Question2

Varun Agrawal

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`

```
MyMLE <- function(x,alfa,beta){
  est <- function(x1 = x,para){
    est = -1*sum(dgamma(x1,shape = para[1], scale = para[2],log = T))
    return(est)
    }
  opt <- optim(par = c(alfa,beta), fn = est,x1 = x)
  return(log(opt$par[1]))
}
```

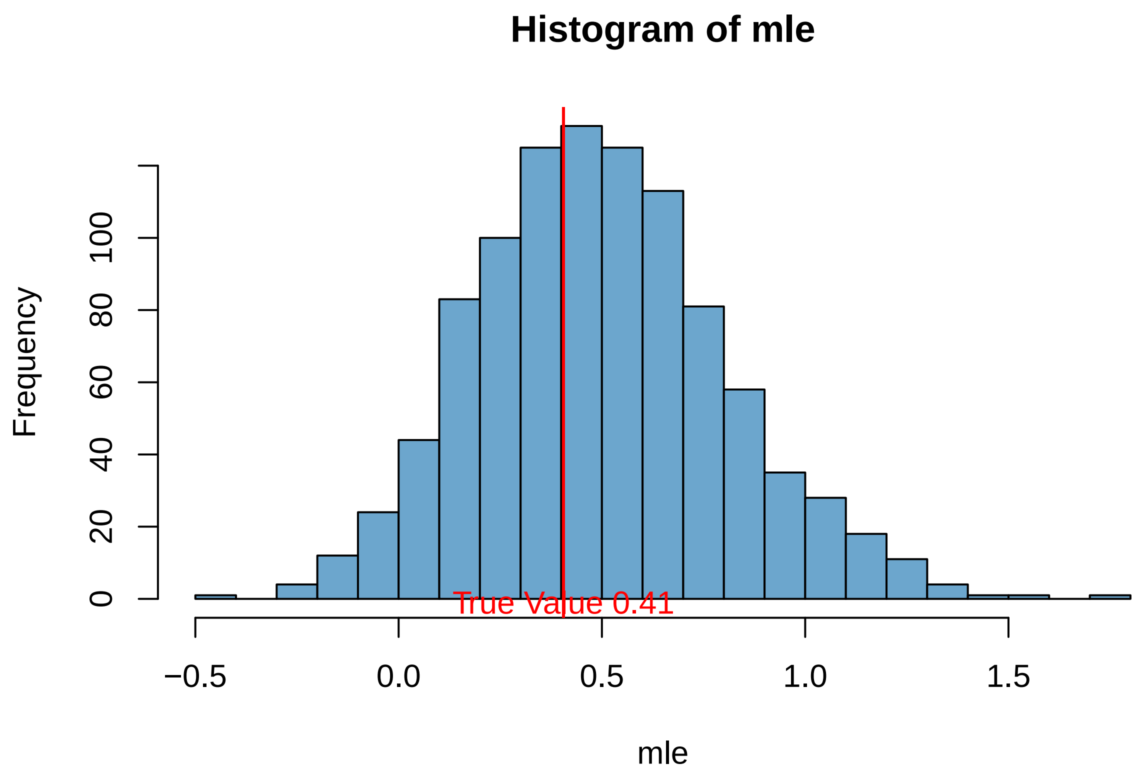
```
Q2_1 <- function(n,N,alfa,beta){
  mle <- c(rep(0,N))
  for(i in 1:N){
    x <- rgamma(n,shape = alfa, scale = beta)
    mle[i] = MyMLE(x,alfa,beta)
  }
  q2_5<- as.character(round(quantile(mle, 0.025),3))
  q97_5<- as.character(round(quantile(mle,0.975),3))
  diff <- as.character(round((quantile(mle,0.975) - quantile(mle, 0.025)),3))

  hist(mle, col = 'skyblue3', breaks = 25)
  abline(v = log(alfa),lwd = 1.5, col = 'red')
  s1 <- as.character(round(log(alfa),2))
  lbl_str <- paste('True Value', s1)
  text(log(alfa),-1,lbl_str,col = 'red')
  print(paste('2.5th quantile = ',q2_5))
  print(paste('97.5th quantile = ',q97_5))
  print(paste('97.5th qunatile - 2.5th quantile = ',diff))
  return(round((quantile(mle,0.975) - quantile(mle, 0.025)),3))
}
```

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.

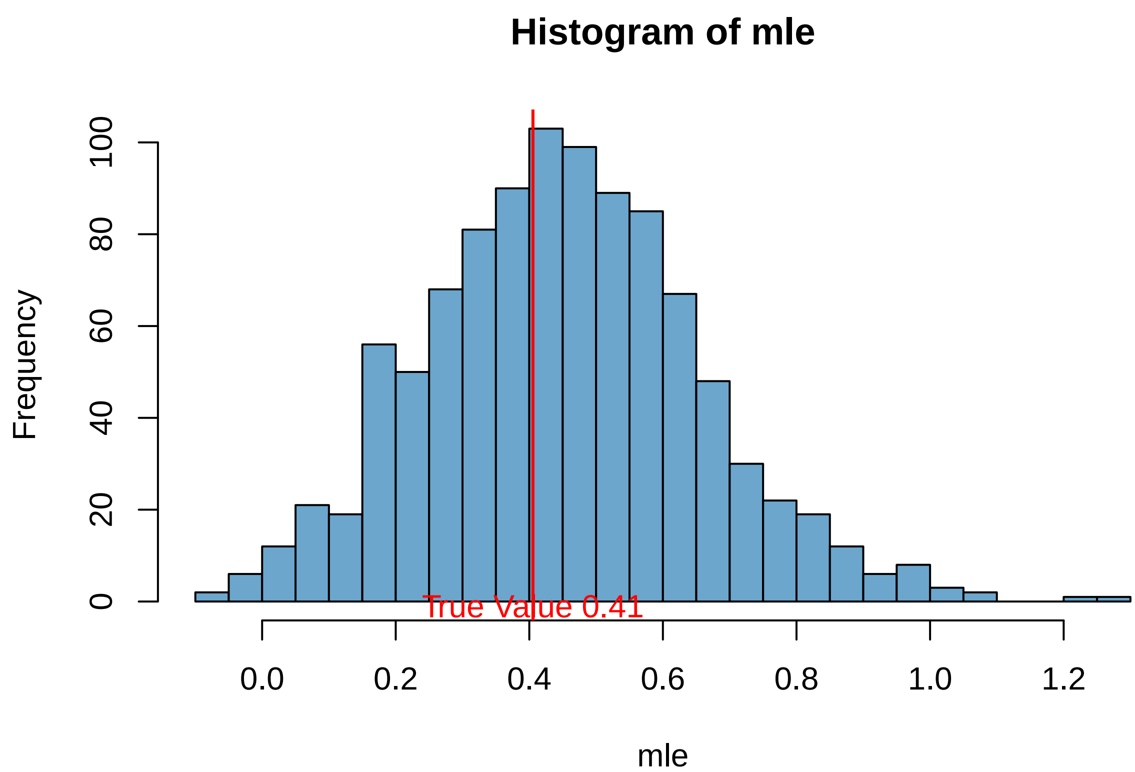
```
q2 <- Q2_1(20,1000,1.5,2.2)
```




```
## [1] "2.5th quantile =  -0.052"
## [1] "97.5th quantile =  1.149"
## [1] "97.5th qunatile - 2.5th quantile =  1.2"
```

3. Choose `n=40`, and `alpha=1.5` and repeat the (2).

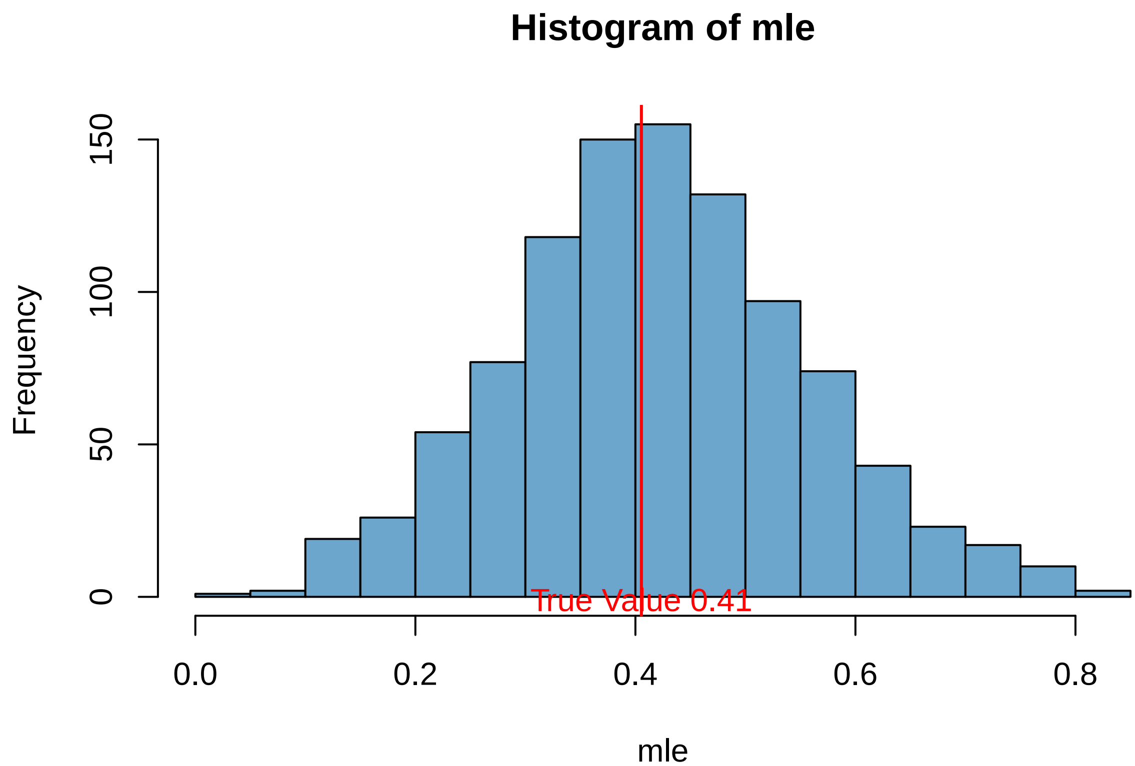
```r
q3 <- Q2_1(40,1000,1.5,2.2)
```




```
## [1] "2.5th quantile =  0.076"
## [1] "97.5th quantile =  0.875"
## [1] "97.5th qunatile - 2.5th quantile =  0.799"
```

4. Choose `n=100`, and `alpha=1.5` and repeat the (2).

```r
q4 <- Q2_1(100,1000,1.5,2.2)
```




```
## [1] "2.5th quantile =  0.163"
## [1] "97.5th quantile =  0.712"
## [1] "97.5th qunatile - 2.5th quantile =  0.549"
```

5. Check if the gap between 2.5 and 97.5-percentile points are shrinking as sample size `n` is increasing?

```
print(q2>q3 && q3>q4)
```

```
## [1] TRUE
```

Hence we can see that the gap between 2.5 and 97.5-percentile points are shrinking as sample size `n` is increasing.

Hint: Perhaps you should think of writing a single `function` where you will provide the values of `n`, `sim_size`, `alpha` and `sigma`; and it will return the desired output.