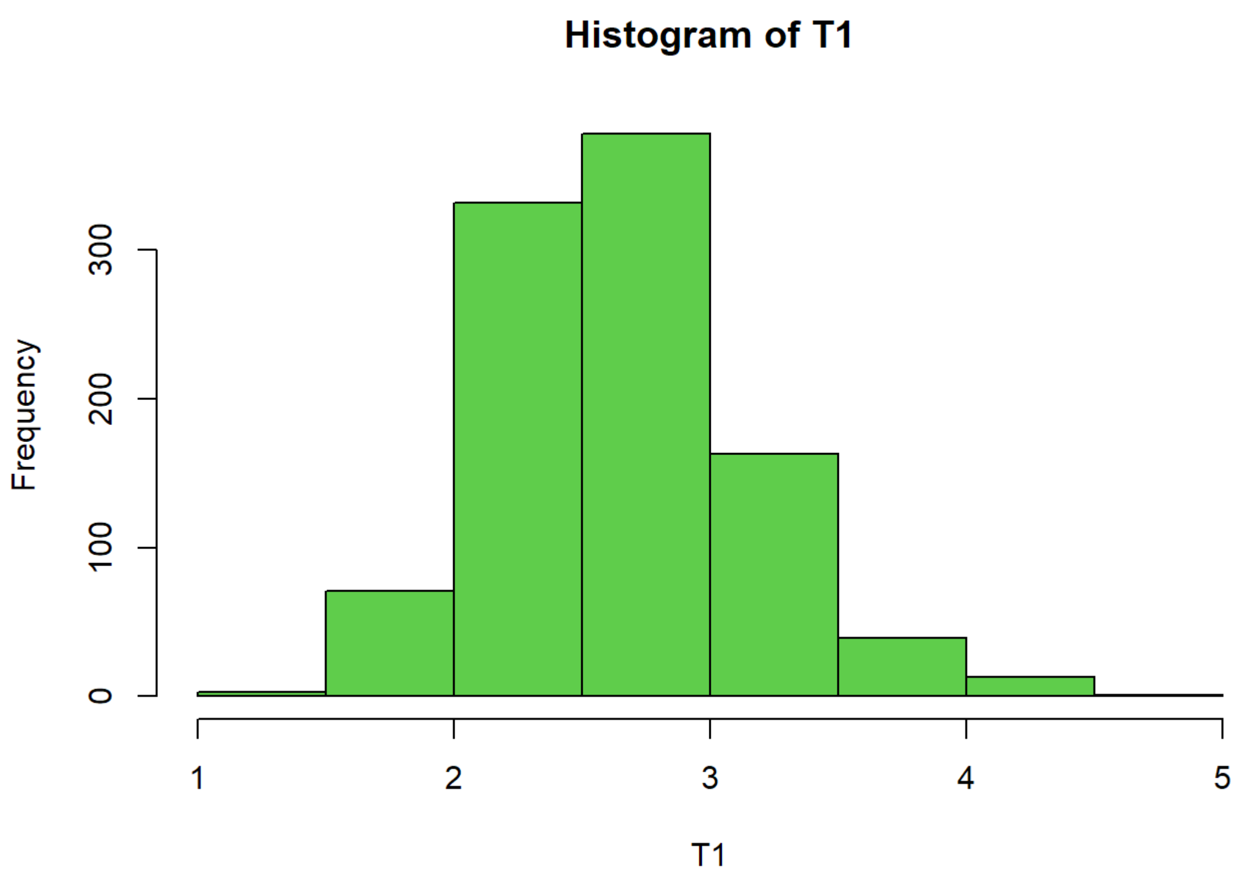
```r
theta <- 2.5;
n <- c(5,15,30);
Mean_T1=Mean_T2=var_T1=var_T2=0
for (i in 1:length(n)){
  y <- matrix(runif(1000*n[i],0,1),1000)
  x <- (1-y)^(-1/theta)
  X_bar <- apply(x,1,mean)
  T1 <- X_bar/(X_bar-1)
  z_bar <- apply(log(x),1,mean)
  T2=1/z_bar
  Mean_T1[i] <- mean(T1)
  var_T1[i] <- var(T1)
  Mean_T2[i] <- mean(T2)
  var_T2[i] <- var(T2)}
cbind(n,Mean_T1,Mean_T2,var_T1,var_T2)
```

```
##       n  Mean_T1  Mean_T2    var_T1    var_T2
## [1,]  5 3.385860 3.157533 3.0872570 3.1785882
## [2,] 15 2.823357 2.699590 0.5951104 0.5916175
## [3,] 30 2.650432 2.572386 0.2456588 0.2364770
```

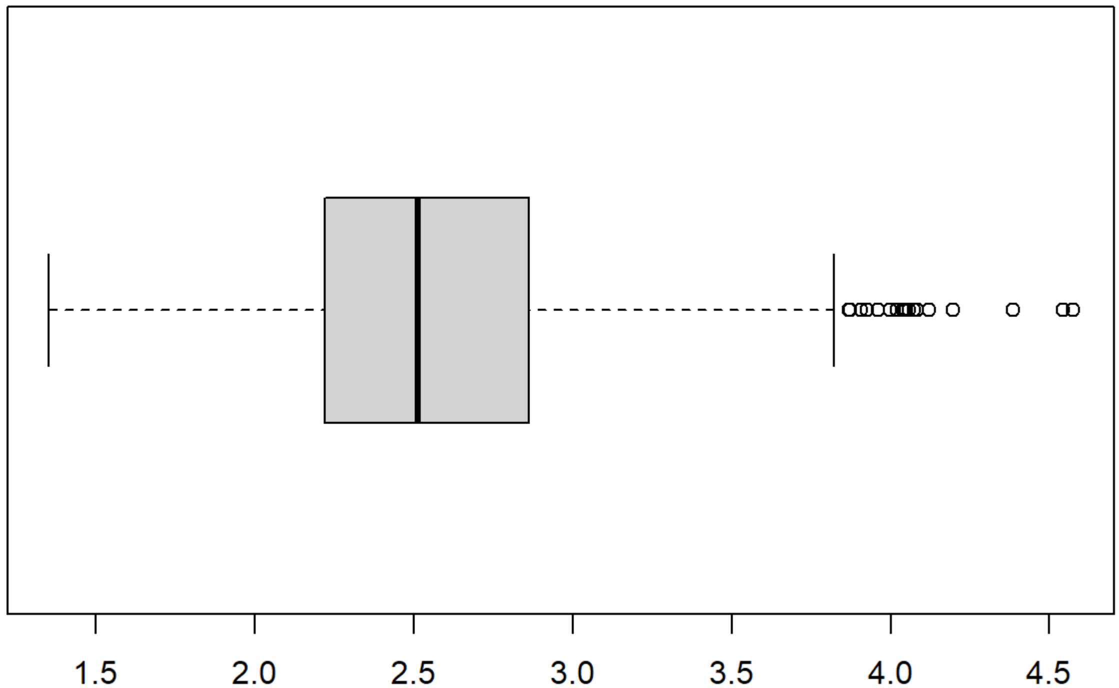
```r
hist(T2,col=2)
```



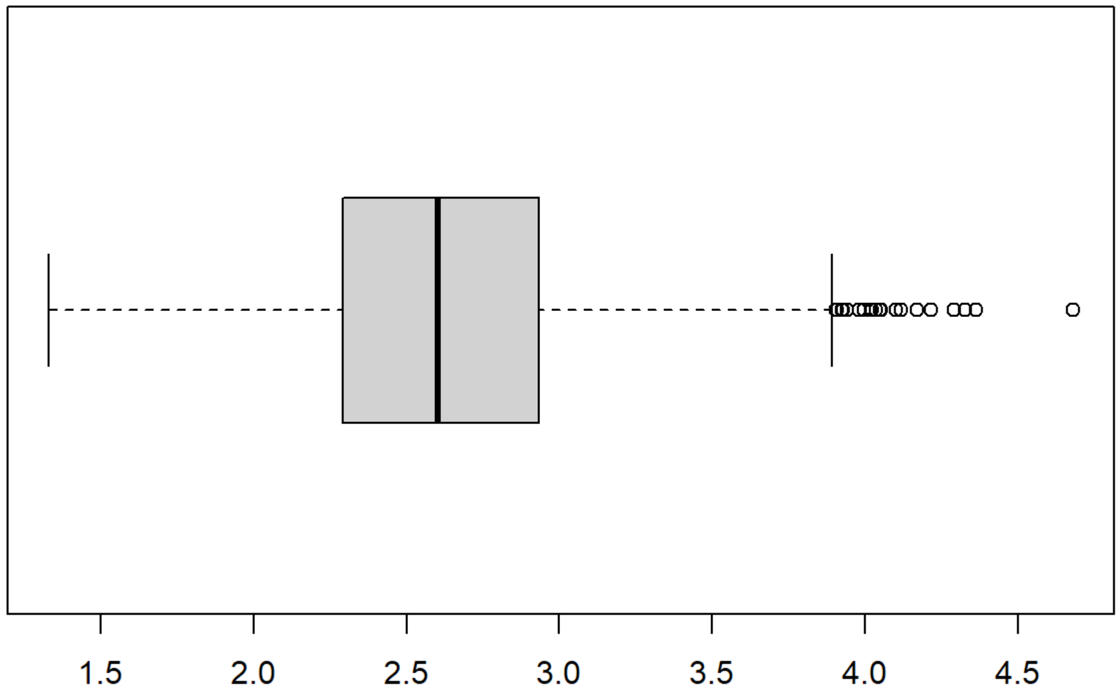
```
hist(T1,col=3)
```




```
boxplot(T2,horizontal=T)
```




```
boxplot(T1,horizontal=T)
```


1.5
2.0
2.5
3.0
3.5
4.0
4.5