

STAT 303 Mathematical Statistics I Term Project: Point Estimation of Gamma Distribution Parameters

Kemal Can Yoloğlu - Student ID: 2614915

January 21, 2026

1 Introduction

This project examines point estimation methods for the parameters of the Gamma distribution. We are going to compare the Method of Moments (MoM) and Maximum Likelihood Estimation (MLE) approaches through both theoretical derivation and extensive simulation studies.

The statistical model assumes that observations X_1, \dots, X_n are independent and identically distributed:

$$X_i \sim \text{Gamma}(k, \theta), \quad k > 0, \theta > 0,$$

with probability density function:

$$f(x|k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad x > 0,$$

where k is the shape parameter and θ is the scale parameter.

2 Data Description

For this analysis, we generated simulated data from a Gamma distribution with known parameters:

- True shape parameter: $k = 3$
- True scale parameter: $\theta = 2$
- Sample size: $n = 50$

2.1 Descriptive Statistics

The histogram of the simulated data is shown in Figure 1. Basic sample statistics are:

- Sample mean: $\bar{X} = 4.9545$
- Sample variance: $S^2 = 8.4787$

The theoretical skewness for $\text{Gamma}(k, \theta)$ is $2/\sqrt{k}$. For our true parameters ($k = 3$), this gives a skewness of approximately 1.155, indicating moderate right-skewness. The histogram confirms this positive skewness, suggesting that the Gamma model is appropriate for these data.

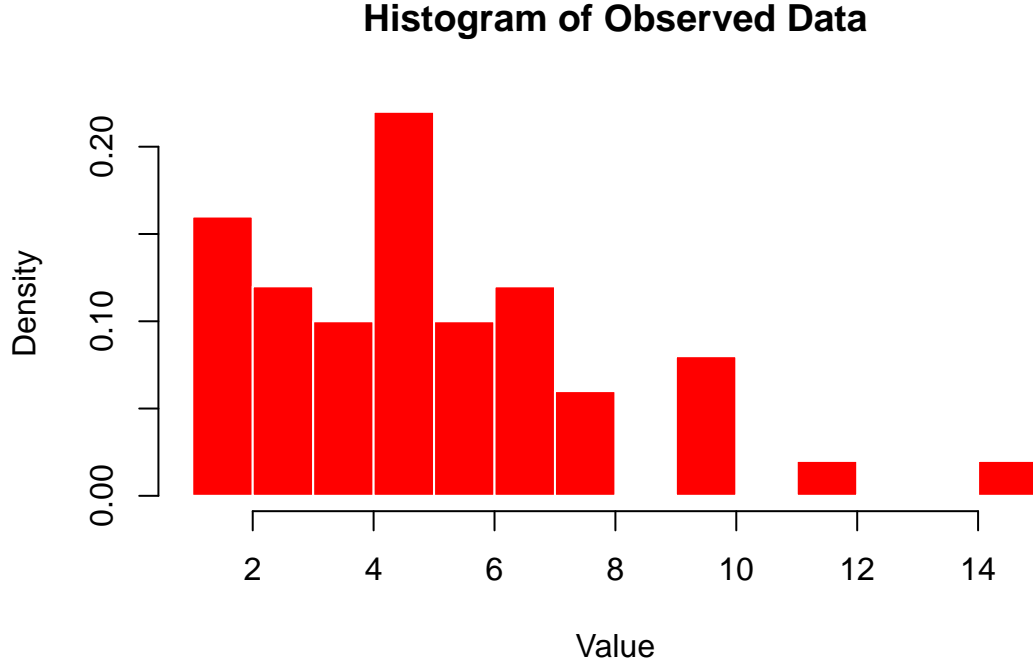


Figure 1: Histogram of simulated Gamma data

3 Point Estimation Methods

3.1 Method of Moments (MoM)

The Method of Moments equates population moments with sample moments. For the Gamma distribution:

$$E(X) = k\theta \quad (1)$$

$$\text{Var}(X) = k\theta^2 \quad (2)$$

Setting these equal to the sample moments \bar{X} and S^2 :

$$\bar{X} = \hat{k}_{\text{MM}}\hat{\theta}_{\text{MM}} \quad (3)$$

$$S^2 = \hat{k}_{\text{MM}}\hat{\theta}_{\text{MM}}^2 \quad (4)$$

From the first equation: $\hat{\theta}_{\text{MM}} = \bar{X}/\hat{k}_{\text{MM}}$. Substituting into the second equation:

$$S^2 = \hat{k}_{\text{MM}} \left(\frac{\bar{X}}{\hat{k}_{\text{MM}}} \right)^2 = \frac{\bar{X}^2}{\hat{k}_{\text{MM}}}$$

Solving for \hat{k}_{MM} :

$$\boxed{\hat{k}_{\text{MM}} = \frac{\bar{X}^2}{S^2}}$$

And therefore:

$$\hat{\theta}_{\text{MM}} = \frac{S^2}{\bar{X}}$$

Estimates for our data:

- $\hat{k}_{\text{MM}} = 2.8951$
- $\hat{\theta}_{\text{MM}} = 1.7113$

3.2 Maximum Likelihood Estimation (MLE)

The log-likelihood function for the Gamma distribution is:

$$\ell(k, \theta) = (k-1) \sum_{i=1}^n \ln x_i - \frac{1}{\theta} \sum_{i=1}^n x_i - nk \ln \theta - n \ln \Gamma(k)$$

Taking partial derivatives:

$$\frac{\partial \ell}{\partial \theta} = \frac{1}{\theta^2} \sum_{i=1}^n x_i - \frac{nk}{\theta} = 0$$

This gives:

$$\hat{\theta}_{\text{MLE}} = \frac{\bar{X}}{\hat{k}_{\text{MLE}}}$$

For k , we have:

$$\frac{\partial \ell}{\partial k} = \sum_{i=1}^n \ln x_i - n \ln \theta - n\psi(k) = 0$$

where $\psi(\cdot)$ is the digamma function. Substituting $\hat{\theta}_{\text{MLE}}$:

$$\ln(\hat{k}) - \psi(\hat{k}) = \ln(\bar{X}) - \frac{1}{n} \sum_{i=1}^n \ln x_i$$

Why no closed-form solution exists: The equation involves the digamma function $\psi(k) = \frac{d}{dk} \ln \Gamma(k)$, which is a transcendental function. This creates a nonlinear equation in k that cannot be solved algebraically and therefore requires numerical methods. In other words, Calculus Methods fails. In this study, the root of the corresponding estimating equation is obtained using numerical root-finding techniques, specifically the `uniroot` function implemented in R.

Estimates for our data:

- $\hat{k}_{\text{MLE}} = 2.8679$
- $\hat{\theta}_{\text{MLE}} = 1.7275$

3.3 Comparison on Observed Data

Table 1 compares the true parameters with both estimation methods:

Table 1: Comparison of parameter estimates

Method	k	θ
True	3.0000	2.0000
MoM	2.8951	1.7113
MLE	2.8679	1.7275

Figure 2 shows the fitted Gamma densities overlaid on the histogram of the data. Both MoM and MLE provide reasonable fits to the data. Compared to the true density, both methods slightly underestimate the parameters, which is consistent with the lower estimated values of k and θ . The fitted curves obtained by MoM and MLE are very close to each other, indicating similar performance for this sample.

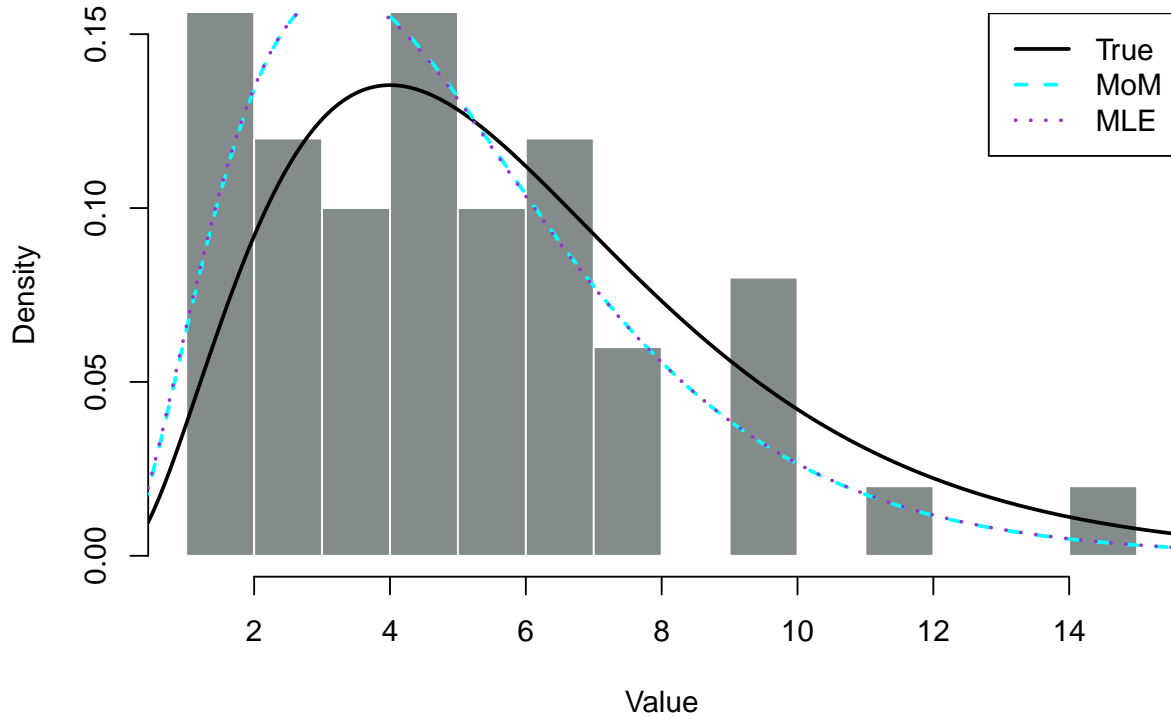


Figure 2: Fitted Gamma densities compared to data

4 Simulation Study

4.1 Design

We conducted a simulation study with:

- **Scenario 1:** $k = 1, \theta = 2$ (high skewness: exponential distribution)
- **Scenario 2:** $k = 5, \theta = 1$ (moderate skewness)

- Sample sizes: $n \in \{20, 50, 100\}$
- Number of replications: $R = 2000$

For each replication, we generated data from the specified Gamma distribution, computed MoM and MLE estimates, and calculated bias, variance, and MSE for each estimator.

4.2 Performance Measures

For an estimator $\hat{\theta}$ of parameter θ :

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta \quad (5)$$

$$\text{Variance}(\hat{\theta}) = \text{Var}(\hat{\theta}) \quad (6)$$

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2 \quad (7)$$

4.3 Results

4.3.1 Scenario 1: High Skewness ($k = 1, \theta = 2$)

Table 2 summarizes the performance of both estimators across different sample sizes for Scenario 1.

Table 2: Simulation results for Scenario 1 ($k = 1, \theta = 2$)

n	Parameter	Method	Bias	Variance	MSE
20	k	MoM	0.2256	0.2486	0.2995
20	k	MLE	0.1516	0.1474	0.1704
20	θ	MoM	-0.0899	0.9473	0.9554
20	θ	MLE	-0.0920	0.5291	0.5375
50	k	MoM	0.0868	0.0770	0.0846
50	k	MLE	0.0534	0.0371	0.0400
50	θ	MoM	-0.0240	0.3700	0.3706
50	θ	MLE	-0.0310	0.1984	0.1993
100	k	MoM	0.0459	0.0365	0.0386
100	k	MLE	0.0239	0.0159	0.0165
100	θ	MoM	-0.0134	0.1894	0.1896
100	θ	MLE	-0.0102	0.0981	0.0982

4.3.2 Scenario 2: Moderate Skewness ($k = 5, \theta = 1$)

Table 3 presents the results for the second scenario.

Table 3: Simulation results for Scenario 2 ($k = 5, \theta = 1$)

n	Parameter	Method	Bias	Variance	MSE
20	k	MoM	0.6625	4.3584	4.7973
20	k	MLE	0.8218	4.2200	4.8953
20	θ	MoM	-0.0029	0.1313	0.1314
20	θ	MLE	-0.0447	0.1002	0.1022
50	k	MoM	0.2552	1.2825	1.3476
50	k	MLE	0.2891	1.1398	1.2233

n	Parameter	Method	Bias	Variance	MSE
50	θ	MoM	-0.0081	0.0482	0.0483
50	θ	MLE	-0.0205	0.0402	0.0406
100	k	MoM	0.1367	0.6244	0.6431
100	k	MLE	0.1466	0.5039	0.5254
100	θ	MoM	-0.0038	0.0254	0.0254
100	θ	MLE	-0.0105	0.0200	0.0201

4.3.3 Graphical Analysis

Figure 3 displays MSE as a function of sample size for both scenarios and both parameters.

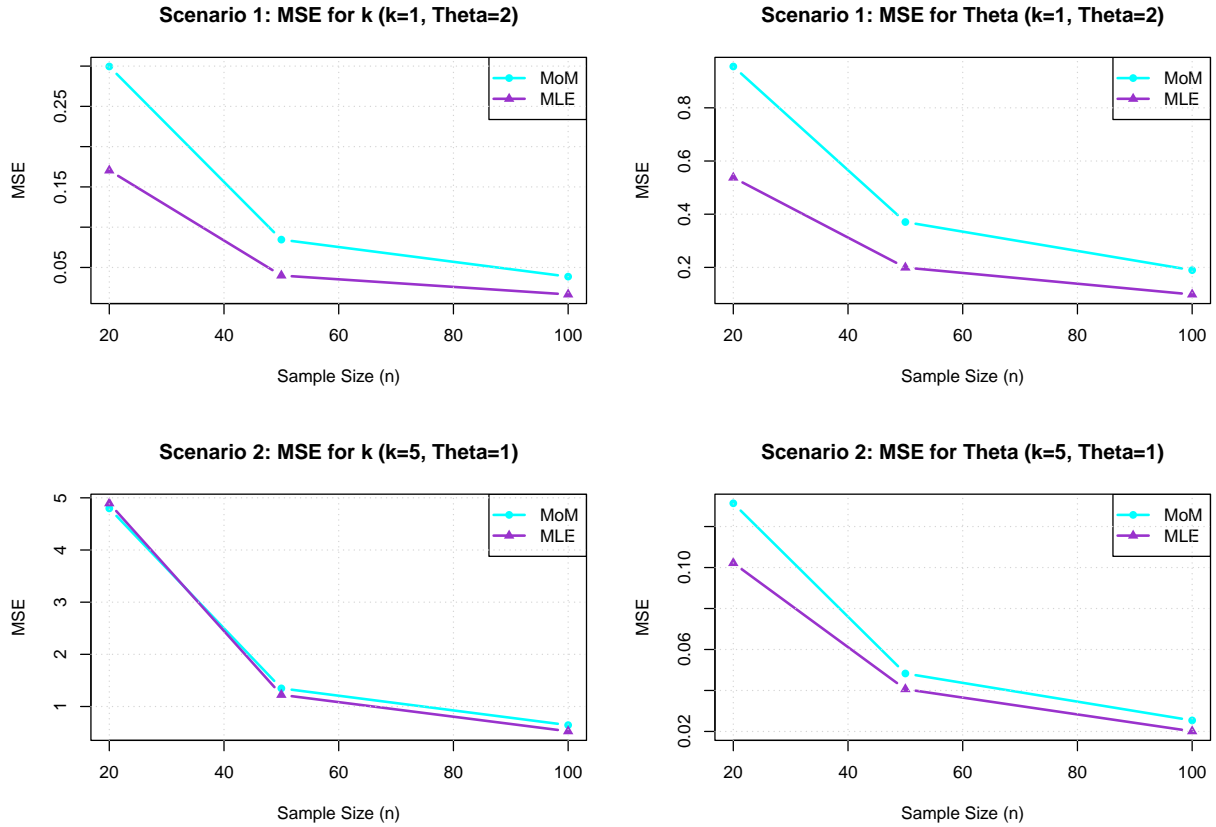


Figure 3: MSE comparison across sample sizes

5 Discussion

5.1 Comparative Performance

The simulation study gives crucial informations about comparing MoM and MLE methods. By looking at the MSE results in Figure 3, for all cases and sample sizes MLE generally get better results than MoM.

For Scale parameter Theta both methods are performs well however MLE is slightly efficient. This results match with our theoretical expectations because MLE uses full likelihood functions and that's way it becomes more efficient than method of moment estimators.

In addition to general accuracy, skewness is effective on bias and efficiency. In high skewness scenario (Scenario 1), both estimators shows larger bias, especially for shape parameter k . MoM exhibits larger bias because it trusts sample moments and they are very sensitive for extreme values.

From an efficiency perspective, larger skewness causes larger variance for both parameters. In contrast, MLE gives continually less variance than MoM, meaning that it gives better efficiency. This efficiency difference is evident especially for k parameter, for theta parameter difference is smaller.

In Scenario 2, when skewness decreases, then both bias and variance also decrease crucially. In that case, efficiency difference between MoM and MLE is getting narrower and MoM becomes more competitive alternative, especially for larger sample sizes.

5.2 Effect of Skewness

Scenario 1 ($k = 1$) represents a highly right-skewed distribution, while Scenario 2 ($k = 5$) shows to moderate skewness. The results indicate that increased skewness largely complicates parameter estimation, causes higher bias and MSE for both estimators.

The advantage of MLE over MoM is more evident under high skewness. In Scenario 1, MoM is more sensitive to extreme observations and it negatively affects performance. However, when skewness decreases in Scenario 2, the performance gap between the two methods are getting narrower, and MoM becomes more competitive, especially for larger sample sizes.

5.3 Effect of Sample Size

As expected, increasing the sample size improves estimation accuracy for both methods. In all scenarios, MSE decreases as the sample size increases from $n = 20$ to $n = 100$, as showed in Figure 3.

The decline in MSE is most occurs when moving from small to moderate sample sizes, indicating that both estimators benefit greatly from extra data. For large samples ($n = 100$), both MoM and MLE perform well, although MLE consistently have lower MSE, confirming its superior finite-sample performance.

5.4 Practical Recommendations

Based on the simulation results, several practical recommendations can be made. MLE should be choosen because it has better statistical features and lower MSE, especially for very skewed distributions and smaller sample sizes.

MoM remains a useful alternative because of its simplicity and closed-form expressions. It can be used for quick exploratory analysis or the source of initial values for numerical MLE optimizations. For small samples with highly skewed data, regardless of the chosen method; larger samples are recommended to ensure more reliable estimation.

6 Conclusion

In this project, we studied and compared two basic point estimation techniques MoM and MLE. The analysis combined theoretical derivations, estimation on observed data, and extensive simulation work under varying skewness levels and sample sizes, providing a comprehensive evaluation of both methods.

The results clearly shows that MLE generally outperforms than MoM for bias, variance and mean square error. This advantage is evident especially for estimating shape parameter k where MLE demonstrates greater stability and efficiency, especially in cases of highly skewed distributions and small sample sizes. This results are matches with classical statistical theory, as the theory states and we discussed in lectures that MLE achieves asymptotic properties by fully using the likelihood function.

Despite MoM performs worse results, it remains a valuable and practical estimation technique. Its simplicity, closed-form solutions, and easier computational make it attractive for exploratory data analysis and for providing reliable starting values in numerical MLE procedures. Under moderate skewness and sufficiently large sample sizes, MoM can produce estimates similar to those obtained with MLE.

This simulation study also highlights the critical role of sample size in parameter estimation. As the sample size increases, both estimators show improved accuracy and reduced mean square error, confirming their consistency. Although the performance difference between MoM and MLE narrows for large sample sizes, MLE consistently maintains a slight advantage in all cases considered.

Generally, the findings of this project highlight the importance of selecting an estimation method based on data characteristics such as skewness and sample size, as well as practical computational considerations. This drives up the pragmatic worth of statistics-based approaches in real world data analysis.