A BRIEF STUDY ON
MAXIMUM LIKELIHOOD ESTIMATES
IN NON-CLOSED FORM

Sreejit Roy

425, 3rd year, Statistics.

Under the guidance of

Dr. Surabhi Dasgupta

I affirm that I have identified all my sources and that no part
Of my dissertation paper uses unacknowledged materials.

CONTENTS

1. INTRODUCTION

Maximum Likelihood Estimation is a method for estimating a parameter. In this method, we maximize the likelihood equation, defined below:

Definition 1.1 (Likelihood)

Let **X** be the realized value of a set of observations with joint density Pr(**X, θ**), where $\boldsymbol{\theta} \coloneqq (\theta_1, \theta_2, \theta_3, \ldots, \theta_p)$, the vector of parameters belongs to $\Theta \subseteq \mathbb{R}_p$. The likelihood equation of **θ**, given the observation **X** is defined as:

$$L(\boldsymbol{\theta}|\boldsymbol{X}) \propto Pr(\boldsymbol{X}, \boldsymbol{\theta})^1 \tag{1.1.1}$$

The principle of maximum likelihood is defined as follows:

Definition 1.2 (Principle of Maximum Likelihood(m.l.))

Accept $\widehat{\boldsymbol{\theta}} = (\widehat{\theta_1}, \widehat{\theta_2}, \widehat{\theta_3}, \ldots, \widehat{\theta_p})$ as the estimate of **θ**, where,

$$L(\widehat{\boldsymbol{\theta}}|\boldsymbol{X}) = \sup_{\theta \in \Theta} L(\boldsymbol{\theta}|\boldsymbol{X}) \tag{1.1.2}$$

Definition 1.3 (Maximum Likelihood Estimator (m.l.e.))

$\widehat{\boldsymbol{\theta}}$**,** as in definition 1.2 is called so.

[1] This definition is in accordance with C.R. Rao (Reference 1). Some authors use equality in equation (1.1.1). That is, they take the proportionality constant to be 1.

Although, practically, it is easier to work with the log-likelihood function,

$$l(\boldsymbol{\theta}|\boldsymbol{X}) := \log L(\boldsymbol{\theta}|\boldsymbol{X}) \tag{1.1.3}$$

Since logarithm is monotonically nondecreasing, it is obvious that $\widehat{\boldsymbol{\theta}}$, as defined in definition 1.3, satisfies the equation

$$l(\widehat{\boldsymbol{\theta}}|\boldsymbol{X}) = \sup_{\theta \in \Theta} l(\boldsymbol{\theta}|\boldsymbol{X}) \tag{1.1.4}$$

When l(**θ**|**X**) is a differentiable function of **θ**, and the supremum $\widehat{\boldsymbol{\theta}}$ is attained at an interior point of Θ, the partial derivatives vanish at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. I.e., $\widehat{\boldsymbol{\theta}}$ is indeed a solution to the equations

$$\frac{\partial\, l(\boldsymbol{\theta}|\boldsymbol{X})}{\partial \theta_i} = 0 \quad \forall i = 1(1)p \tag{1.1.5}$$

Equations (1.1.5) are called maximum likelihood equations, and any solution of them are known as a maximum likelihood estimate.

Although the method of maximum likelihood is of tremendous importance, equations (1.1.5) are easy to solve mathematically only if they are in a closed form. Otherwise, we cannot easily evaluate m.l. estimate, given a likelihood equation.

In those situations, one may resort to numerical methods for evaluating the m.l. estimates. These solutions, although approximate, are indeed useful.

One such numerical method is Newton Raphson method, defined below:

Definition 1.4 (Newton Raphson Method)

For evaluating root of the equation

$$f(x) = 0 \tag{1.1.6}$$

provided an initial guess of the root is x_0, a better approximation of the root of equation (1.1.6) is given by:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \qquad (1.1.7)$$

Repeat this process until desired accuracy is achieved.

In general, for nth guess x_n, the (n+1)th guess is obtained using:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (1.1.8)$$

We use this method due to its greater applicability.

In this dissertation, we wish to consider some distributions, for which equations (1.1.5) are not in a closed form. We then use numerical methods to obtain an approximation of m.l.e., so that we can further study some properties exhibited by the numerical estimates.

2. METHODOLOGY

For each of the distributions considered, we proceed through the following steps:

1) Obtain equations (1.1.5) explicitly for the distribution concerned.
2) Generate a random sample[2] of size n for the given distribution with parameter $\boldsymbol{\theta}$.
3) Apply Newton Raphson method (definition 1.4) for obtaining estimate of $\widehat{\boldsymbol{\theta}}$.

It is obvious to question whether $\widehat{\boldsymbol{\theta}}$ is asymptotically unbiased. For that,

4) Increase n and perform step 3, until $\left|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right| < \varepsilon$, for some fixed $\varepsilon > 0$.

Another question that arises is whether the variance of the estimator attains the Fréchet Cramér Rao lower bound (FCRLB)[3] or not. For that,

5) We perform nonparametric bootstrapping[4] for variance as follows:
   i. Draw k resamples with replacement from the sample of size n.
   ii. Calculate $\widehat{\boldsymbol{\theta}}$ for each of the resamples.
   iii. Calculate the variance of the k $\widehat{\boldsymbol{\theta}}$s. This variance, say V_{bs}, is an estimate of the variance of the estimator.

[2] For generating a random sample in Python, NumPy library is useful.

[3] FCR Lower Bound is a lower bound for the variance of an estimator. See appendix A5.

[4] It is the practice of estimating properties of some estimator by measuring them while sampling from population under consideration.

6) Increase resampling size (k) and check closeness of V_{bs} and FCRLB of $\boldsymbol{\theta}$.

We will consider two distributions, Cauchy and Gamma, and for each of them, will consider ten different values of the parameter(s). For each of the distribution parameter pair, we will run through steps 1 to 6.

3. RESULTS AND DISCUSSION

We consider the distributions separately.

3.1 CAUCHY DISTRIBUTION

A Random Variable (RV) X is said to follow Cauchy distribution with parameters θ and σ if its probability density function (pdf) is of the form:

$$f(x) = \frac{\sigma}{\pi \{\sigma^2 + (x-\theta)^2\}} \qquad \theta \in \mathbb{R}, \sigma \in \mathbb{R}^+; x \in \mathbb{R} \tag{3.1.1}$$

We write X ~ Cauchy (θ, σ).

For the sake of simplicity, we take σ=1.

Clearly, for a random sample of size n, the likelihood function of θ is given by:

$$L(\theta) = \frac{1}{\pi^n \prod_{i=1}^{n} \{1 + (x_i - \theta)^2\}} \tag{3.1.2}$$

The log-likelihood function of θ is given by:

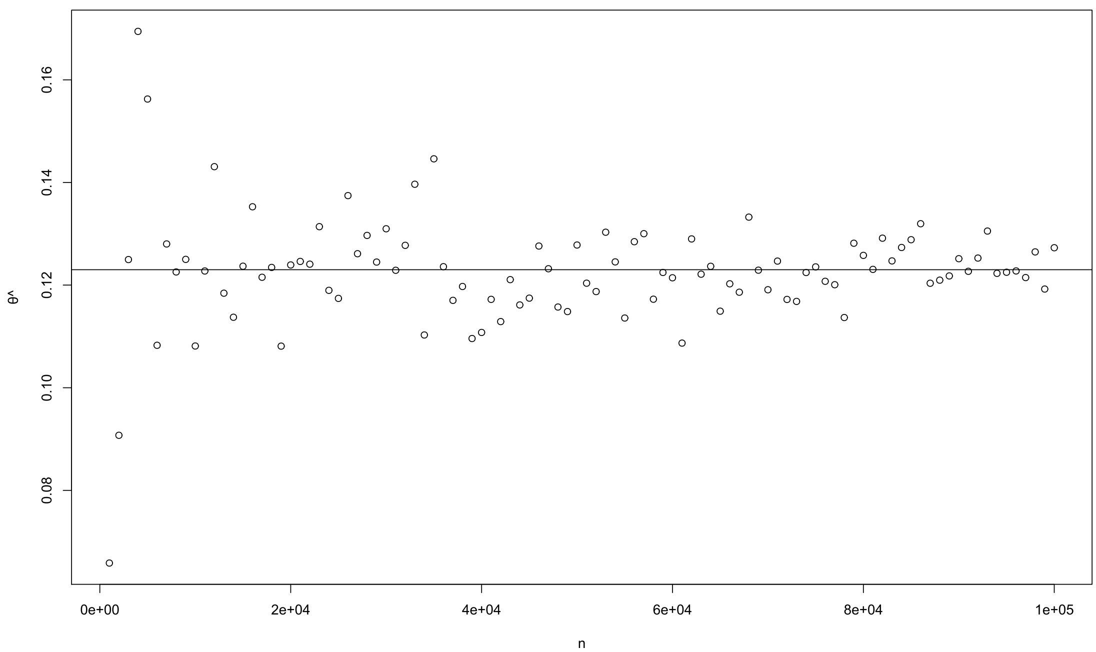
$$l(\theta) = -nlog\pi - \sum_{i=1}^{n} \{1 + (x_i - \theta)^2\} \tag{3.1.3}$$

Therefore, the maximum likelihood equation of θ is:

$$\sum_{i=1}^{n} \frac{(x_i - \theta)}{1 + (x_i - \theta)^2} = 0 \tag{3.1.4}$$

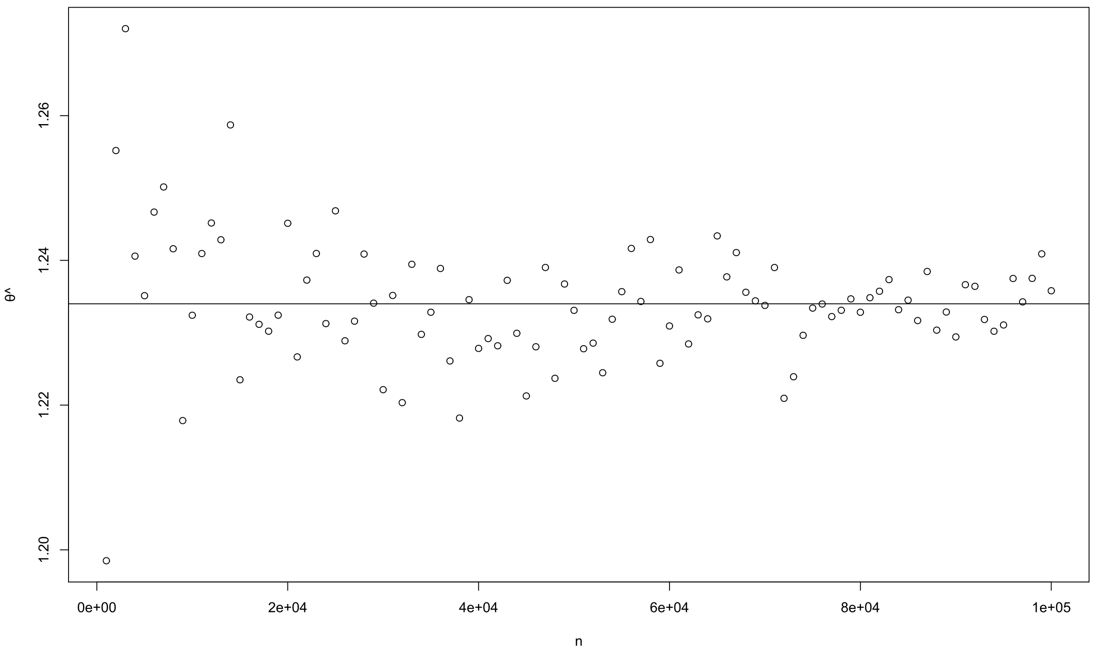
Equation (3.1.4) is not of closed form. We thus apply Newton Raphson method for estimating $\boldsymbol{\hat{\theta}}$.

For ten different values of θ, we estimate $\boldsymbol{\hat{\theta}}$ using Newton Raphson method[5] for hundred different sample sizes (n=1000(1000)100000). we obtain the following graphs through simulation:

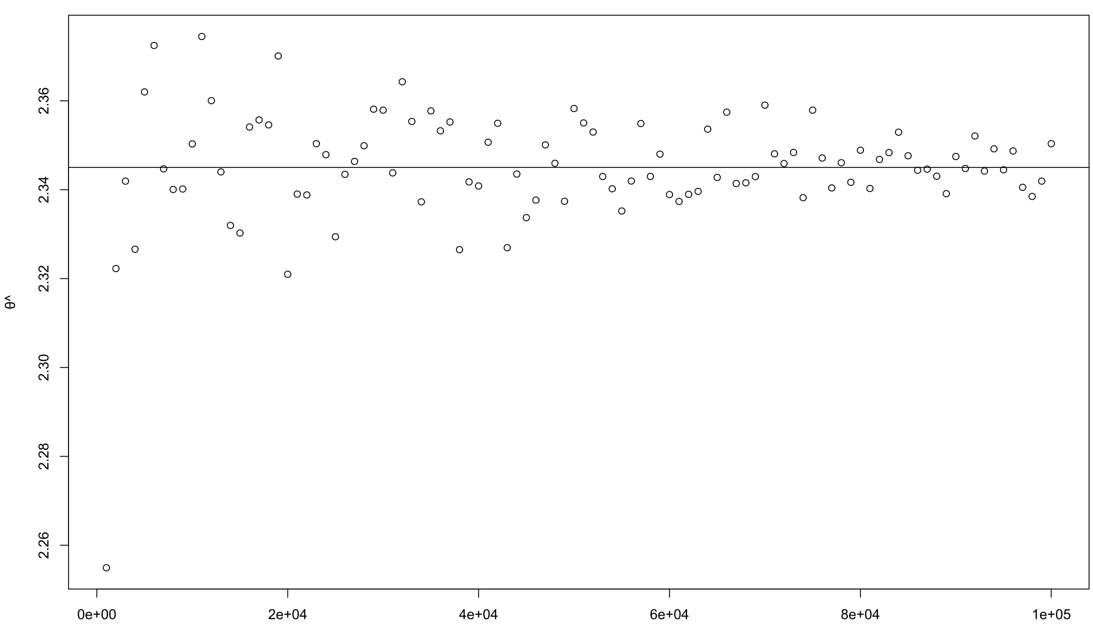


(figure 3.1.1) Samples from Cauchy (0.123, 1)

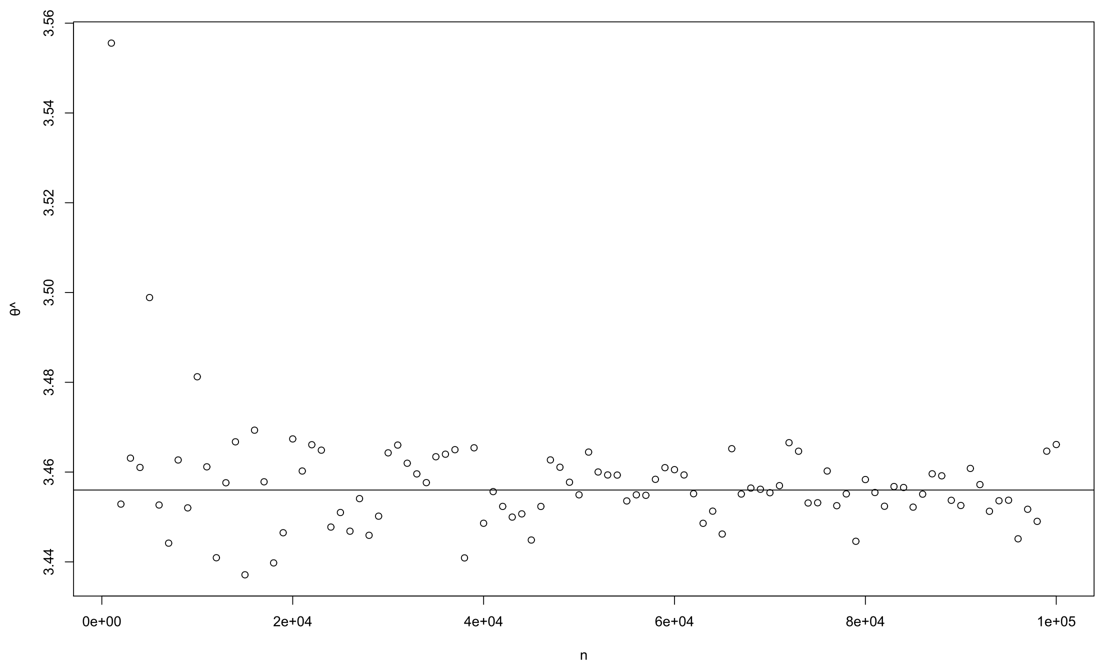
[5] For relevant R code, refer to appendix A1.



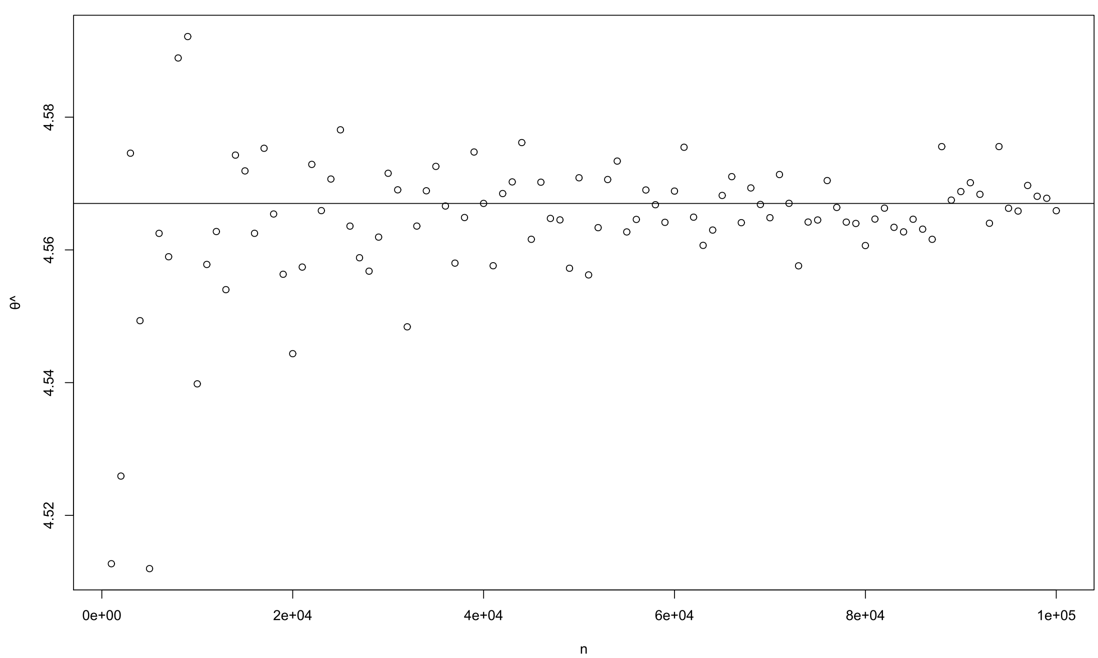
(figure 3.1.2) Samples from Cauchy (1.234, 1)



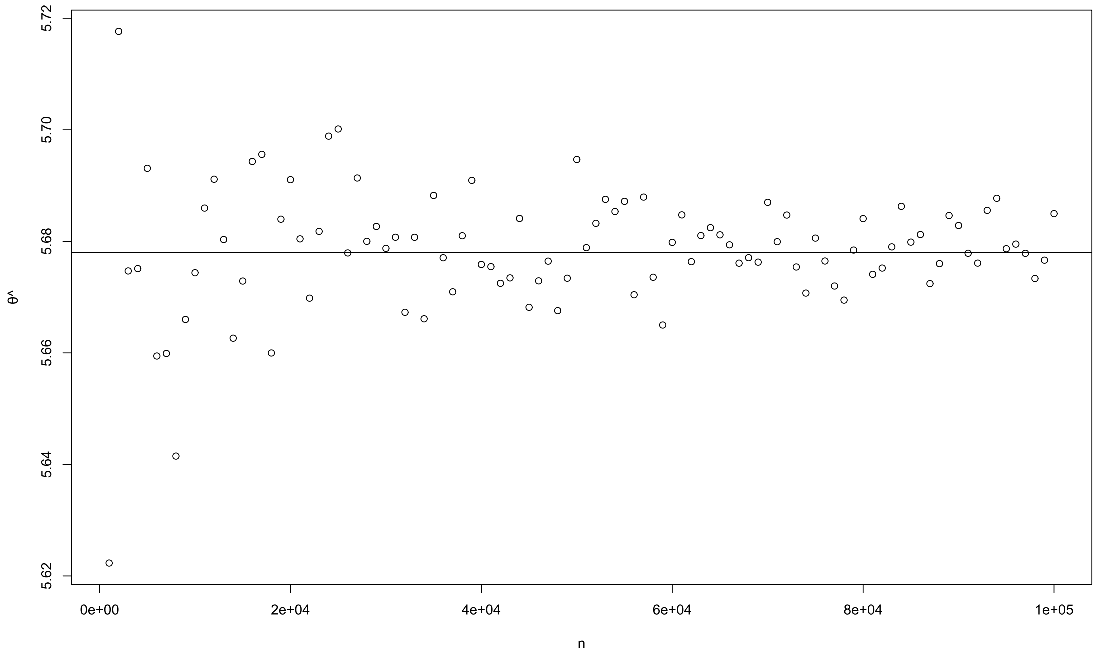
(figure 3.1.3) Samples from Cauchy (2.345, 1)



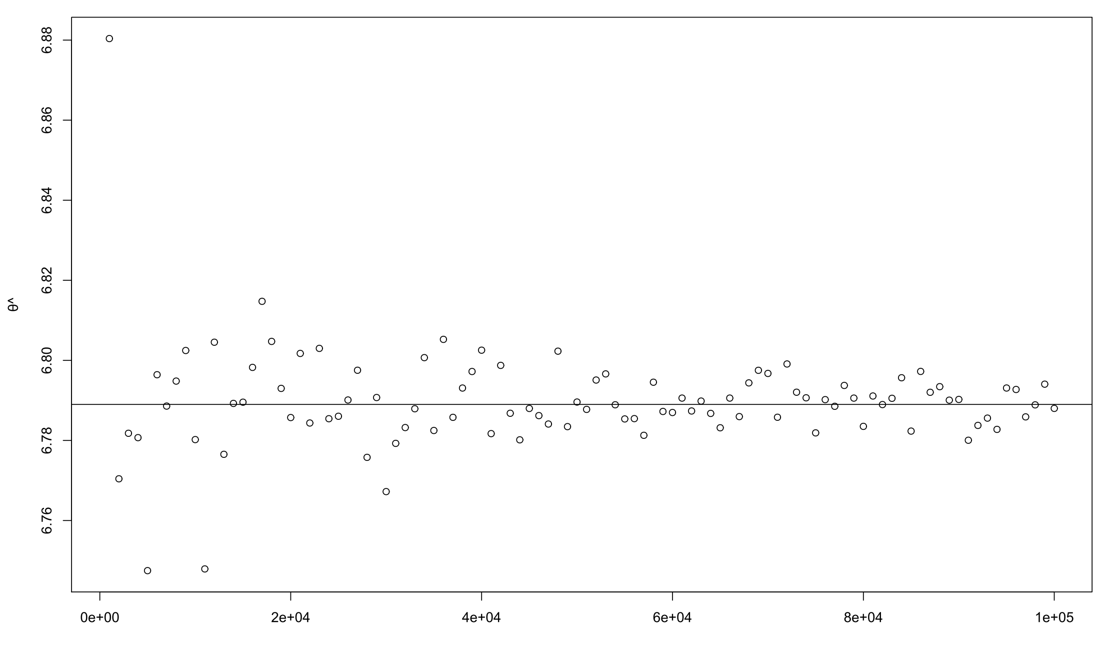
(figure 3.1.4) Samples from Cauchy (3.456, 1)



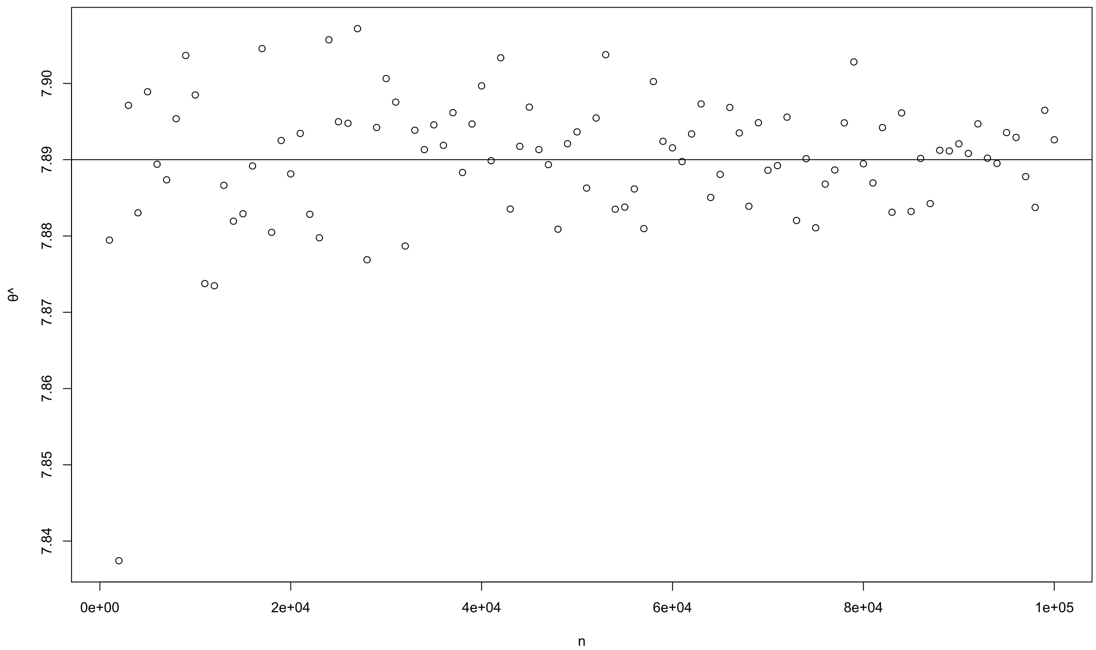
(figure 3.1.5) Samples from Cauchy (4.567, 1)



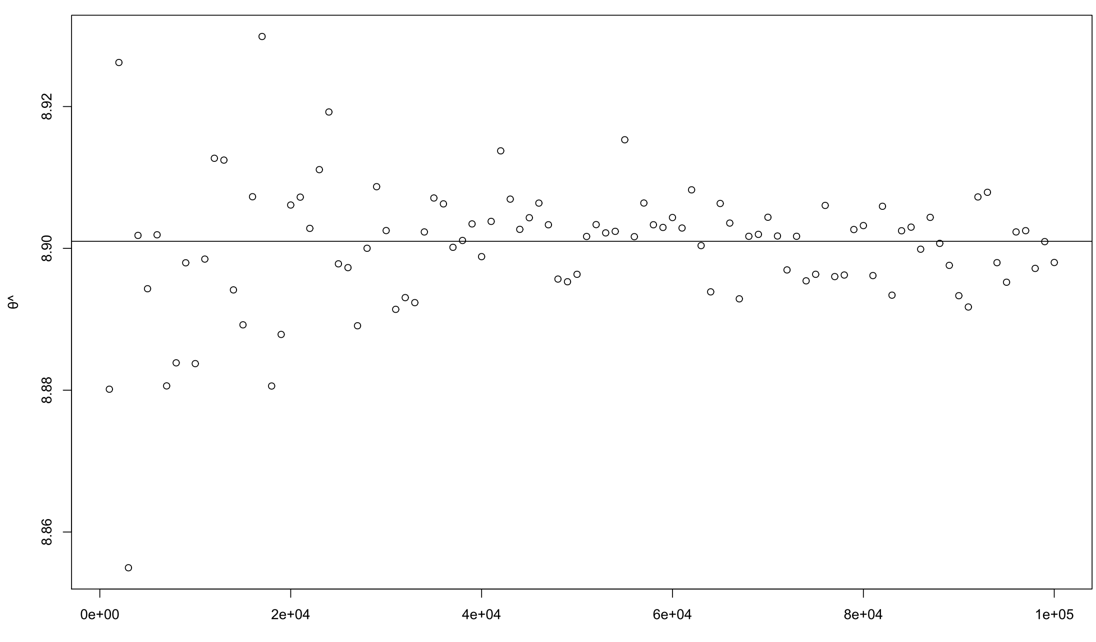
(figure 3.1.6) Samples from Cauchy (5.678, 1)



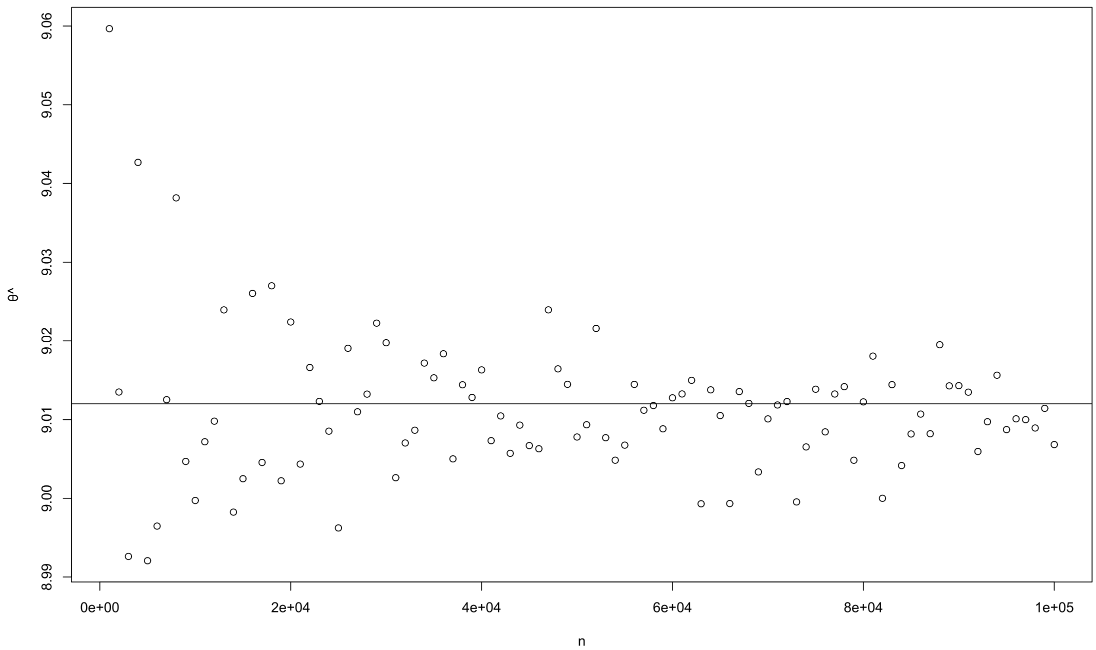
(figure 3.1.7) Samples from Cauchy (6.789, 1)



(figure 3.1.8) Samples from Cauchy (7.89, 1)



(figure 3.1.9) Samples from Cauchy (8.901, 1)



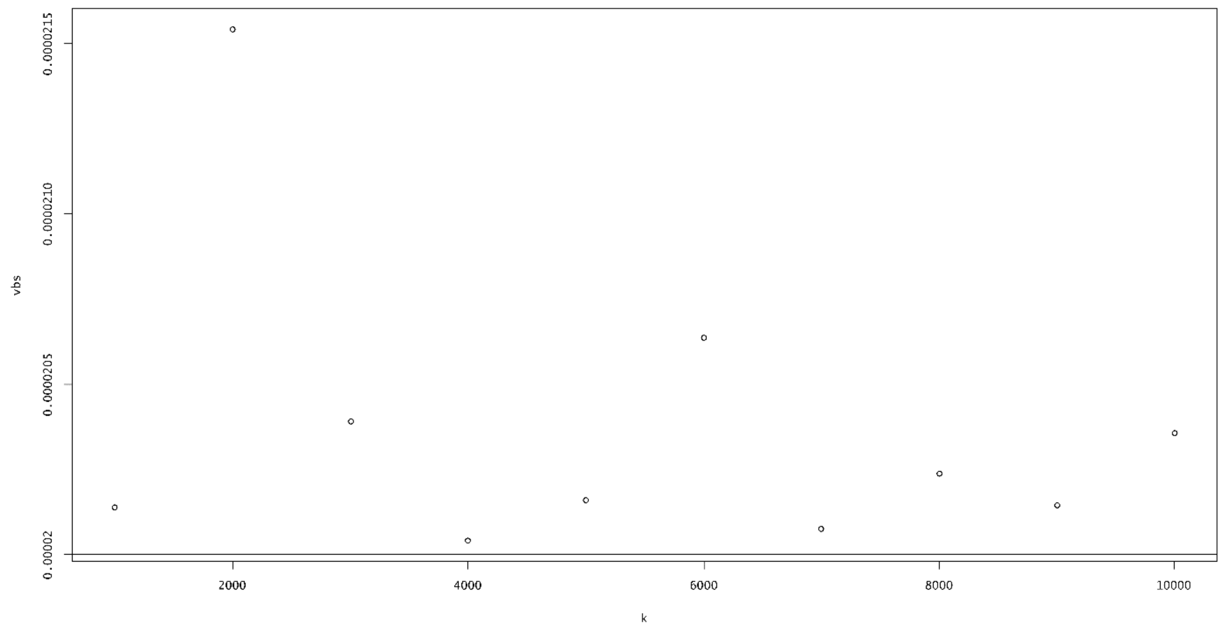
(figure 3.1.10) Samples from Cauchy (9.012, 1)

It is clear from the figures that the estimates of $\widehat{\boldsymbol{\theta}}$, obtained using Newton Raphson method, are asymptotically unbiased.

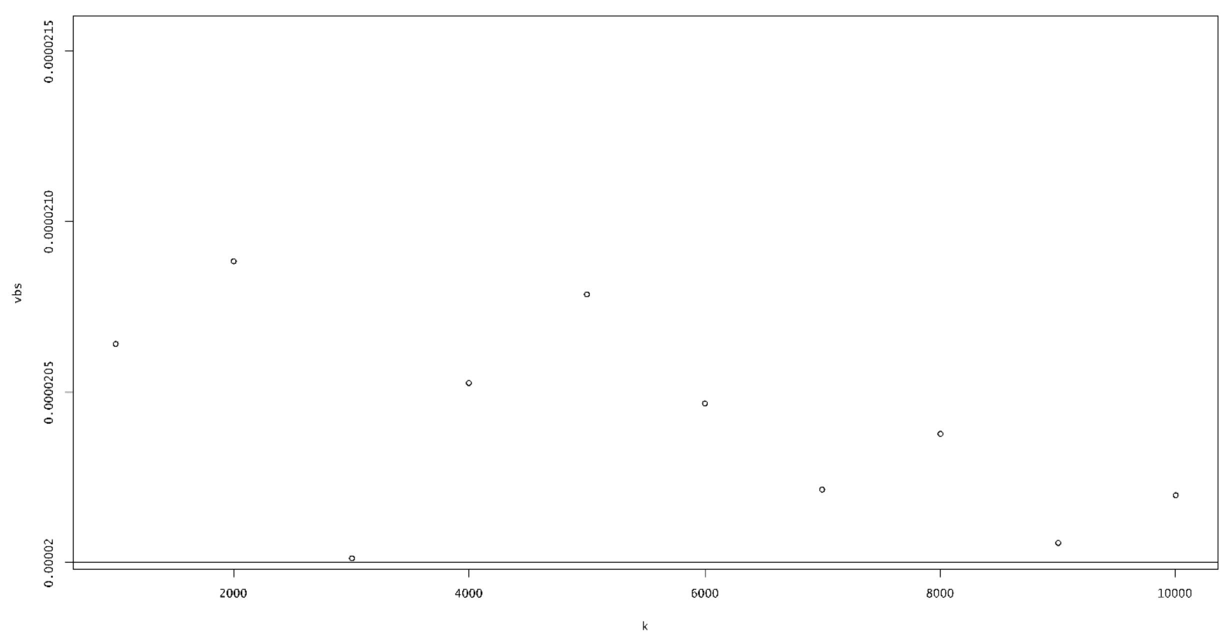
We now perform nonparametric bootstrapping for estimating the variance of the estimator.

For ten different values of θ, we estimate $var(\widehat{\boldsymbol{\theta}})$ using Nonparametric Bootstrapping[6] for ten different resample sizes (k=1000(1000)10000). we obtain the following graphs through simulation:

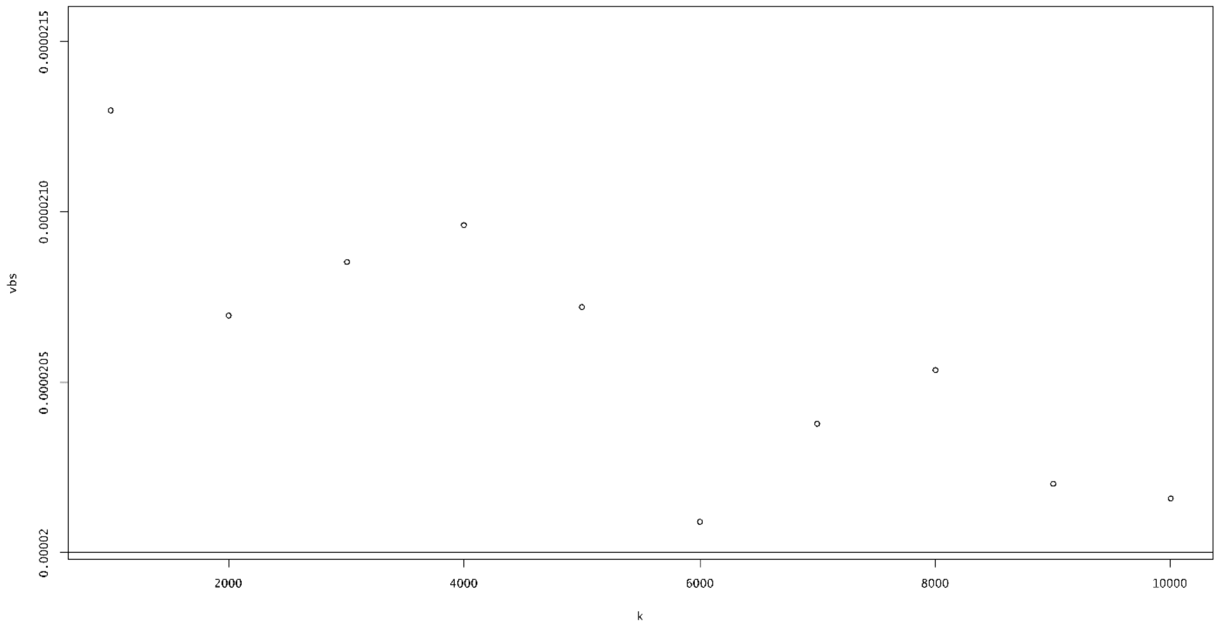
[6] For relevant R code, refer to appendix A2.



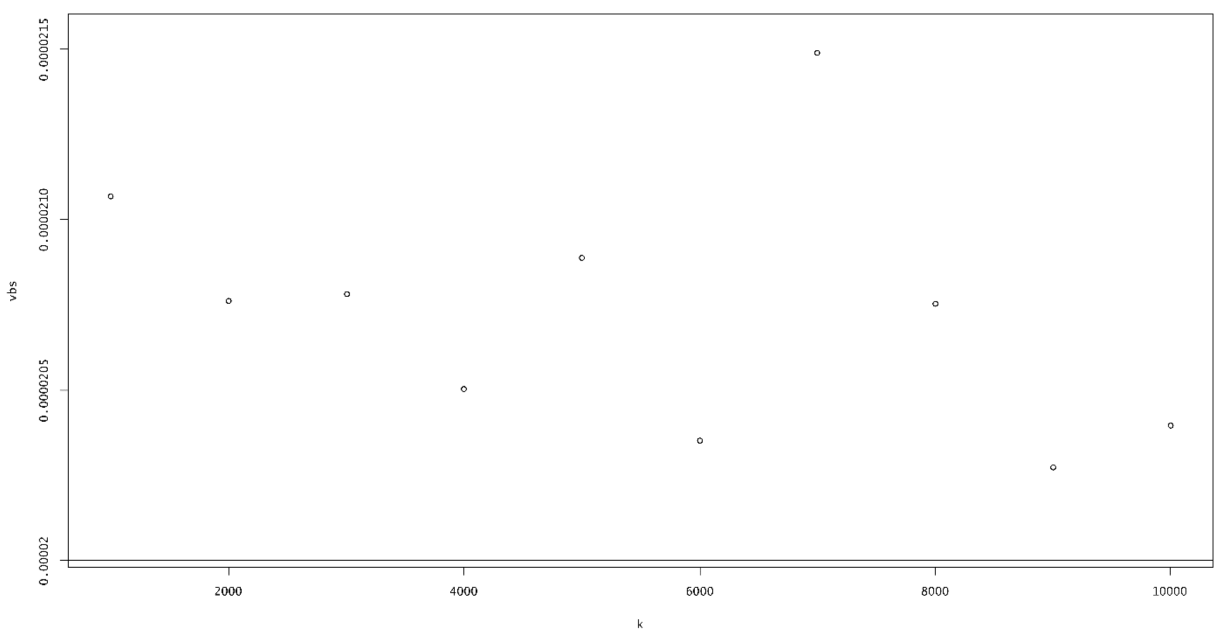
(figure 3.1.11) Variance for different resamples from Cauchy (0.123, 1)



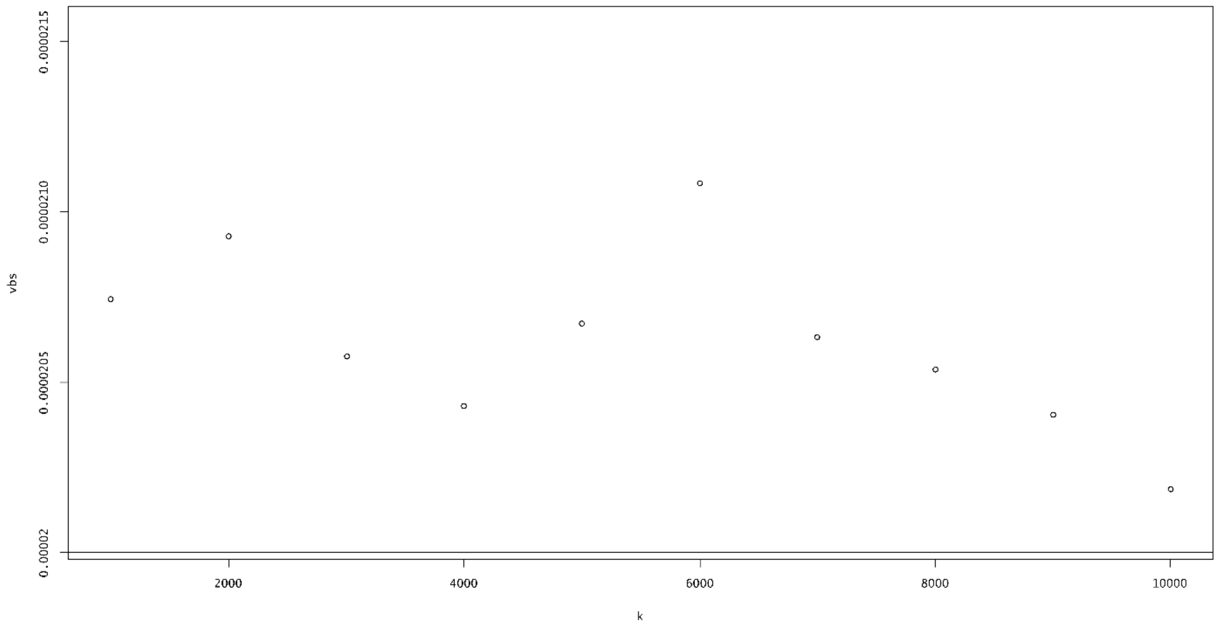
(figure 3.1.12) Variance for different resamples from Cauchy (1.234, 1)



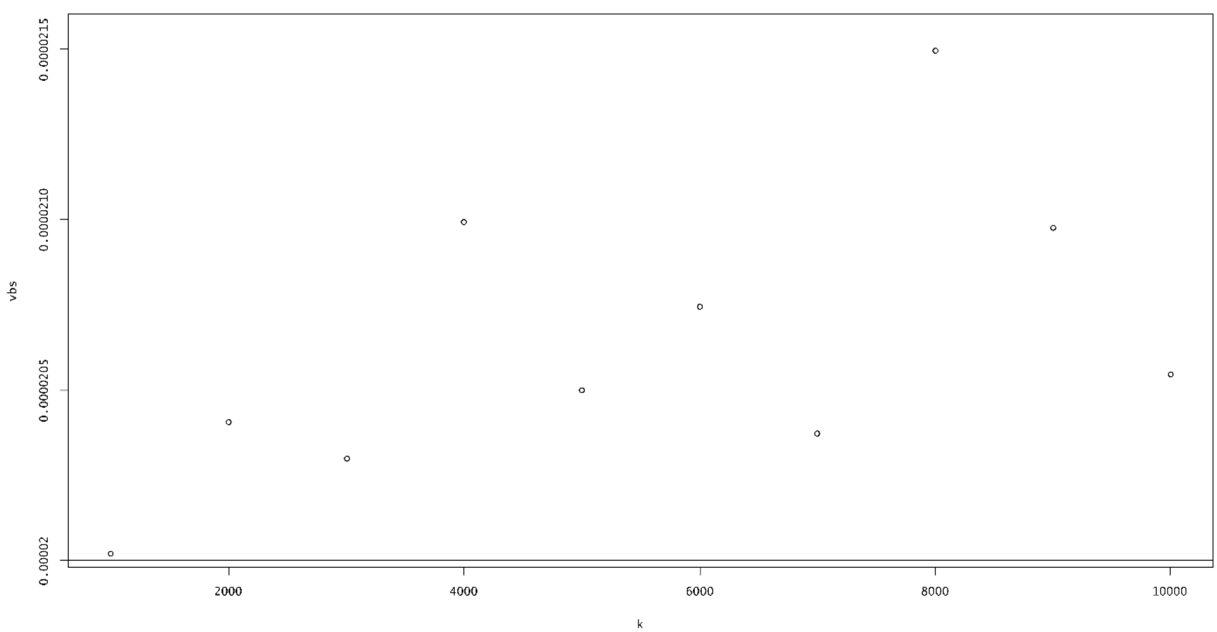
(figure 3.1.13) Variance for different resamples from Cauchy (2.345, 1)



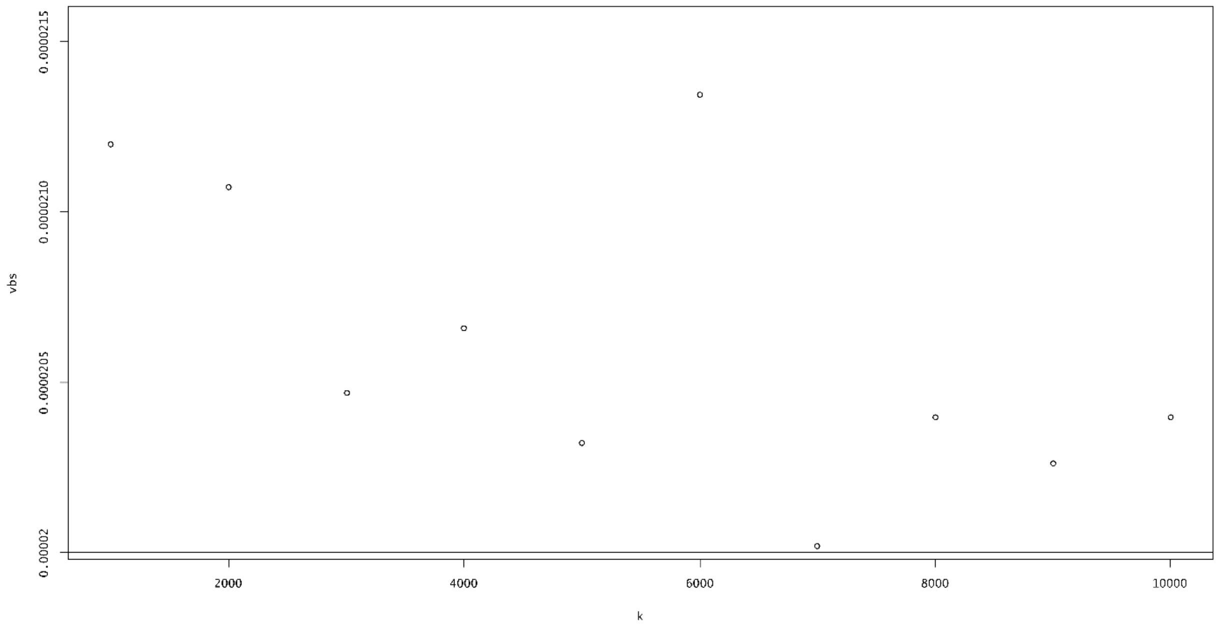
(figure 3.1.14) Variance for different resamples from Cauchy (3.456, 1)



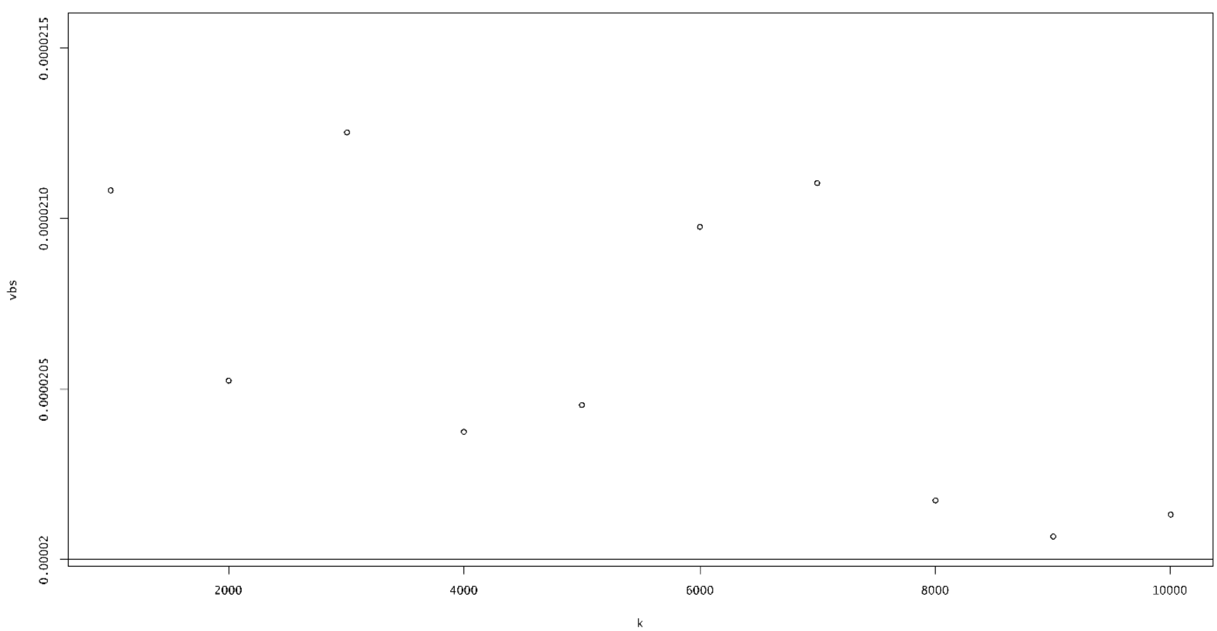
(figure 3.1.15) Variance for different resamples from Cauchy (4.567, 1)



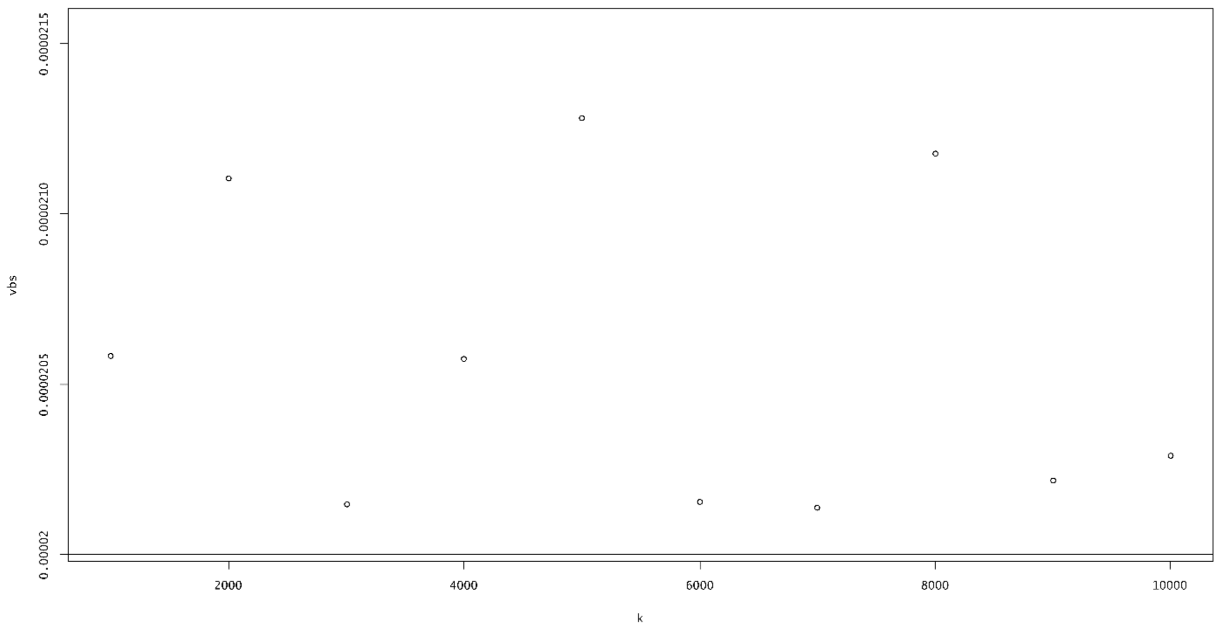
(figure 3.1.16) Variance for different resamples from Cauchy (5.678, 1)



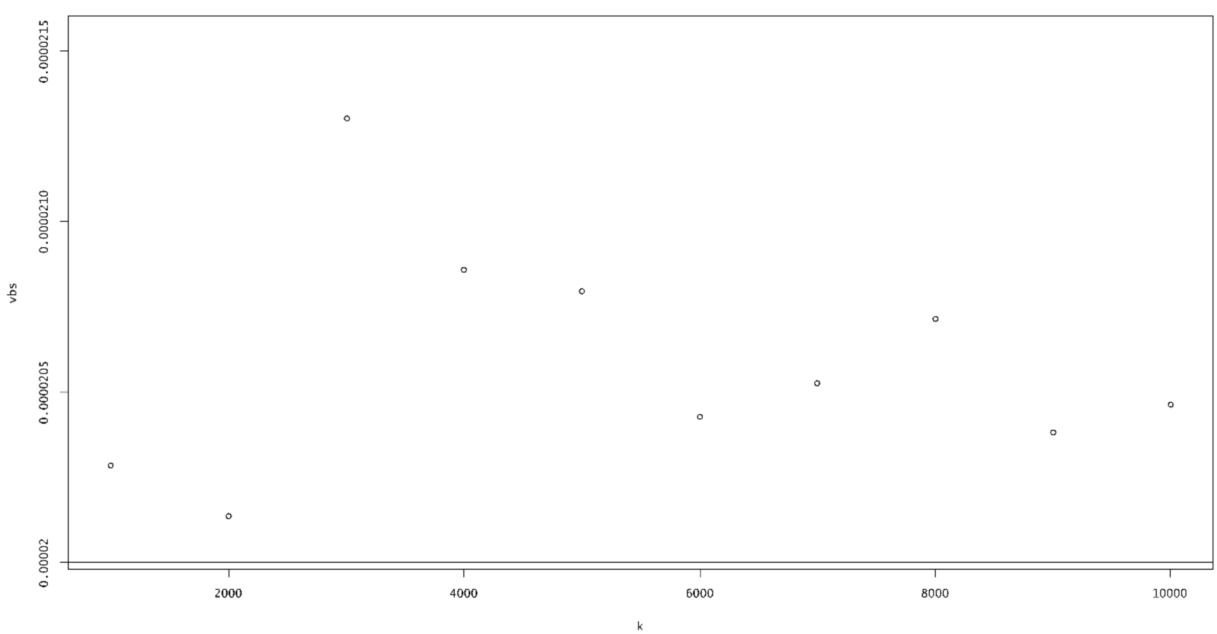
(figure 3.1.17) Variance for different resamples from Cauchy (6.789, 1)



(figure 3.1.18) Variance for different resamples from Cauchy (7.89, 1)



(figure 3.1.19) Variance for different resamples from Cauchy (8.901, 1)



(figure 3.1.20) Variance for different resamples from Cauchy (9.012, 1)

It is clear from the figures that the variance of the estimates of $\hat{\boldsymbol{\theta}}$, obtained using the method of Nonparametric Bootstrapping, tends to the FCRLB.

3.2 GAMMA DISTRIBUTION

An RV X is said to follow Gamma distribution with parameters a and b if its pdf is of the form:

$$f(x) = \frac{be^{-bx}(bx)^{a-1}}{\Gamma(a)} \qquad a \in \mathbb{R}^+, b \in \mathbb{R}^+; x \in \mathbb{R}^+ \tag{3.2.1}$$

We write X ~ Gamma (a, b).

Clearly, for a random sample of size n, the likelihood function of (a, b) ` is given by:

$$L(a,b) = \frac{b^{an} e^{-(b\sum_{i=1}^{n} x_i)} (\prod_{i=1}^{n} x_i)^{a-1}}{\Gamma(a)^n} \tag{3.2.2}$$

The log-likelihood function is given by:

$$l(a,b) = anlogb - nlog\Gamma(a) - b\sum_{i=1}^{n} x_i + (a-1)\sum_{i=1}^{n} logx_i \tag{3.2.3}$$

Therefore, the maximum likelihood equations are:

$$logb - \psi(a) + \frac{\sum_{i=1}^{n} logx_i}{n} = 0 \qquad (3.2.4a)^7$$

[7] $\psi(.)$ is the digamma function, defined to be the first order derivative of log-gamma function.

$$\frac{an}{b} - \sum_{i=1}^{n} x_i = 0 \qquad (3.2.4b)$$

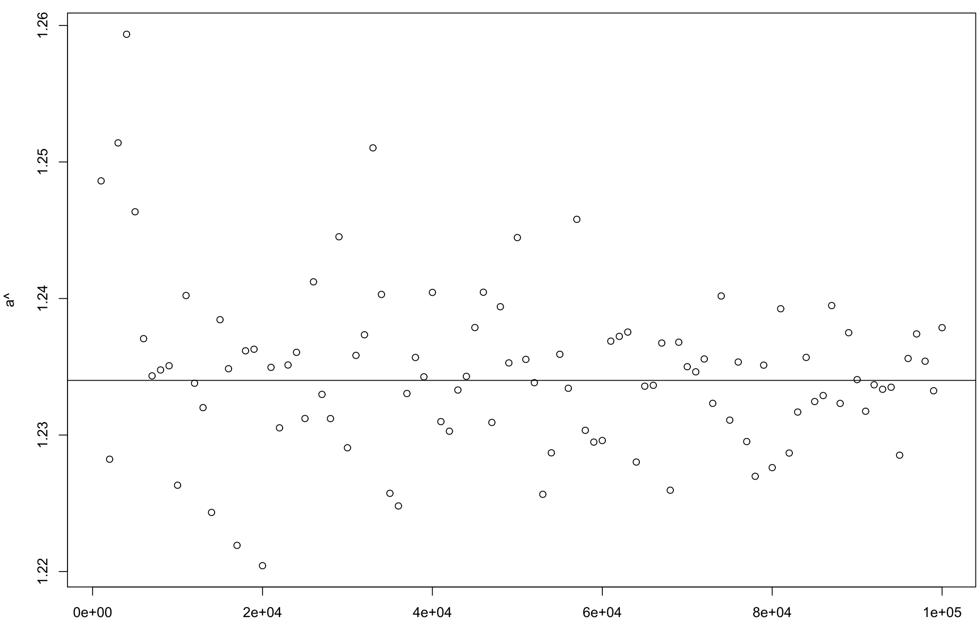
For the sake of simplicity, we take b=1.

Equation (3.2.4a) then reduces to

$$-\psi(a) + \frac{\sum_{i=1}^{n} log x_i}{n} = 0 \qquad (3.2.5)$$

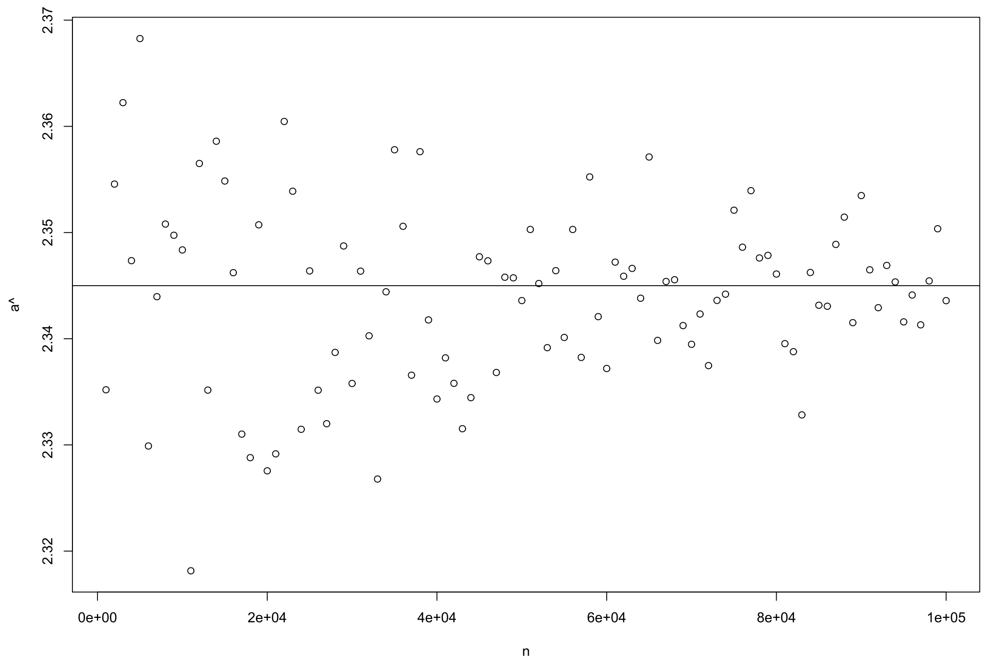
Equation (3.2.5) is not of closed form. We thus apply Newton Raphson method for estimating $\hat{\boldsymbol{a}}$.

For ten different values of a, we estimate $\hat{\boldsymbol{a}}$ using Newton Raphson method[8] for hundred different sample sizes (n=1000(1000)100000). we obtain the following graphs through simulation:

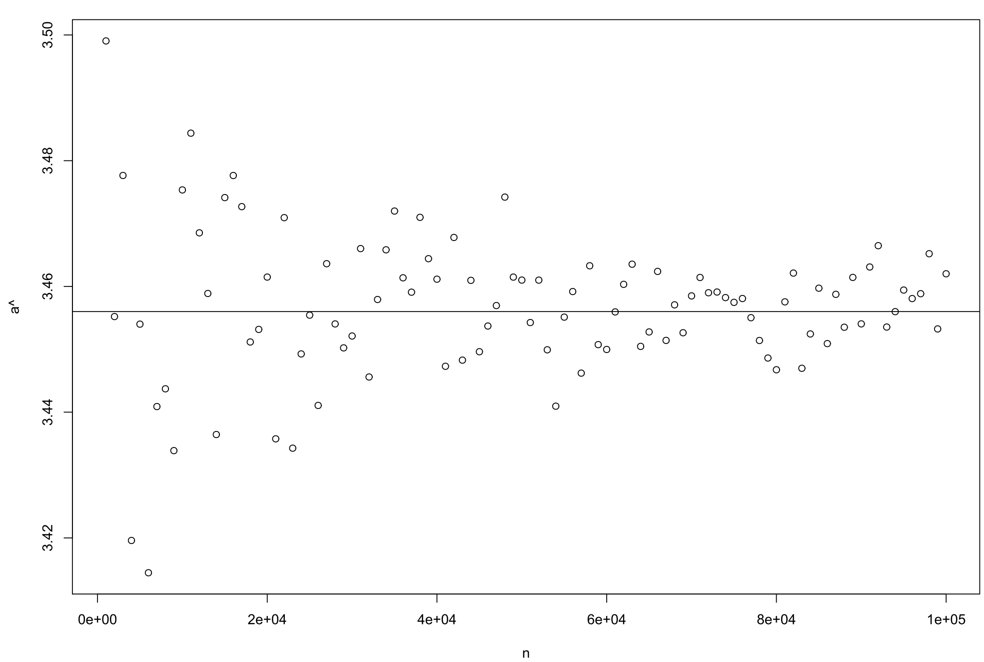


(figure 3.2.1) Samples from Gamma (1.234, 1)

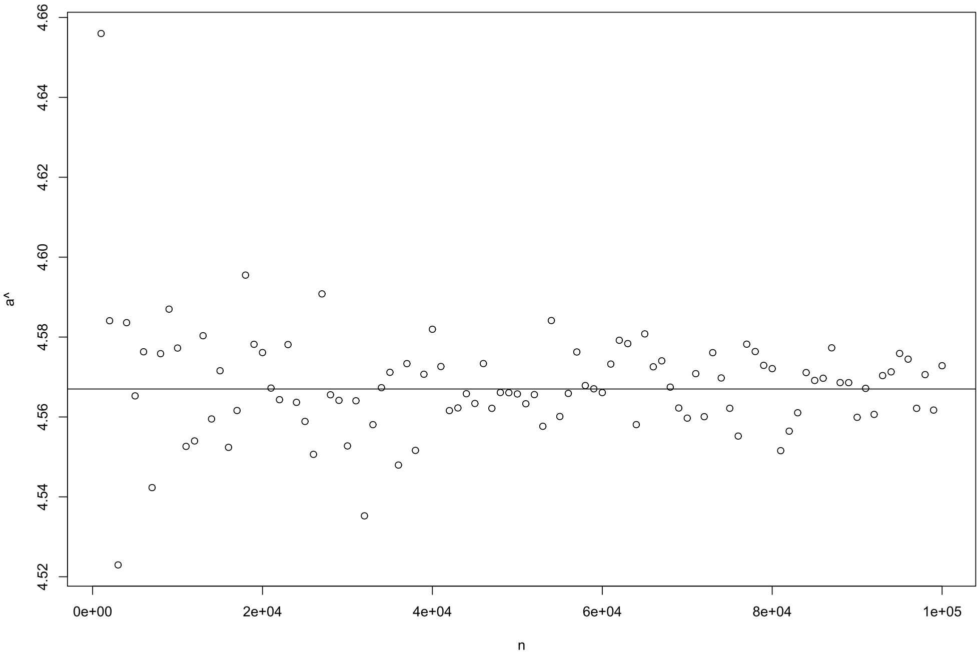
[8] For relevant R code, refer to appendix A3.



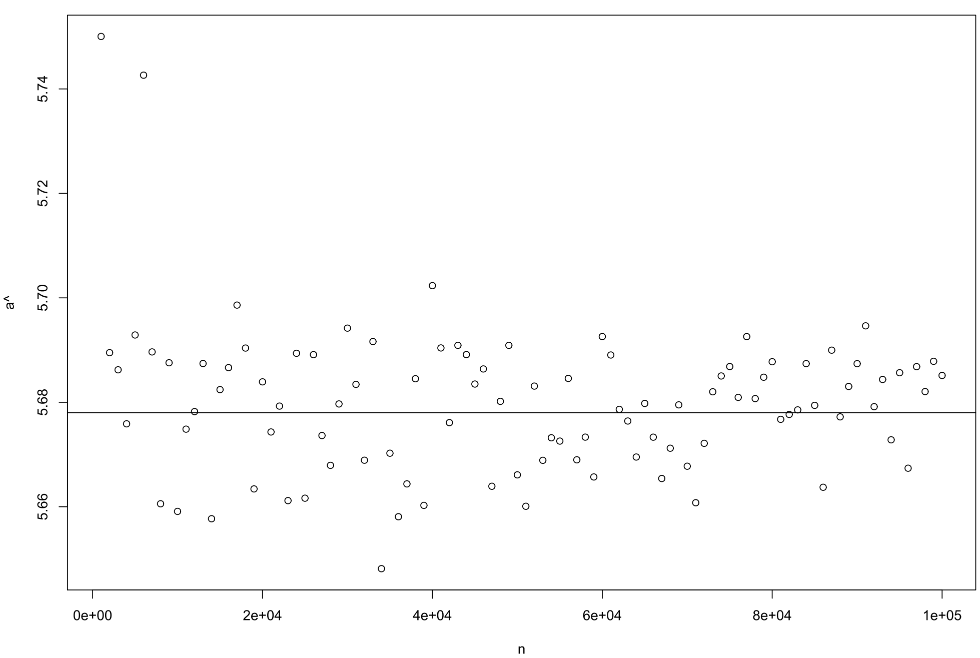
(figure 3.2.2) Samples from Gamma (2.345, 1)



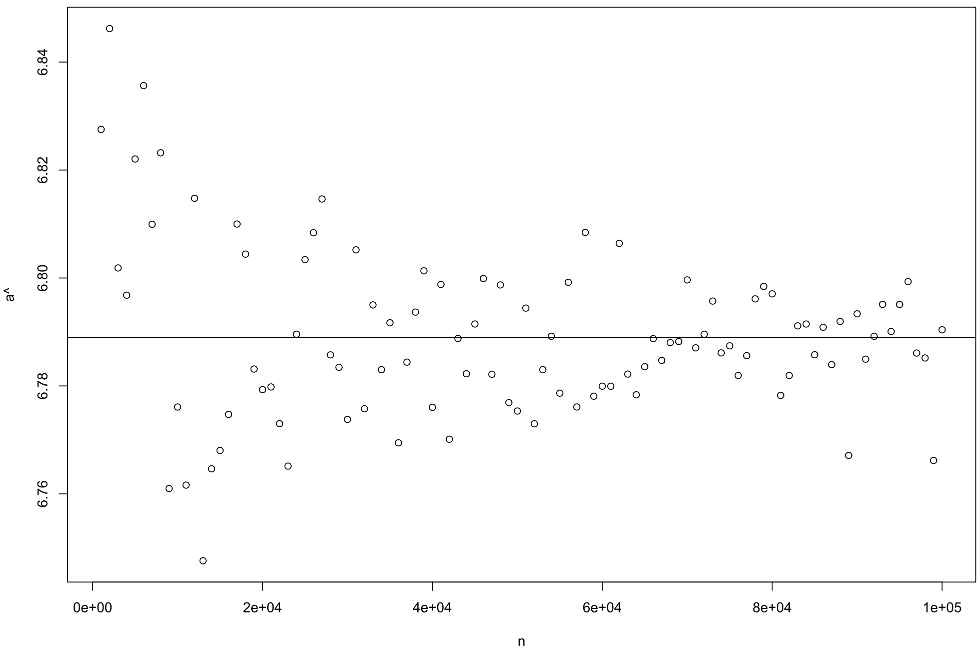
(figure 3.2.3) Samples from Gamma (3.456, 1)



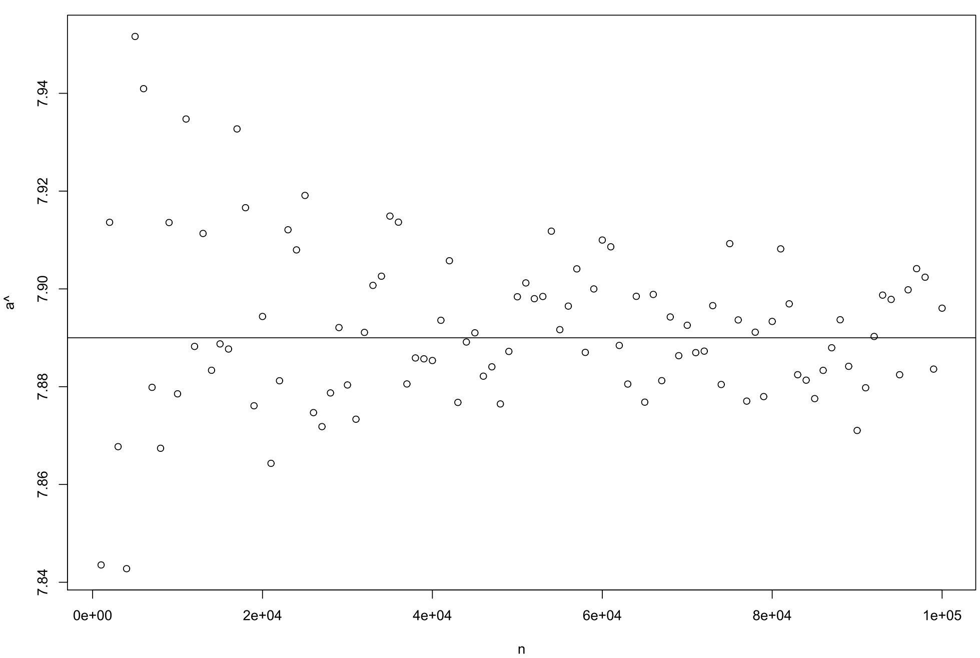
(figure 3.2.4) Samples from Gamma (4.567, 1)



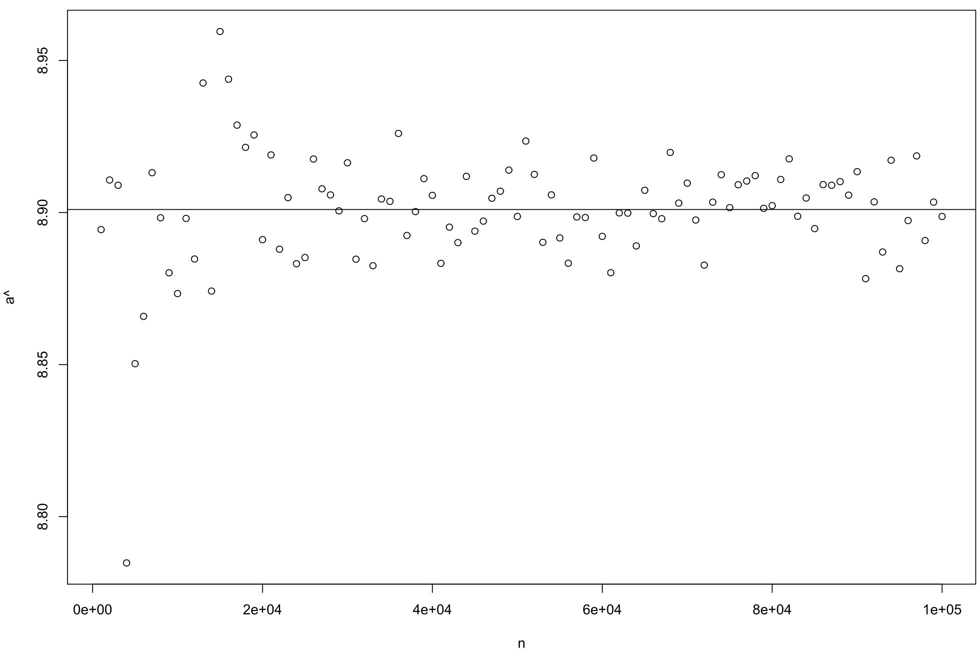
(figure 3.2.5) Samples from Gamma (5.678, 1)



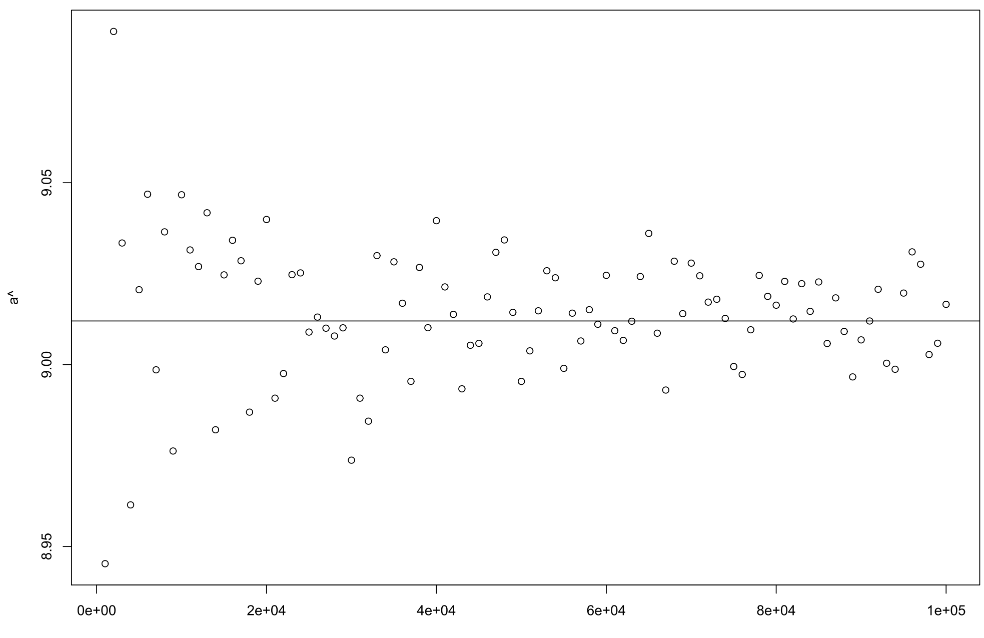
(figure 3.2.6) Samples from Gamma (6.789, 1)



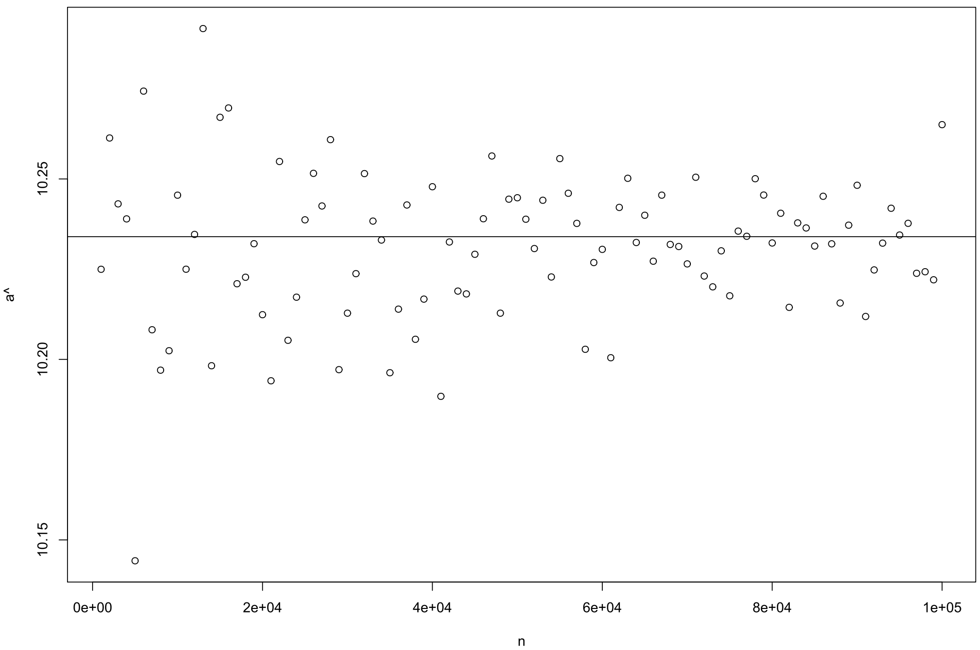
(figure 3.2.7) Samples from Gamma (7.89, 1)



(figure 3.2.8) Samples from Gamma (8.901, 1)



(figure 3.2.9) Samples from Gamma (9.012, 1)



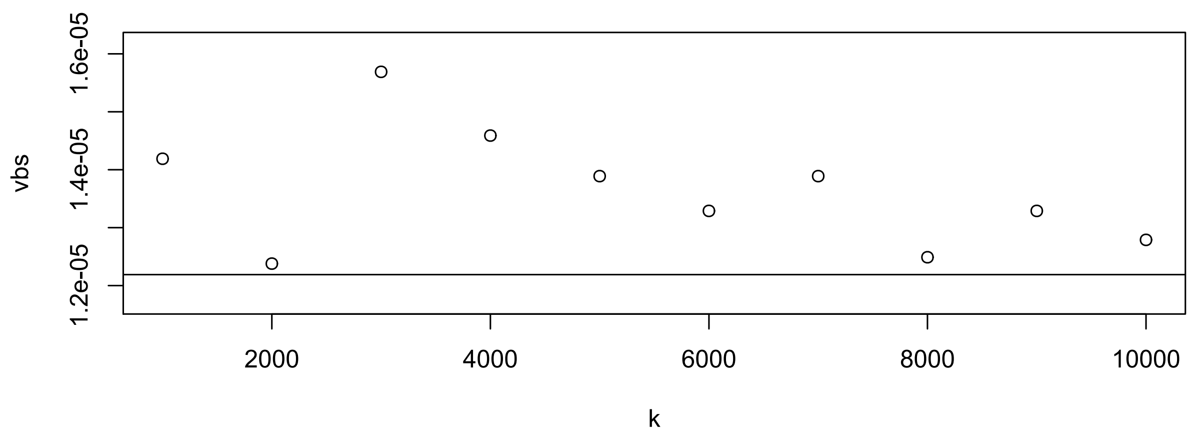
(figure 3.2.10) Samples from Gamma (10.234, 1)

It is clear from the figures that the estimates of $\hat{\boldsymbol{a}}$, obtained using Newton Raphson method, are asymptotically unbiased.

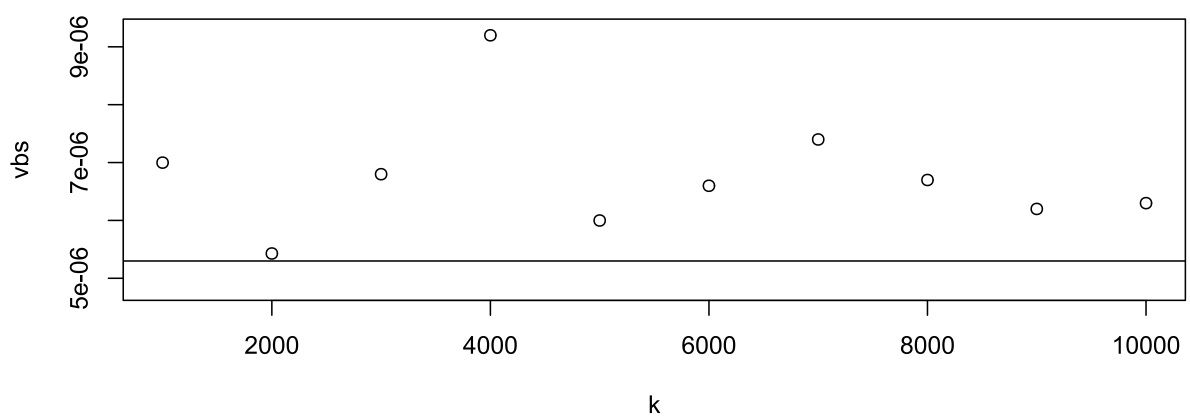
We now perform nonparametric bootstrapping for estimating the variance of the estimator.

For ten different values of a, we estimate $var(\hat{\boldsymbol{a}})$ using Nonparametric Bootstrapping[9] for ten different resample sizes (k=1000(1000)10000). we obtain the following graphs through simulation:

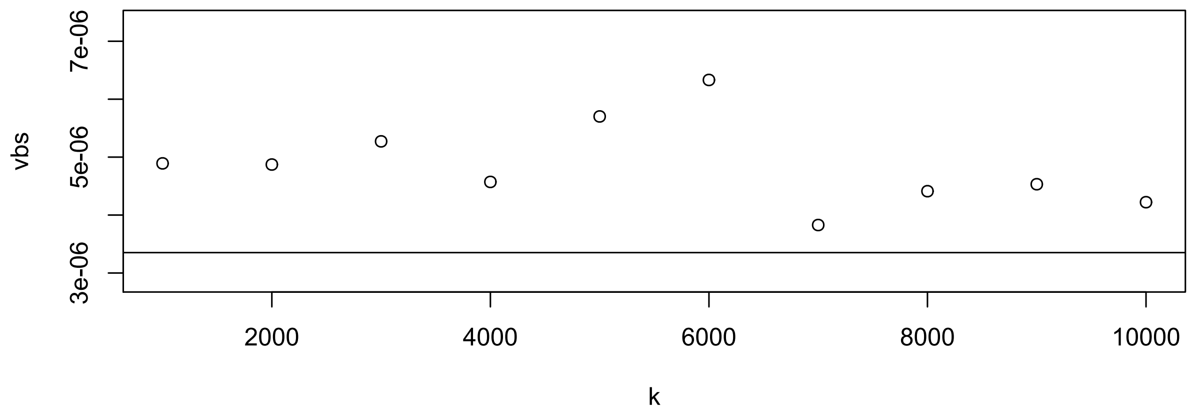
[9] For relevant R code, refer to appendix A4.



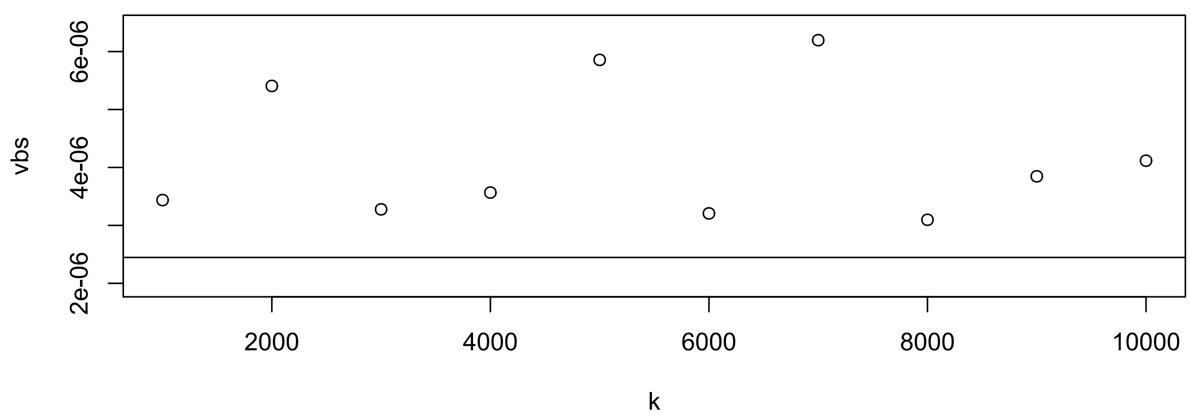
(figure 3.2.11) Variance for different resamples from Gamma (1.234, 1)



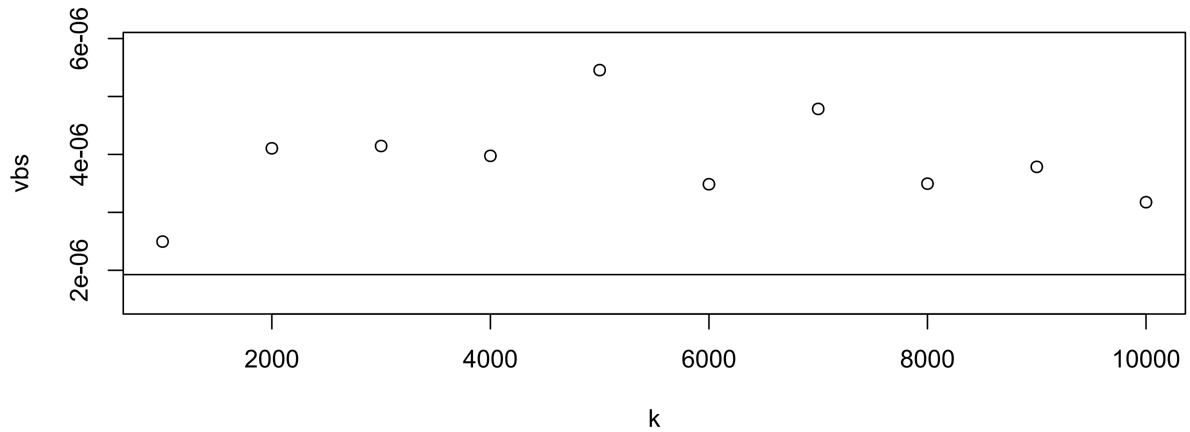
(figure 3.2.12) Variance for different resamples from Gamma (2.345, 1)



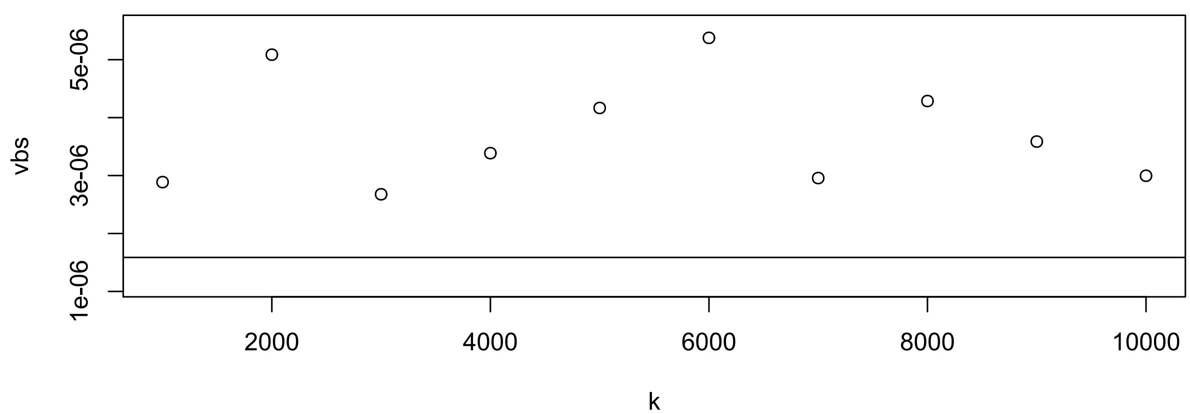
(figure 3.2.13) Variance for different resamples from Gamma (3.456, 1)



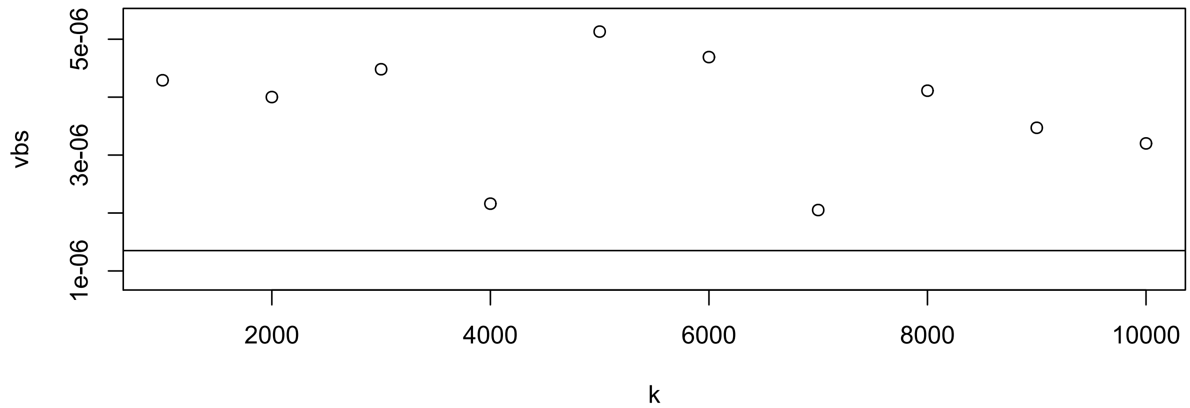
(figure 3.2.14) Variance for different resamples from Gamma (4.567, 1)



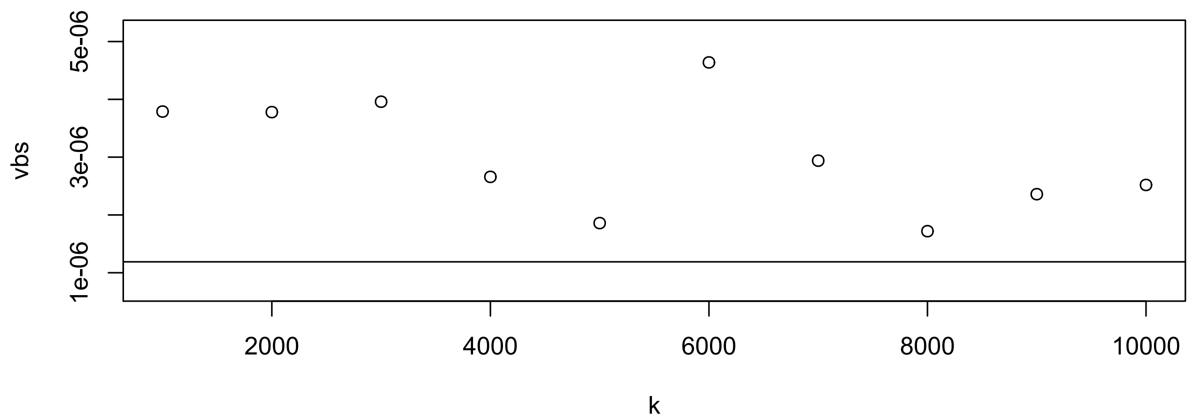
(figure 3.2.15) Variance for different resamples from Gamma (5.678, 1)



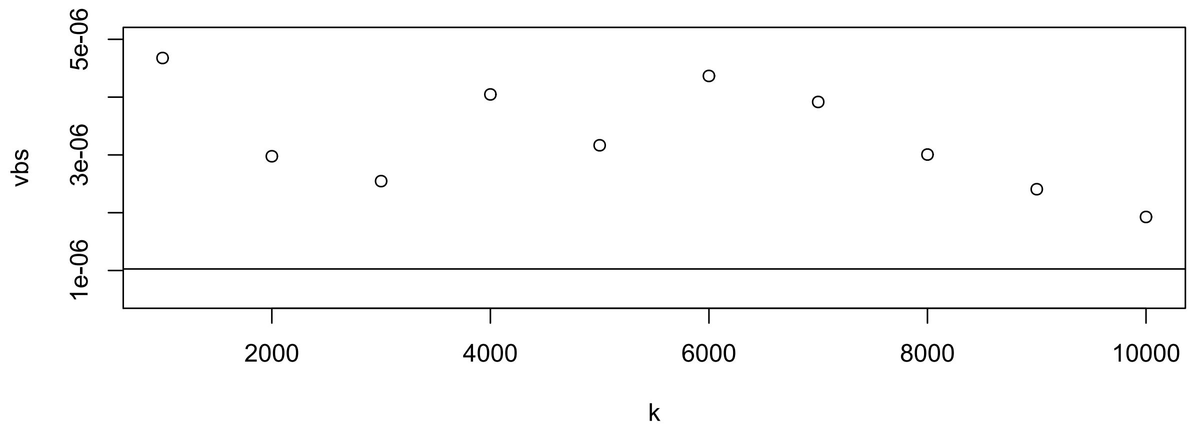
(figure 3.2.16) Variance for different resamples from Gamma (6.789, 1)

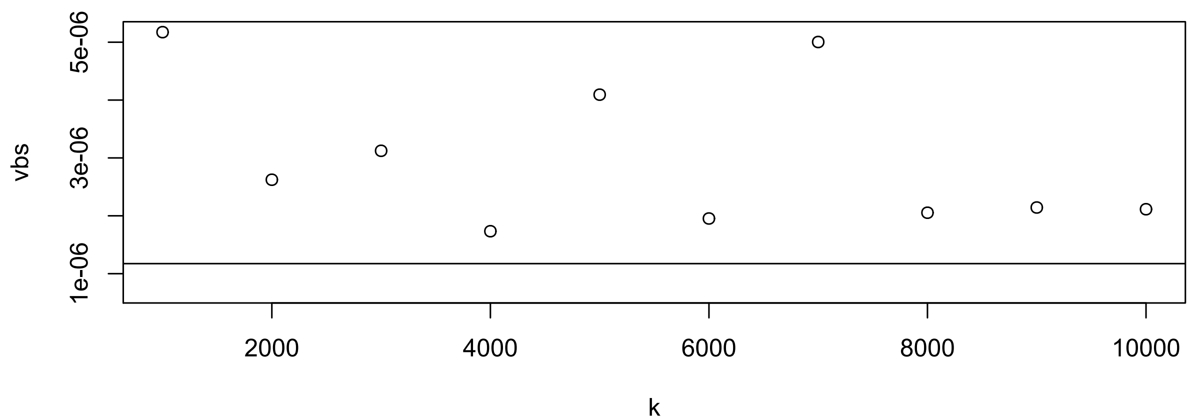


(figure 3.2.17) Variance for different resamples from Gamma (7.89, 1)



(figure 3.2.18) Variance for different resamples from Gamma (8.901, 1)





(figure 3.2.19) Variance for different resamples from Gamma (9.012, 1)

(figure 3.2.20) Variance for different resamples from Gamma (10.234, 1)

It is clear from the figures that the variance of the estimates of $\hat{\boldsymbol{a}}$, obtained using the method of Nonparametric Bootstrapping, tends to the FCRLB.

4. CONCLUSION

In light of the aforementioned simulations, we can conclude that:

1. The m.l.e. of the location parameter of Cauchy distribution obtained using Newton Raphson method is asymptotically unbiased.
2. The variance of that m.l.e., estimated using nonparametric bootstrapping, tends to FCRLB as resampling size increases.
3. The m.l.e. of the shape parameter of Gamma distribution obtained using Newton Raphson method is asymptotically unbiased.
4. The variance of that m.l.e., estimated using nonparametric bootstrapping, tends to FCRLB as resampling size increases.

5. LIMITATIONS

A major limitation of the whole simulation procedure is, obviously, the lack of computational power. For example, the code A2 takes close to three hours to compile for a single instance in a 3.9 GHz OC 6 core processor. Future improvements in core clock speeds would reduce these computation times significantly.

Although the process of **jackknifing** would reduce the computation time, it would be significantly less accurate.

6. FURTHER SCOPE

1. For the sake of simplicity, the took the scale parameter of Cauchy distribution to be unity. Similarly, one can set the location to 0, and find m.l. estimate of the scale parameter in a similar manner. Although significantly harder, one can solve for both parameters simultaneously.
2. For the sake of simplicity, the took the rate parameter of Gamma distribution to be unity. From equation (3.2.4b), it is clear that evaluating m.l. estimate of the rate parameter is quite easy and does not require any numerical method, whatsoever. Although, solving for both parameters simultaneously is significantly harder.
3. For improving computational efficiency, one can implement **bag of little bootstraps**, given proper coding knowledge.
4. The codes can easily be modified to estimate higher order moments.
5. Given enough computational power, one can simply generalise the codes to evaluate m.l. estimates of multidimensional parameters of multidimensional distributions.

7. REFERENCES AND BIBLIOGRAPHY

LIKELIHOOD

1. Rao, C.R. (1973), *Linear Statistical Inference and its Applications*, Wiley, 353-355.
2. Pawitan, Yudi (2001), *In All Likelihood*, Oxford Science Publications.
3. Rohatgi, Vijay K., and Saleh, A.K. Md. Ehsanes (2001), *An Introduction to Probability and Statistics 2nd ed.*, Wiley, 409-421.

NUMERICAL METHODS

4. Atkinson, Kendall E. (1978), *An Introduction to Numerical Analysis*, Wiley.

DISTRIBUTION THEORY

5. Johnson, N.L., Kotz, S., and Balakrishnan, N. (1994), *Continuous Univariate Distributions 1*, Wiley.
6. Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995), *Continuous Univariate Distributions 2*, Wiley.
7. Johnson, N.L., Kotz, S., and Kemp, A.W. (1992), *Univariate Discrete Distributions*, Wiley.

BOOTSTRAPPING

8. Efron, Bradley, and Tibshirani, Robert J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall / CRC.

9. Casella, George, and Berger, Roger L. (2002), *Statistical Inference 2nd ed.*, Duxbury, 478-481.

R, PYTHON etc. FOR STATISTICS

10. Wickham, Hadley, and Grolemund, Garrett (2017), *R for Data Science*, O'Reilly.
11. McKinney, Wes (2012), *Python for Data Analysis*, O'Reilly.
12. Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome (2001), *The Elements of Statistical Learning 2nd ed.*, Springer.
13. James, Gareth, Witten, Daniela, Hastie, Trevor, Tibshirani, Robert (2013), *An Introduction to Statistical Learning: with Applications in R*, Springer.
14. Sweigart, Al (2015), *Automate the Boring Stuff with Python*, No Starch Press.

8. ACKNOWLEDGEMENT

I would like to thank Fr. Principal and the college authorities for their support and logistics. I would also like to thank each and every professor of our department. My project guide Dr. Surabhi Dasgupta continuously helped me by clearing my doubts and suggesting methodologies, procedures, and references. Therefore, without her support and vision, the project would have been incomplete. My friends also helped me in some aspects of the project. Hence, they too are worthy of a mention. Maa kept on waking numerous times in the night just to ensure my wellbeing whenever I studied throughout the night. I will cherish these memories forever. Last but not the least, my entire family kept their unquestionable faith in me, as always.

9. APPENDIX

A1.

R CODE FOR NEWTON RAPHSON METHOD ESTIMATION OF CAUCHY PARAMETER

```
n = seq(1000, 100000, 1000)

th = c(0.123, 1.234, 2.345, 3.456, 4.567, 5.678, 6.789, 7.89, 8.901, 9.012)



g = function(x, p) {

  s = 0

  for (i in 1:length(x)) {

    s = s + {

      (x[i] - p) / (1 + (x[i] - p) ^ 2)

    }

  }

  s

}
```

```
g1 = function(x, p) {
  s = 0
  for (i in 1:length(x)) {
    s = s + {
      ((x[i] - p) ^ 2 - 1) / (1 + (x[i] - p) ^ 2) ^ 2
    }
  }
  s
}

itf = function(th) {
  thhat = array(0)
  for (i in n) {
    x = rcauchy(i, th, 1)
    t = array(0)
    t[1] = floor(th)
    t[2] = t[1] - (g(x, t[1]) / g1(x, t[1]))
```

```
    while (abs(t[2] - t[1]) > 0.0000000000001) {
      t[1] = t[2]
      t[2] = t[2] - (g(x, t[2]) / g1(x, t[2]))
    }
    thhat[i / 1000] = t[2]
  }
  thhat
}

for (i in 1:length(th)) {
  plot(n, itf(th[i]), ylab = "θ^")
  abline(h = th[i])
}
```

A2.

R CODE FOR NONPARAMETRIC BOOTSTRAPPING ESTIMATION OF VARIANCE OF ESTIMATION CAUCHY PARAMETER

```
n = 100000
```

```
for (m in 1:length(th)) {
  y = rcauchy(n, th[m], 1)
  k = seq(1000, 10000, 1000)
  vbs = array(0)
  for (i in 1:length(k)) {
    thhat = array(0)
    for (j in 1:k[i]) {
      x = sample(y, length(y), T)
      t = array(0)
      t[1] = floor(th)
      t[2] = t[1] - (g(x, t[1]) / g1(x, t[1]))
      while (abs(t[2] - t[1]) > 0.0000000001) {
        t[1] = t[2]
        t[2] = t[2] - (g(x, t[2]) / g1(x, t[2]))
      }
      thhat[j] = t[2]
    }
    vbs[i] = var(thhat)
```

```
  }
  plot(k, vbs)
  abline(h = 2 / n)
}
```

A3.

R CODE FOR NEWTON RAPHSON METHOD ESTIMATION OF GAMMA PARAMETER

```
n = seq(1000, 100000, 1000)
th = c(1.234, 2.345, 3.456, 4.567, 5.678, 6.789, 7.89, 8.901, 9.012, 10.234)

g = function(x, p) {
  s = -digamma(p)
  for (i in 1:length(x)) {
    s = s + log(x[i]) / length(x)
  }
  s
}
```

```
itf = function(th) {
  thhat = array(0)
  for (i in n) {
    x = rgamma(i, th, 1)
    t = array(0)
    t[1] = floor(th)
    t[2] = t[1] + (g(x, t[1]) / trigamma(t[1]))
    while (abs(t[2] - t[1]) > 0.0000000000001) {
      t[1] = t[2]
      t[2] = t[2] + (g(x, t[2]) / trigamma(t[2]))
    }
    thhat[i / 1000] = t[2]
  }
  thhat
}

for (i in 1:length(th)) {
```

```
  plot(n, itf(th[i]), ylab = "a^")
  abline(h = th[i])
}
```

A4.

R CODE FOR NONPARAMETRIC BOOTSTRAPPING ESTIMATION OF VARIANCE OF ESTIMATION GAMMA PARAMETER

```
n = 100000
for (m in 1:length(th)) {
  y = rgamma(n, th[m], 1)
  k = seq(1000, 10000, 1000)
  vbs = array(0)
  for (i in 1:length(k)) {
    thhat = array(0)
    for (j in 1:k[i]) {
      x = sample(y, length(y), T)
      t = array(0)
      t[1] = floor(th)
```

```
    t[2] = t[1] + (g(x, t[1]) / trigamma(t[1]))
    while (abs(t[2] - t[1]) > 0.0000000001) {
      t[1] = t[2]
      t[2] = t[2] + (g(x, t[2]) / trigamma(t[2]))
    }
    thhat[j] = t[2]
   }
   vbs[i] = var(thhat)
  }
  plot(k, vbs)
  abline(h = 1 / n * trigamma(th[m]))
}
```

A5.

FRÉCHET CRAMÉR RAO LOWER BOUND

The following theorem provides a lower bound for the variance of an estimator, as well as establishes the concept of Fréchet Cramér Rao lower bound:

Th 1.(Fréchet, Cramér, Rao) For an open interval $\Theta \subseteq \mathbb{R}$, suppose the family $\{f_\theta : \theta \subseteq \Theta\}$ satisfies the following regularity conditions:

i. The support $\mathcal{X} \coloneqq \{x : f_\theta(x) > 0\}$ does not depend on θ.

ii. For $x \in \mathcal{X}$ and $\theta \in \Theta$, $\frac{\partial}{\partial\theta}\log f_\theta(x)$ exists, and is finite.

iii. For any statistic h(.)with finite $E_\theta|h(X)|$ for all θ, integration (summation) and differentiation with respect to θ can be interchanged in $E_\theta(h(X))$. I.e.,

$$\frac{\partial}{\partial\theta}\int h(x)f_\theta(x)dx = \int h(x)\frac{\partial}{\partial\theta}f_\theta(x)dx \qquad (A5.1)$$

whenever the R.H.S. of equation (A5.1) is finite.

Let T(X) be such that $var_\theta(T(X))$ is finite for all θ. Let $\psi(\theta)=E_\theta(X)$.

If

$$I(\theta) \coloneqq E_\theta\left\{\frac{\partial}{\partial\theta}\log f_\theta(x)\right\}^2 \qquad (A5.2)$$

satisfies $I(\theta) \in \mathbb{R}$, then,

$$var_\theta\big(T(X)\big) \geq \frac{\{\psi'(\theta)\}^2}{I(\theta)} \qquad (A5.3)$$

The R.H.S. of equation (A5.3) is called the FCRLB of variance.

Result 1. For X ~ Cauchy (θ, σ), FCRLB of θ is 2/n.

Result 2. For X ~ Gamma (a, b), FCRLB of a is 1/{n trigamma(a)}.