Probability and Statistics in R

Assignment 2 Problem 2

Problem 2 : Simulation Study to Understand Sampling Distribution

Part A Suppose $X_1, X_2, \cdots, X_n \overset{iid}{\sim} Gamma(\alpha, \sigma)$, with pdf as

$$f(x|\alpha, \sigma) = \frac{1}{\sigma^{\alpha}\Gamma(\alpha)} e^{-x/\sigma} x^{\alpha-1}, \quad 0 < x < \infty,$$

The mean and variance are $E(X) = \alpha\sigma$ and $Var(X) = \alpha\sigma^2$. Note that `shape` = α and `scale` = σ.

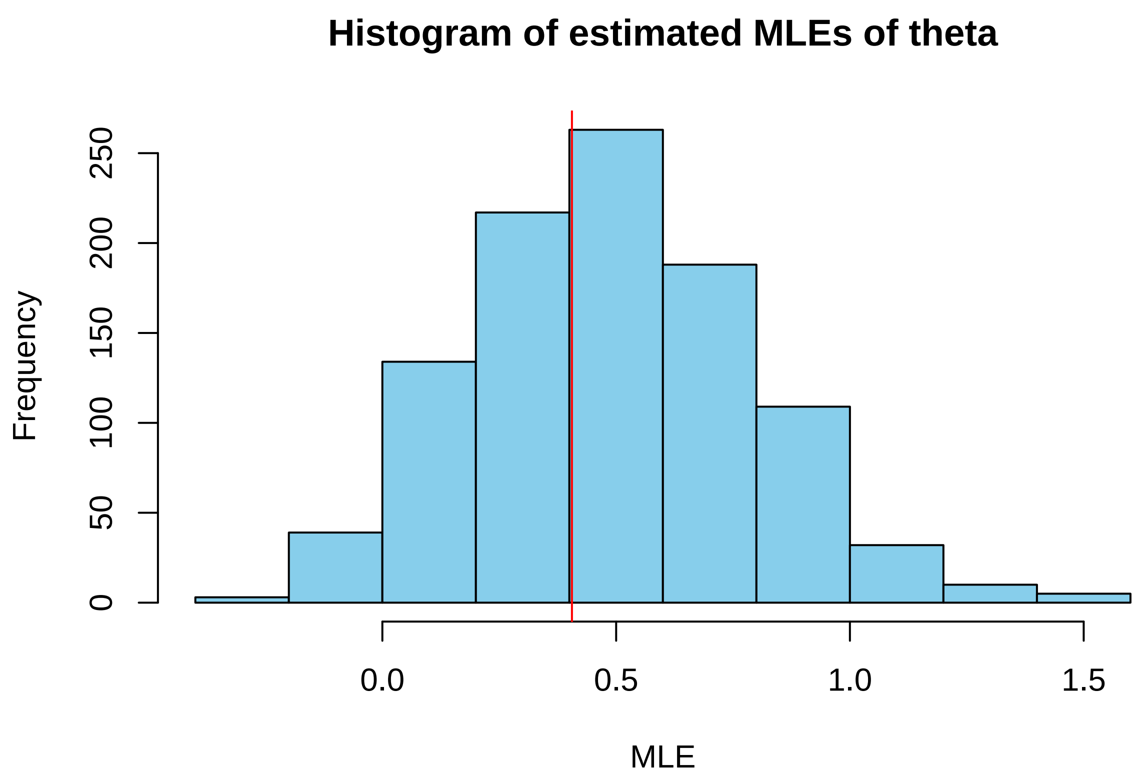
1. Write a `function` in `R` which will compute the MLE of $\theta = \log(\alpha)$ using `optim` function in `R`. You can name it `MyMLE`
2. Choose `n=20`, and `alpha=1.5` and `sigma=2.2`
   (i) Simulate $\{X_1, X_2, \cdots, X_n\}$ from `rgamma(n=20,shape=1.5,scale=2.2)`
   (ii) Apply the `MyMLE` to estimate θ and append the value in a vector
   (iii) Repeat the step (i) and (ii) 1000 times
   (iv) Draw histogram of the estimated MLEs of θ.
   (v) Draw a vertical line using `abline` function at the true value of θ.
   (vi) Use `quantile` function on estimated θ's to find the 2.5 and 97.5-percentile points.
3. Choose `n=40`, and `alpha=1.5` and repeat the (2).
4. Choose `n=100`, and `alpha=1.5` and repeat the (2).
5. Check if the gap between 2.5 and 97.5-percentile points are shrinking as sample size `n` is increasing?

```
simulation <- function(n, alpha, sigma, N){

  theta = log(alpha)

  logl <- function(par, sample){
    loglike = sum(dgamma(sample, shape = par[1], scale = par[2], log = T))
    return(-loglike)
  }

  mymles = c()
  for(i in 1:N){
    sample = rgamma(n = n, shape = alpha, scale = sigma)
    par = c(1.5, 2.2)
    fit = optim(par, logl, sample = sample)
    alpha_hat = fit$par[1]
    mymle = log(alpha_hat)
    mymles = c(mymles, mymle)
  }

  hist(mymles, col = "skyblue", xlab = "MLE",
       main = "Histogram of estimated MLEs of theta")
```

```
  abline(v = theta, col = "red")
  quantiles = quantile(mymles, probs = c(0.025, 0.975))
  print("Quantiles")
  print.table(quantiles)
  diffq = quantiles[2] - quantiles[1]
  print(paste("Gap between 2.5 and 97.5-percentile points", diffq))

}
```

For sample size:20

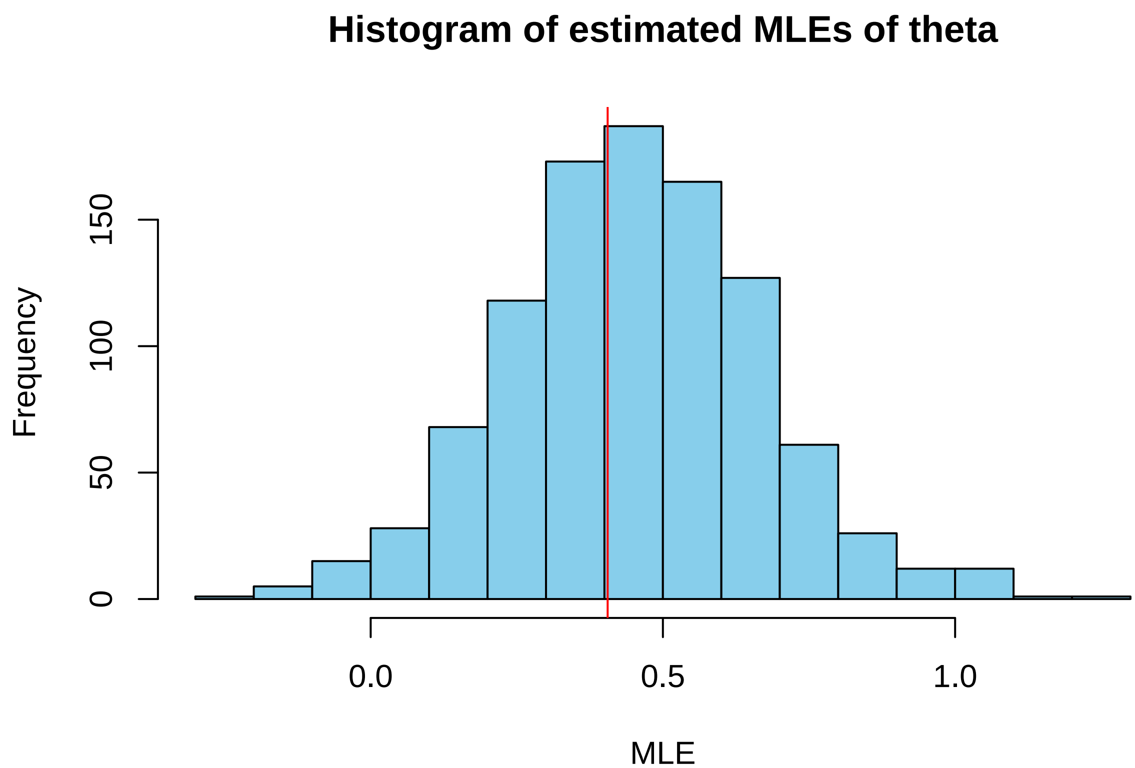
```
simulation(20, 1.5, 2.2, 1000)
```




```
## [1] "Quantiles"
##         2.5%       97.5%
## -0.05492501  1.11907380
## [1] "Gap between 2.5 and 97.5-percentile points 1.17399880177436"
```

For sample size:40

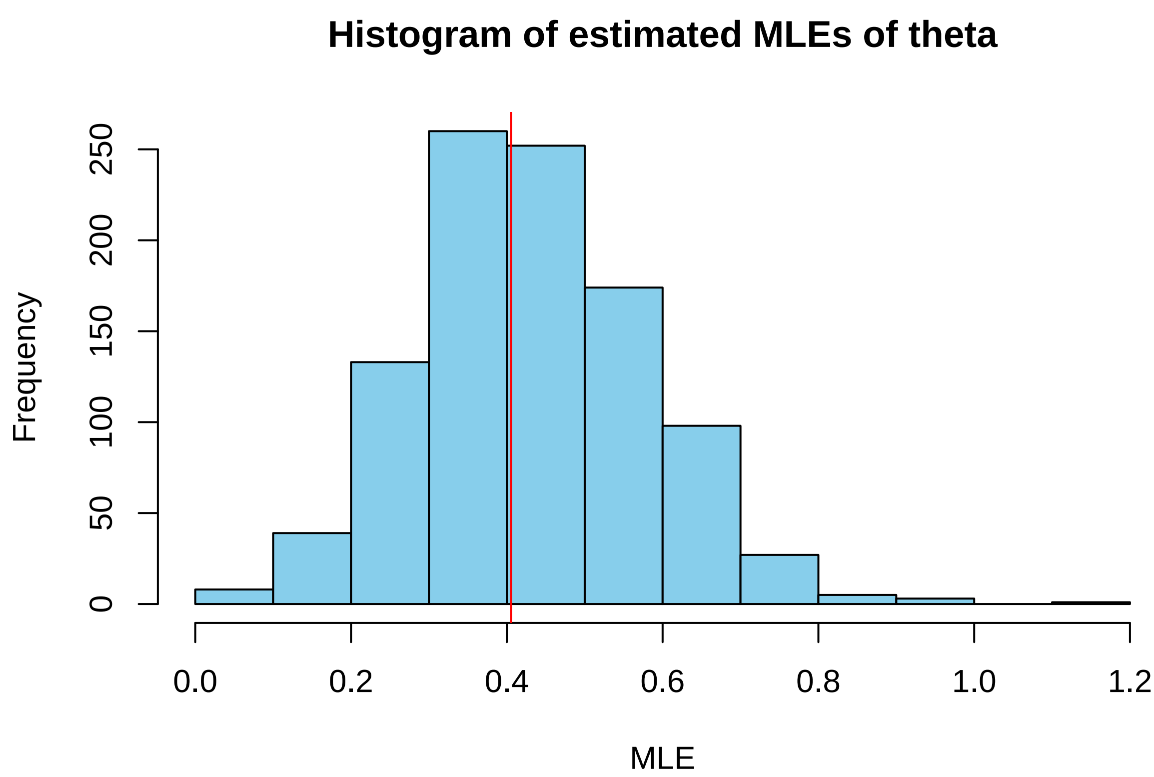
```
simulation(40, 1.5, 2.2, 1000)
```




```
## [1] "Quantiles"
##       2.5%      97.5%
## 0.02165716 0.91477201
## [1] "Gap between 2.5 and 97.5-percentile points 0.893114851766961"
```

For sample size:80

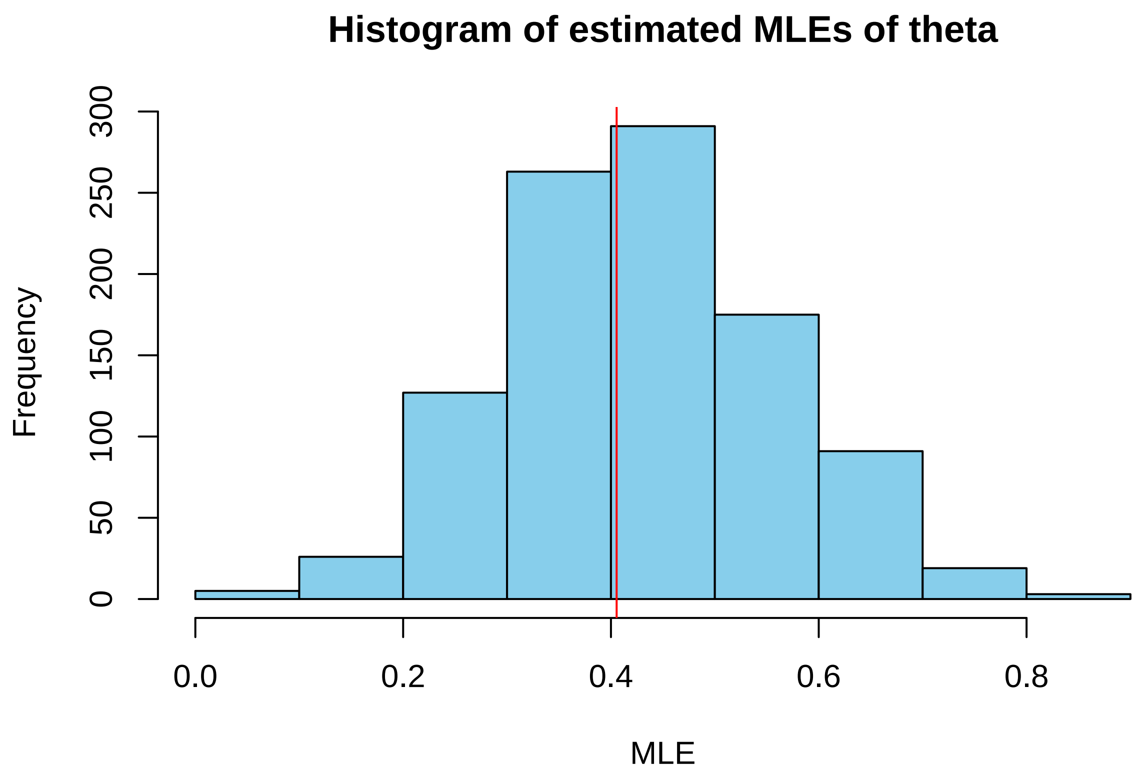
```
simulation(80, 1.5, 2.2, 1000)
```




```
## [1] "Quantiles"
##      2.5%     97.5%
## 0.1611103 0.7272741
## [1] "Gap between 2.5 and 97.5-percentile points 0.566163808405529"
```

For sample size:100 and simulation size:1000

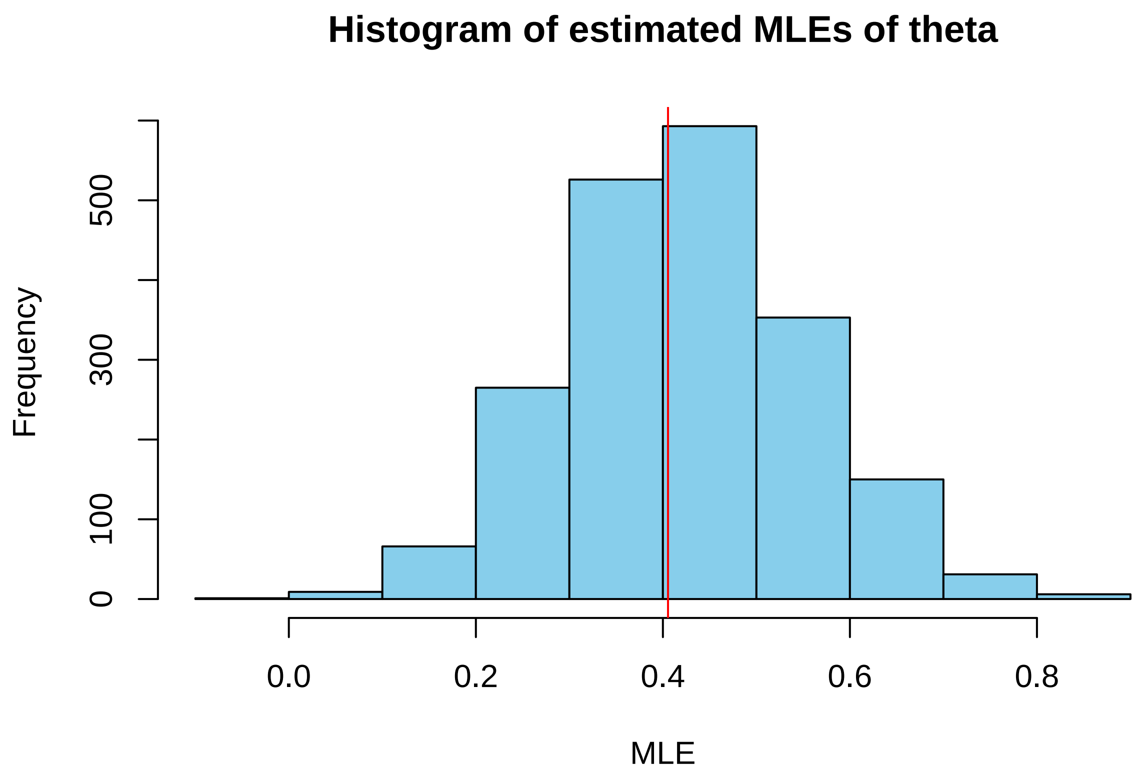
```
simulation(100, 1.5, 2.2, 1000)
```




```
## [1] "Quantiles"
##      2.5%     97.5%
## 0.1883622 0.6930597
## [1] "Gap between 2.5 and 97.5-percentile points 0.504697483725652"
```

For sample size:100 and simulation size:2000

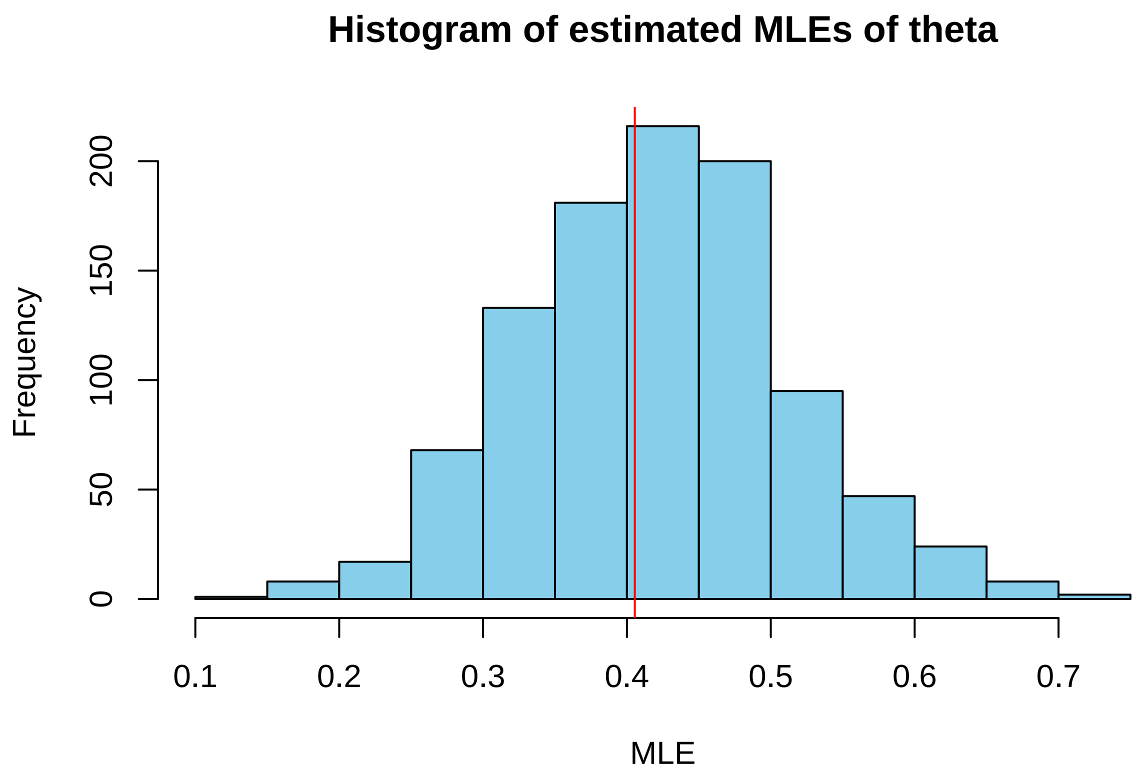
```
simulation(100, 1.5, 2.2, 2000)
```




```
## [1] "Quantiles"
##      2.5%     97.5%
## 0.1783566 0.6820748
## [1] "Gap between 2.5 and 97.5-percentile points 0.503718141705482"
```

For sample size:200 and simulation size:1000

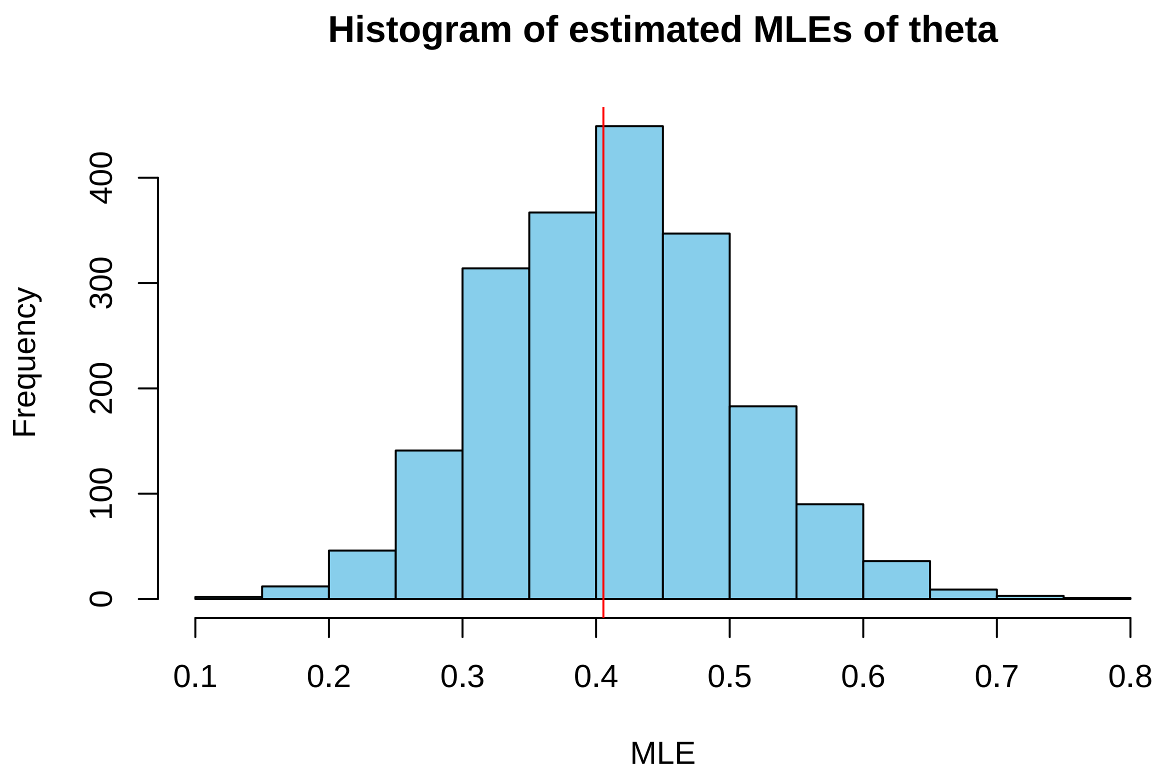
```
simulation(200, 1.5, 2.2, 1000)
```




```
## [1] "Quantiles"
##      2.5%     97.5%
## 0.2481390 0.6110496
## [1] "Gap between 2.5 and 97.5-percentile points 0.362910539582713"
```

For sample size:200 and simulation size:2000

```
simulation(200, 1.5, 2.2, 2000)
```




```
## [1] "Quantiles"
##      2.5%     97.5%
## 0.2457437 0.5971154
## [1] "Gap between 2.5 and 97.5-percentile points 0.35137171810132"
```

From the above results we can observe that as the sample size n is increasing, the gap between 2.5 and 97.5 percentile points is shrinking.