Statistics CH2

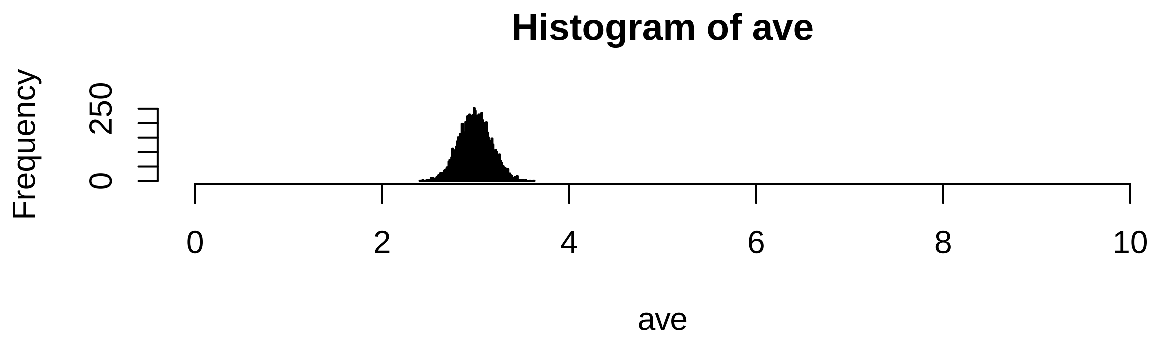
Xiang Li

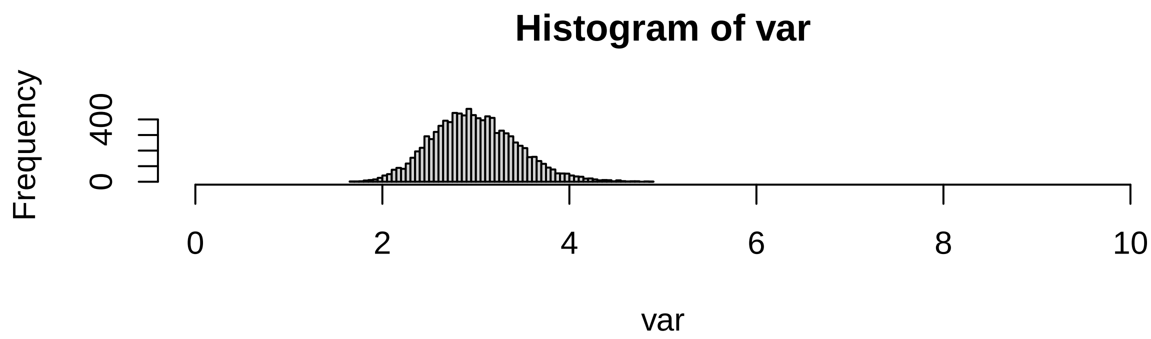
2023/10/19

2.1.0.3 Exercise

Strictly speaking, we only know now that the sample average is a much better estimator of λ than the sample variance when $n = 20$ and $\lambda = 5$. It is quite possible that that's not true for other values of n and λ. Make similar histograms as above but for different values of n and λ. What do you conclude?

```r
k = 10000
ave = var = numeric(k)
for (i in 1:k){
  x = rpois(n = 100,lambda = 3)  # data
  ave[i] = mean(x)          # sample average
  var[i] = var(x)           # sample variance
}
par(mfrow = c(2,1))
hist(ave,100,xlim=c(0,10))
hist(var,100,xlim=c(0,10))
```



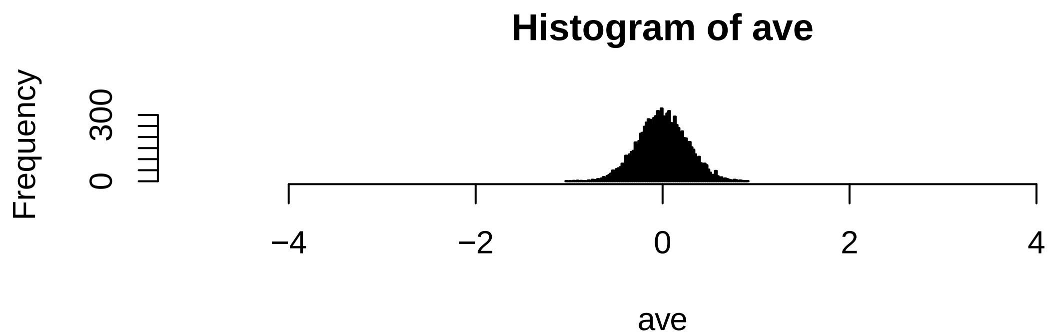


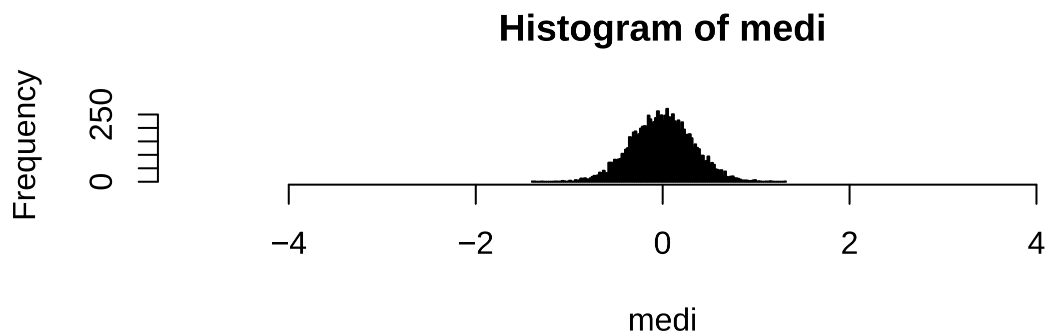
The sample average is better.

2.1.0.4 Exercise

Investigate if the sample average or the sample median is better. Use the approach we just applied to the Poisson distribution.

```
k = 10000
ave = medi = numeric(k)
for (i in 1:k){
  x = rnorm(n = 15, mean = 0, sd = 1)   # data
  ave[i] = mean(x)           # sample average
  medi[i] = median(x)            # sample median
}
par(mfrow = c(2,1))
hist(ave,100,xlim=c(-5,5))
hist(medi,100,xlim=c(-5,5))
```






```
mean(ave)
```

```
## [1] -0.001320008
```

```
sd(ave)
```

```
## [1] 0.2606139
```

```
mean(medi)
```

```
## [1] -0.003182591
```

```
sd(medi)
```

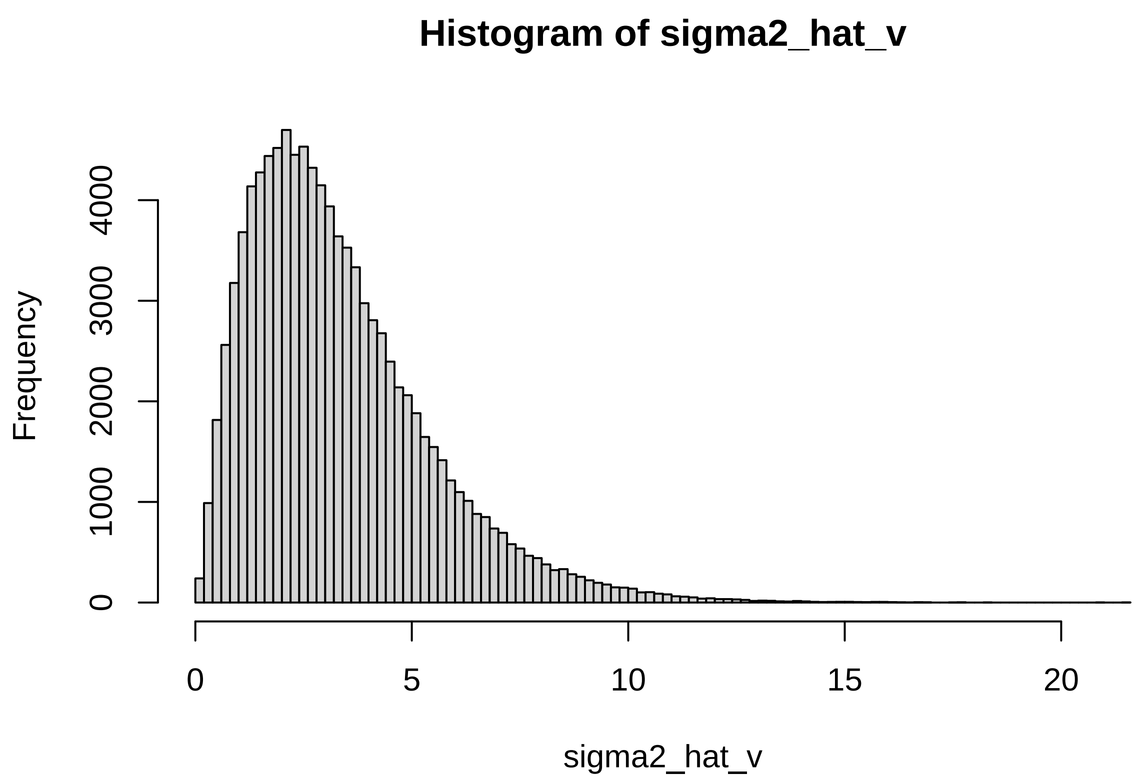
```
## [1] 0.3228413
```

Both ave and medi are unbiased. But ave is less variable.

2.2.0.1 Exercise

Verify numerically that this estimator is biased; it is systematically too small. Choose a sample of size $n = 6$ from the normal distribution with $\mu = 0$ and $\sigma^2 = 4$. Then check mean of the sampling distribution.

```r
sigma2_hat_v = replicate(1e5, {
  X = rnorm(6, mean = 0, sd = 2)
  sigma2_hat = mean((X-mean(X)) ** 2)
  return(sigma2_hat)
})
hist(sigma2_hat_v, 100)
```



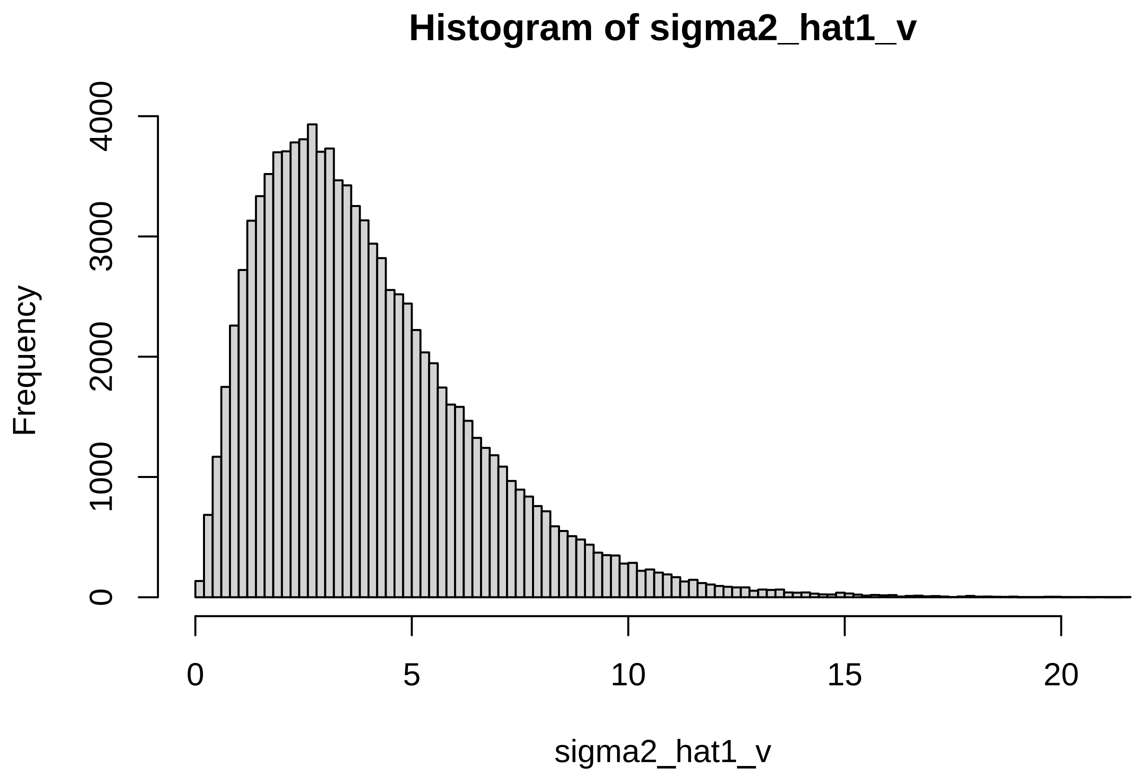

```r
mean(sigma2_hat_v)
```

```
## [1] 3.332767
```

2.2.0.2 Exercise

Verify numerically that S^2 is an unbiased estimator of the variance. Choose a sample of size $n = 6$ from the normal distribution with $\mu = 0$ and $\sigma = 2$. Then check the mean of the sampling distribution.

```r
sigma2_hat1_v = replicate(1e5, {
  X = rnorm(6, mean = 0, sd = 2)
  sigma2_hat1 = sum((X-mean(X)) ** 2)/(6-1)
  return(sigma2_hat1)
})
hist(sigma2_hat1_v, 100)
```




```
mean(sigma2_hat1_v)
```

```
## [1] 4.005883
```

2.2.0.4 Exercise (from Dekking et al)

Suppose the enemy has N=5000 tanks with serial numbers 1,2,. . . ,5000. We do not know N, but so far we have captured 10 tanks with numbers $X_1, X_2, ..., X_10$. Commander Bond of the secret service proposed the following estimator of N based on these numbers

$T_1 = 2\bar{X} - 1.$

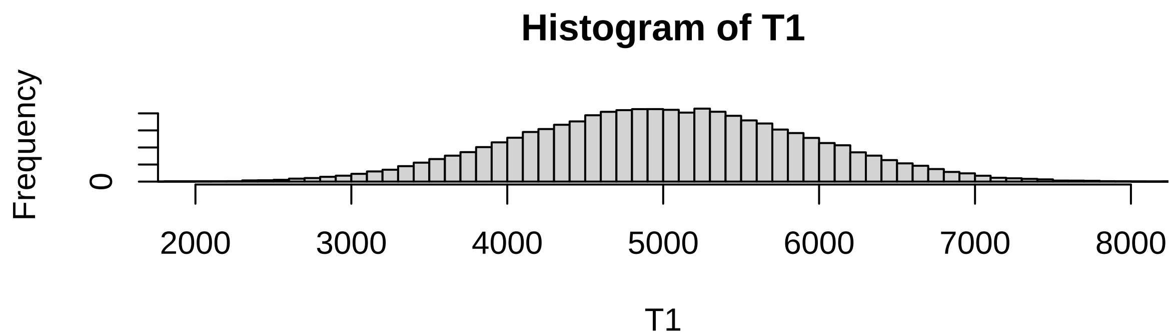
An unnamed person in department Q has a different proposal

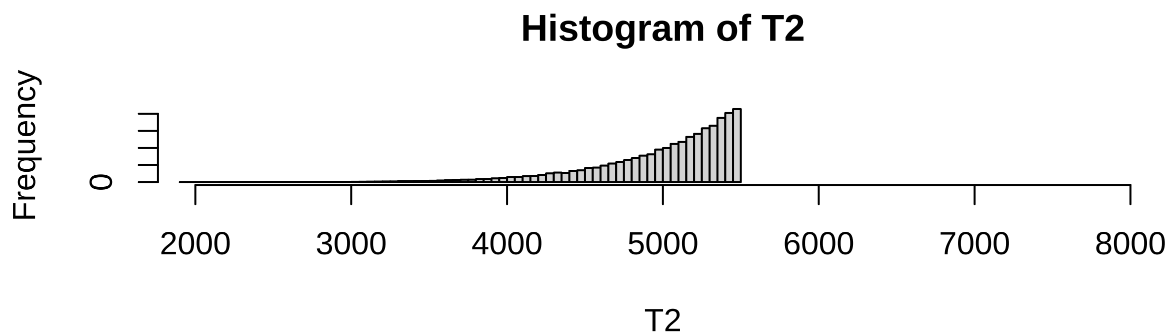
$T_2 = \frac{11}{10} max(X_1, X_2, ..., X_10) - 1$

a. Make histograms of the sampling distributions of T_1 and T_2. Use par(mfrow=c(2, 1)) to align them vertically. You can use the command sample(5000, 10) to sample 10 out of 5000 without replacement.

```
k = 1e5
T1 = numeric(k)
T2 = numeric(k)
for (i in 1:k) {
  X = sample(5000, 10)
  T1[i] = 2*mean(X)-1
```

```r
  T2[i] = 11/10*max(X)-1
}
par(mfrow = c(2, 1))
hist(T1, 100, xlim = c(2000, 8000))
hist(T2, 100, xlim = c(2000, 8000))
```





b. Both estimators T1 and T2 are unbiased. Computing the mean of their sampling distributions.

```r
mean(T1)
```

```
## [1] 4998.721
```

```r
mean(T2)
```

```
## [1] 4998.261
```

c. Which estimator has the smaller mean squared error?

```r
mean((T1-5000) ** 2)
```

```
## [1] 836742.4
```

```
mean((T2-5000) ** 2)
```

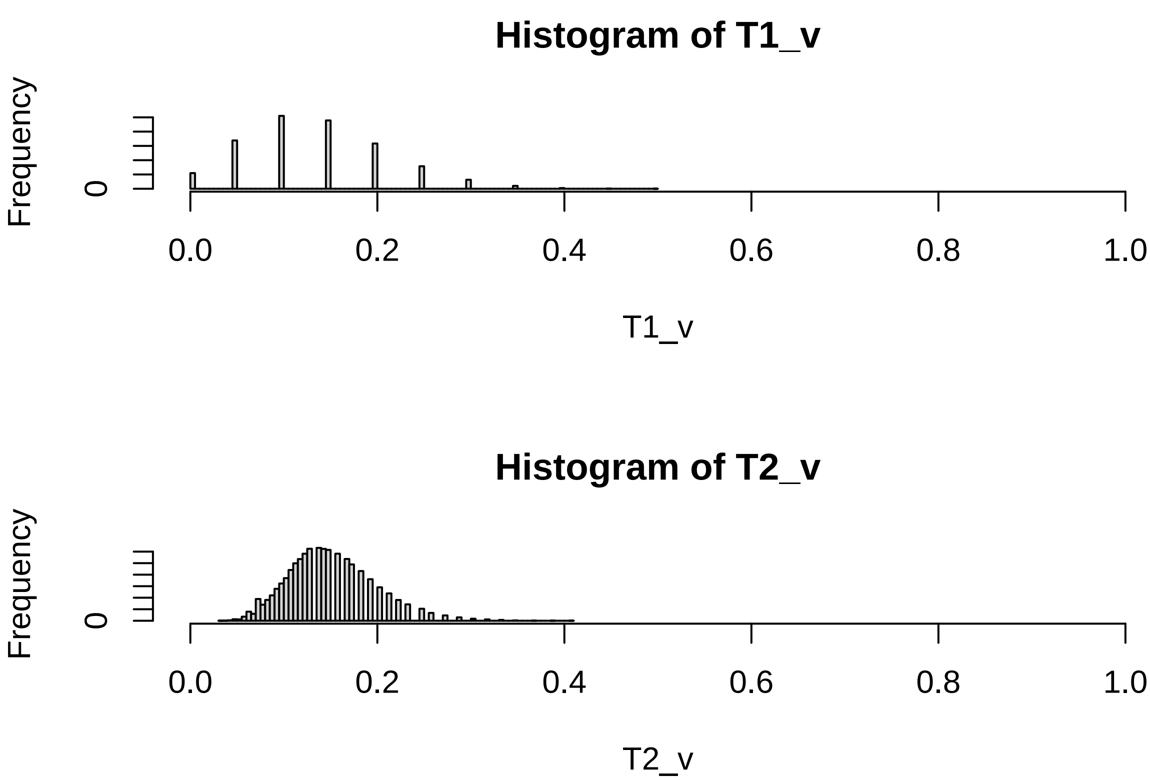
```
## [1] 208064.6
```

T_2 has smaller MSE.

2.2.0.5 Exercise (from Dekking et al)

b. Make histograms of the sampling distributions of the two estimators, and align them vertically.

```
k = 1e5
T1_v = T2_v = numeric(k)
for (i in 1:k) {
  X = rpois(20, 2)
  T1_v[i] = mean(X == 0)
  T2_v[i] = exp(-1*mean(X))
}
par(mfrow = c(2, 1))
hist(T1_v, 100, xlim = c(0, 1))
hist(T2_v, 100, xlim = c(0, 1))
```



c. Check that T_1 is unbiased, but T_2 is slightly biased.

```
p_0 = exp(-2)
mean(T1_v) - p_0
```

```
## [1] 9.421676e-05
```

```
mean(T2_v) - p_0
```

```
## [1] 0.006674715
```

d. Which estimator has the smaller MSE?

```
mse1 = mean(T1_v-p_0)
mse2 = mean(T2_v-p_0)
mse1
```

```
## [1] 9.421676e-05
```

```
mse2
```

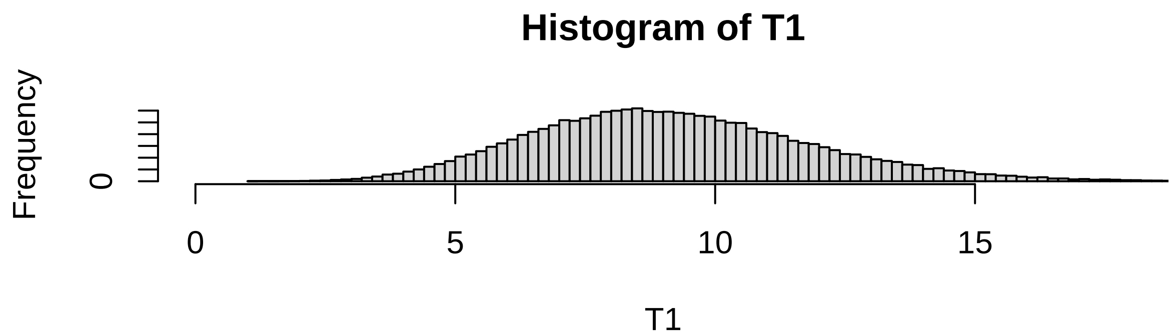
```
## [1] 0.006674715
```

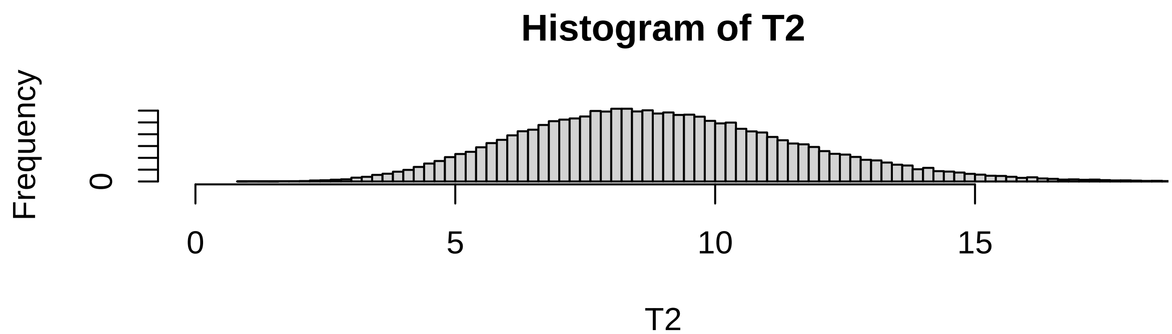
T_1 has smaller MSE.

2.2.0.6 Exercise

c. Use simulation to compare the bias and MSE of the unbiased estimator and $\bar{X}^2$.

```
k = 1e5
T1 = numeric(k)
T2 = numeric(k)
for (i in 1:k) {
  X = rnorm(20, mean = 3, sd = 2)
  T1[i] = mean(X) ** 2
  T2[i] = mean(X) ** 2 - var(X)/20
}
par(mfrow = c(2, 1))
hist(T1, 100, xlim = c(0, 18))
hist(T2, 100, xlim = c(0, 18))
```






```r
bias1 = mean(T1) - 3**2
bias1
```

```
## [1] 0.200764
```

```r
bias2 = mean(T2) - 3**2
bias2
```

```
## [1] 0.0007794478
```

```r
mse1 = mean((T1-3**2) ** 2)
mse1
```

```
## [1] 7.332434
```

```r
mse2 = mean((T2-3**2) ** 2)
mse2
```

```
## [1] 7.296134
```

2.2.0.7 Exercise

```r
sig2_hat = replicate(1e5, {
  X = rnorm(6, mean = 0, sd = 2)
  n = 6
  return(var(X)*(n-1)/n)
})
bias = mean(sig2_hat) - 2**2
mse = mean((sig2_hat - 2**2)**2)
n = length(sig2_hat)
var = var(sig2_hat)*(n-1)/n
c(mse, var+bias**2)
```

```
## [1] 4.899191 4.899191
```