Question_2

Tanmoy Paul

2022-11-15

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

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

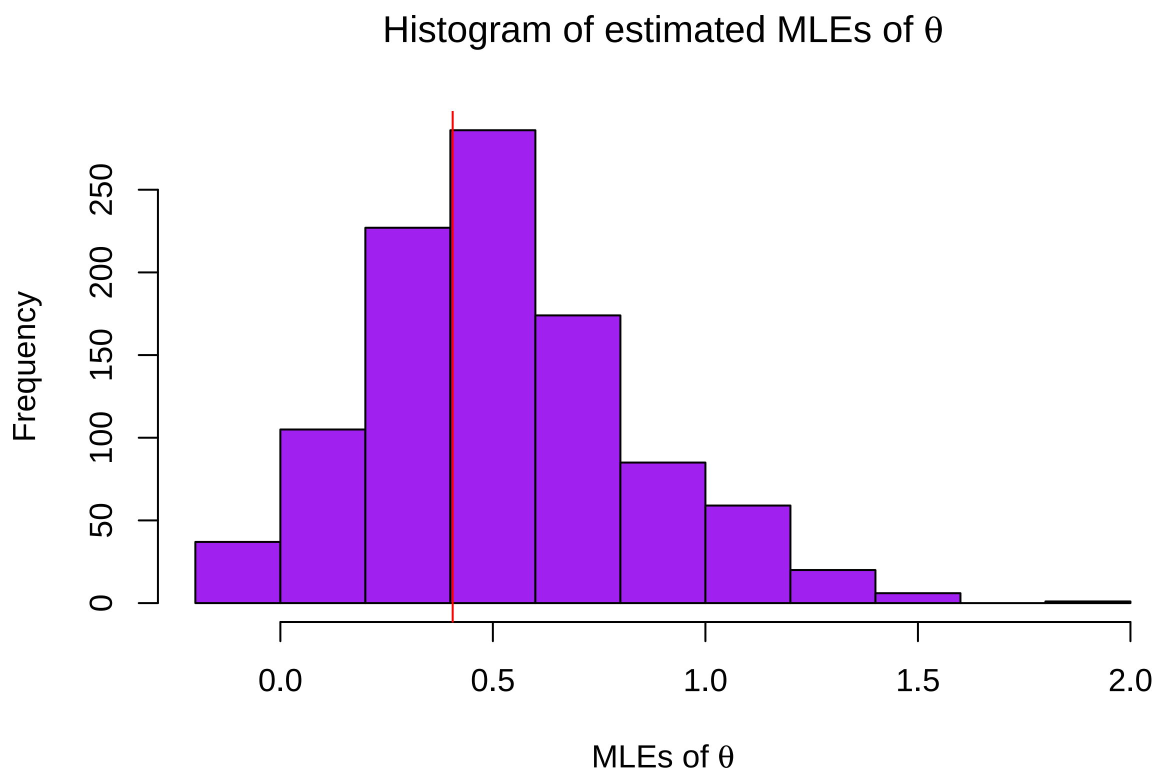
The mean and variance are $E(X) = \alpha\sigma$ and $Var(X) = \alpha\sigma^2$. Note that `shape` = α and `scale` = σ.

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?

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.

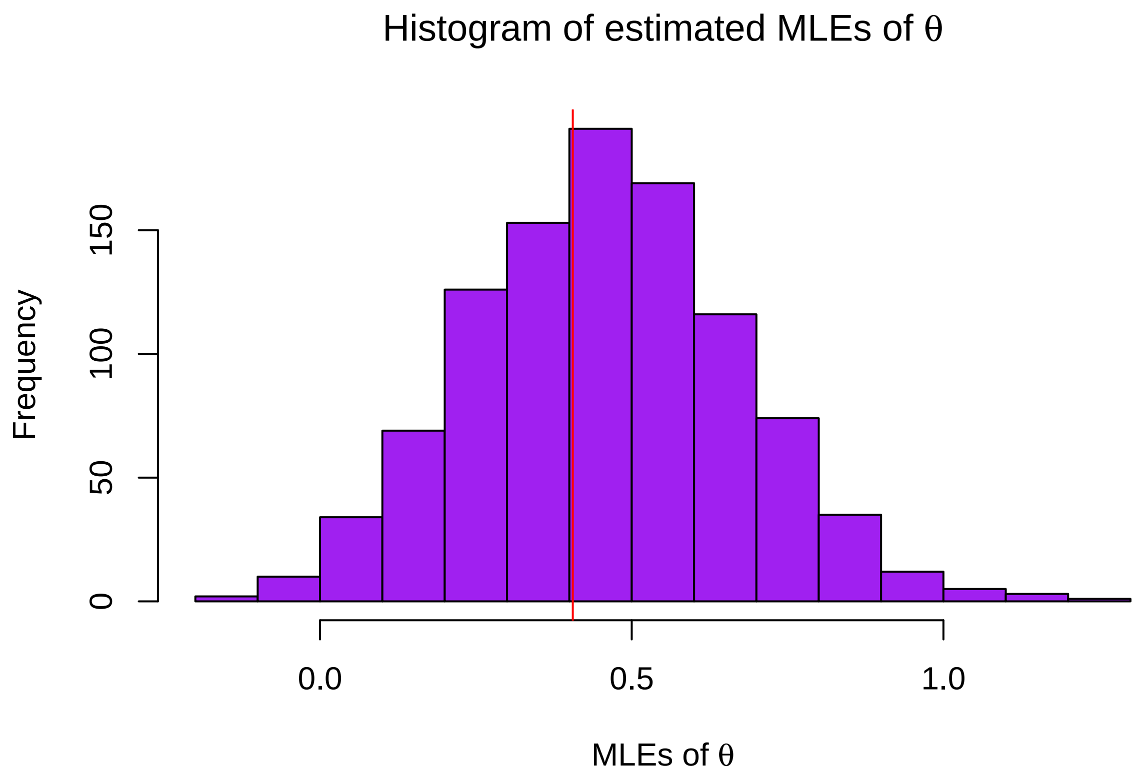
```
Question2<-function(n,sim_size,alpha,sigma)
{
  LogLikeFunc=function(data,para)
  {
    l =sum(dgamma(data,shape = para[1],scale = para[2],log=T))
    return(-l)
  }
  MyMLE=function(data){
    s=var(data)/mean(data)
    a=mean(data)/s
    initial=c(a,s)
    fit=optim(initial,LogLikeFunc,data=data)
    return(log(fit$par[1]))
  }
  v=c()
  for(i in 1:sim_size){
    x=rgamma(n=n,shape = alpha,scale=sigma)
    v=append(v,MyMLE(x),after = length(v))
  }
  hist(v,col="purple",main=expression(paste("Histogram of estimated MLEs of ",theta)),
```

```
        xlab=expression(paste("MLEs of ",theta)))
  abline(v=log(alpha),col="red")
  q=quantile(v,probs = c(2.5,97.5)/100)
  r=as.data.frame(q)
  d=r$q[2]-r$q[1]
  print('gap between 2.5 and 97.5 percentile points')
  d
}
Question2(20,1000,1.5,2.2)
```



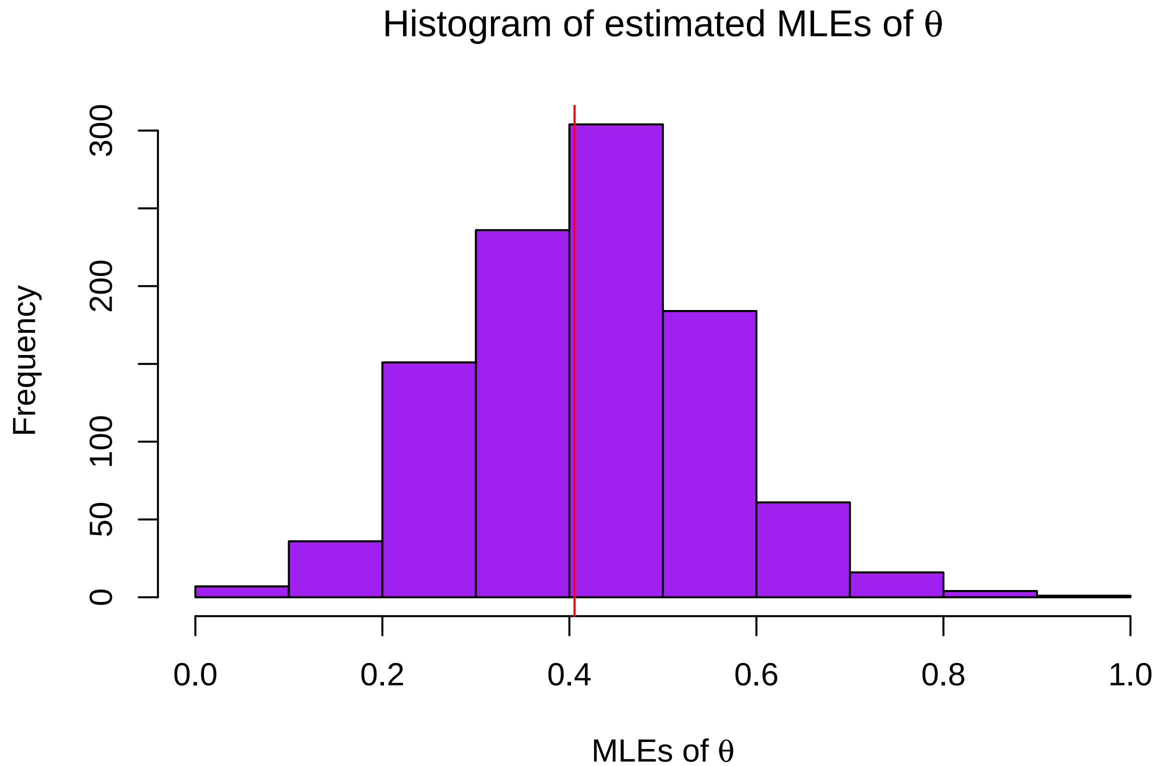

```
## [1] "gap between 2.5 and 97.5 percentile points"

## [1] 1.275892
```

```
Question2(40,1000,1.5,2.2)
```




```
## [1] "gap between 2.5 and 97.5 percentile points"
## [1] 0.8205174
```

```r
Question2(100,1000,1.5,2.2)
```




```
## [1] "gap between 2.5 and 97.5 percentile points"
## [1] 0.5193157
```

5. Yes,the gap between 2.5 and 97.5 percentile points are shrinking as sample size `n` is increasing