Point Estimation of Gamma Distribution Parameters

EFE ANIK 2614436

12 January 2026

1 Introduction

This report is about the point estimation of Gamma distribution parameters using the Method of Moments and Maximum Likelihood Estimation. Firstly, we analyze an observed dataset and then conduct a simulation study to compare estimator performance across different sample sizes and two levels of skewness. This report will consider point estimation, and will evaluate performance using bias, variance, and mean squared error. The Gamma model is an appropriate choice for strictly positive and right skewed data. It is flexible enough to represent different skewness levels through the shape parameter k. I compare MoM and MLE because they represent two common estimation properties. MoM is simple and closed form. MLE is likelihood based and generally more efficient. MLE is computationally heavier. The goal is not only to compute the estimates. Also to understand how the choice of estimator changes the fitted shape and how performance acts when the sample size is limited.

1.1 Statistical model

Throughout the report, I assume $X_1, \ldots, X_n$ are i.i.d. and follow a Gamma distribution with shape $k > 0$ and scale $\theta > 0$:

$$X_i \sim \text{Gamma}(k, \theta), \qquad i = 1, \ldots, n.$$

The density is

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

Parameter k mainly controls the skewness and θ scales the distribution.

2 Observed Data Analysis

2.1 Data description

The observed dataset was simulated from a Gamma distribution with true parameters $k = 2$ and $\theta = 18$, using a sample size of $n = 100$ and a fixed random seed for reproducibility. All observations are positive. That makes the Gamma model appropriate. I used a simulated observed dataset to keep the study fully reproducible. Also to ensure the data satisfy the model assumptions. Choosing moderate values $(k, \theta) = (2, 18)$ produces a visibly skewed sample with a meaningful right tail. This helps to illustrate how MoM and MLE can lead to different fitted shapes on the same dataset.

2.2 Descriptive statistics

For the observed sample ($n = 100$), the sample mean and sample variance are

$$\bar{x} = 33.0468, \qquad s^2 = 425.8687.$$

The histogram in Figure 1 shows a right skewed distribution. Its consistent with the Gamma family. Also, the large sample variance compared to the mean supports the existence of a right tail. Since for a Gamma model $\mathrm{Var}(X) = k\theta^2$, matching the variance is crucial for capturing tail thickness. Which is why MoM and MLE can differ mainly in skewness and tail behavior rather than the center.

2.3 Parameter estimation: MoM and MLE

Method of Moments (MoM): For $X \sim \mathrm{Gamma}(k, \theta)$ with shape $k > 0$ and scale $\theta > 0$, we have

$$\mathbb{E}[X] = k\theta, \qquad \mathrm{Var}(X) = k\theta^2.$$

Equating these with the sample moments $\bar{x}$ and s^2 gives the MoM estimators

$$\hat{k}_{\mathrm{MoM}} = \frac{\bar{x}^2}{s^2}, \qquad \hat{\theta}_{\mathrm{MoM}} = \frac{s^2}{\bar{x}}.$$

Using $\bar{x} = 33.0468$ and $s^2 = 425.8687$, we obtain

$$\hat{k}_{\mathrm{MoM}} = 2.5644, \qquad \hat{\theta}_{\mathrm{MoM}} = 12.8868.$$

Maximum Likelihood Estimation (MLE): The log-likelihood for iid. data $x_1, , , , , x_n$ from $\mathrm{Gamma}(k, \theta)$ is:

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

Maximizing with respect to θ yields:

$$\hat{\theta}_{\mathrm{MLE}} = \frac{\bar{x}}{\hat{k}_{\mathrm{MLE}}}.$$

Substituting back, $\hat{k}_{\mathrm{MLE}}$ must satisfy the nonlinear equation:

$$\ln(\hat{k}_{\mathrm{MLE}}) - \psi(\hat{k}_{\mathrm{MLE}}) = \ln(\bar{x}) - \frac{1}{n} \sum_{i=1}^{n} \ln x_i,$$

where $\psi(\cdot)$ is the digamma function. Since this equation has no closed form solution for k, $\hat{k}_{\mathrm{MLE}}$ is obtained numerically. For the observed sample,

$$\hat{k}_{\mathrm{MLE}} = 2.1552, \qquad \hat{\theta}_{\mathrm{MLE}} = 15.3332.$$

In practice, we solve this equation with a numerical root finding method. To make the iteration stable, I use the MoM estimate as an initial value. Because it is fast to compute and usually close to the optimum when n is not extremely small. This step explains why MLE is more computationally demanding. In each dataset we must numerically solve for $k > 0$ before computing $\hat{\theta}_{\mathrm{MLE}} = \bar{x}/\hat{k}_{\mathrm{MLE}}$.
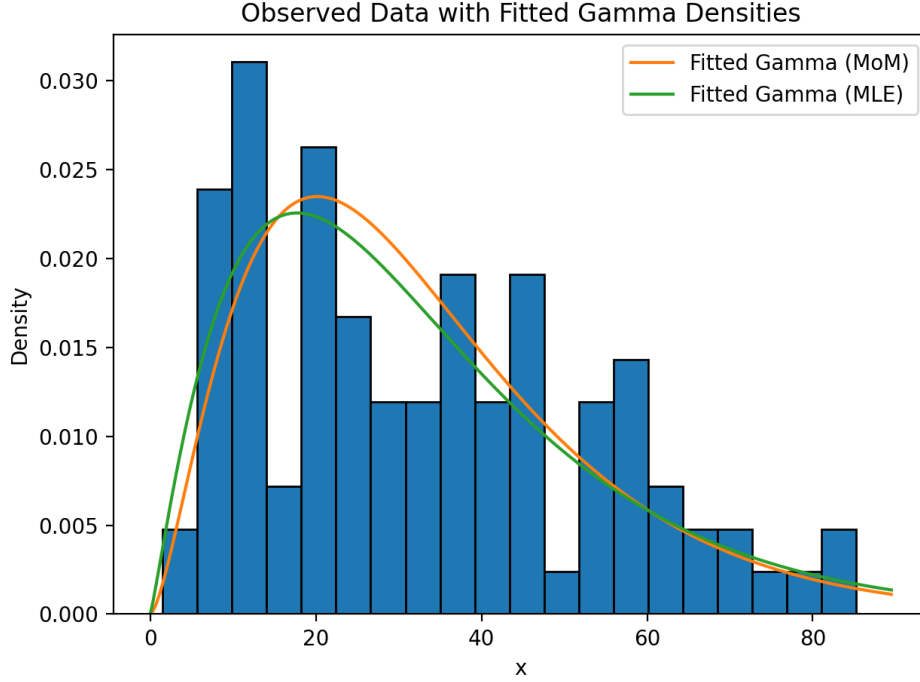
2.4 Fitted densities

Figure 1: Observed sample ($n = 100$): histogram with fitted Gamma densities using MoM and MLE parameter estimates.

2.5 Interpretation of MoM and MLE on the observed sample

Both MoM and MLE match the fitted mean to the sample mean. For MoM, $\hat{k}_{\mathrm{MoM}}\hat{\theta}_{\mathrm{MoM}} = \bar{x}$, and for MLE the identity $\hat{\theta}_{\mathrm{MLE}} = \bar{x}/\hat{k}_{\mathrm{MLE}}$ implies $\hat{k}_{\mathrm{MLE}}\hat{\theta}_{\mathrm{MLE}} = \bar{x}$. So, the main difference between two fitted models is how they represent the shape.

In our results, MoM yields a larger shape estimate ($\hat{k}_{\mathrm{MoM}} = 2.5644$) than MLE ($\hat{k}_{\mathrm{MLE}} = 2.1552$). Since larger k typically means less right skewness, the MoM fit is slightly less skewed. In Figure 1, the two curves are close around the bulk of the data. However, can differ in the right tail. A few large observations affect the overall fit.

MLE uses the full likelihood and is more efficient in finite samples. MoM depends only on the first two sample moments and can be more sensitive to variability in s^2. The cost of MLE is computational. The estimating equation for k has no closed form solution, so numerical methods are required.

3 Simulation Study

3.1 Design

Two Gamma distributions are considered to represent different skewness levels:

$$\text{Case 1: } k, \theta = 1, 2 \quad \text{higher skewness,} \qquad \text{Case 2: } k, \theta = 5, 1 \quad \text{moderate skewness.}$$

For each case and each $n \in \{20, 50, 100\}$, we generate $R = 2000$ independent samples from $\mathrm{Gamma}(k, \theta)$. For every sample, we compute both MoM and MLE estimates of k and θ.

Estimator performance:

$$\mathrm{Bias}(\hat{\eta}) = \mathbb{E}[\hat{\eta}] - \eta, \qquad \mathrm{Var}(\hat{\eta}) = \mathbb{E}\left[(\hat{\eta} - \mathbb{E}[\hat{\eta}])^2\right], \qquad \mathrm{MSE}(\hat{\eta}) = \mathbb{E}\left[(\hat{\eta} - \eta)^2\right],$$

for $\eta \in \{k, \theta\}$. Monte Carlo averages over the R replications are used to approximate these quantities. These two cases were chosen to represent a difficult estimation setting and a more regular setting. The sample sizes $n \in \{20, 50, 100\}$ were selected to show small, medium, and larger sample behavior with a realistic range. I used $R = 2000$ replications to reduce Monte Carlo noise so that differences between MoM and MLE are not driven by random simulation error.

3.2 Tables of Bias, Variance, and MSE

Table 1: Simulation results for the shape parameter k:

Case	n	Bias (MLE)	Bias (MoM)	MSE (MLE)	MSE (MoM)	Var (MLE)	Var (MoM)
Case 1 (k=1, theta=2)	20	0.151168	0.221615	0.159774	0.272119	0.136991	0.223117
Case 1 (k=1, theta=2)	50	0.051680	0.082259	0.039160	0.082327	0.036507	0.075598
Case 1 (k=1, theta=2)	100	0.026162	0.051974	0.017510	0.039187	0.016834	0.036504
Case 2 (k=5, theta=1)	20	0.867349	0.712111	5.194299	5.018968	4.444227	4.514122
Case 2 (k=5, theta=1)	50	0.294752	0.248125	1.325233	1.493948	1.238974	1.433098
Case 2 (k=5, theta=1)	100	0.144671	0.127119	0.529783	0.637023	0.509108	0.621174

Table 2: Simulation results for the scale parameter θ:

Case	n	Bias (MLE)	Bias (MoM)	MSE (MLE)	MSE (MoM)	Var (MLE)	Var (MoM)
Case 1 (k=1, theta=2)	20	-0.105666	-0.103884	0.486110	0.845448	0.475182	0.835073
Case 1 (k=1, theta=2)	50	-0.038867	-0.028213	0.202494	0.363796	0.201084	0.363182
Case 1 (k=1, theta=2)	100	-0.024401	-0.038022	0.099576	0.176562	0.099030	0.175204
Case 2 (k=5, theta=1)	20	-0.054614	-0.014705	0.104650	0.132643	0.101718	0.132493
Case 2 (k=5, theta=1)	50	-0.016526	0.001289	0.042588	0.055586	0.042336	0.055612
Case 2 (k=5, theta=1)	100	-0.010866	-0.002934	0.019793	0.024842	0.019684	0.024845

Tables 1 and 2 separate the error into bias and variance components. This helps interpret why MSE changes. In many entries, the variance term is the main contributor. So increasing n reduces MSE by shrinking variance. Another clear pattern is the scale parameter θ is much harder in the Case 1. In Case 2 the MSE values for θ are already small and decrease quickly with n.
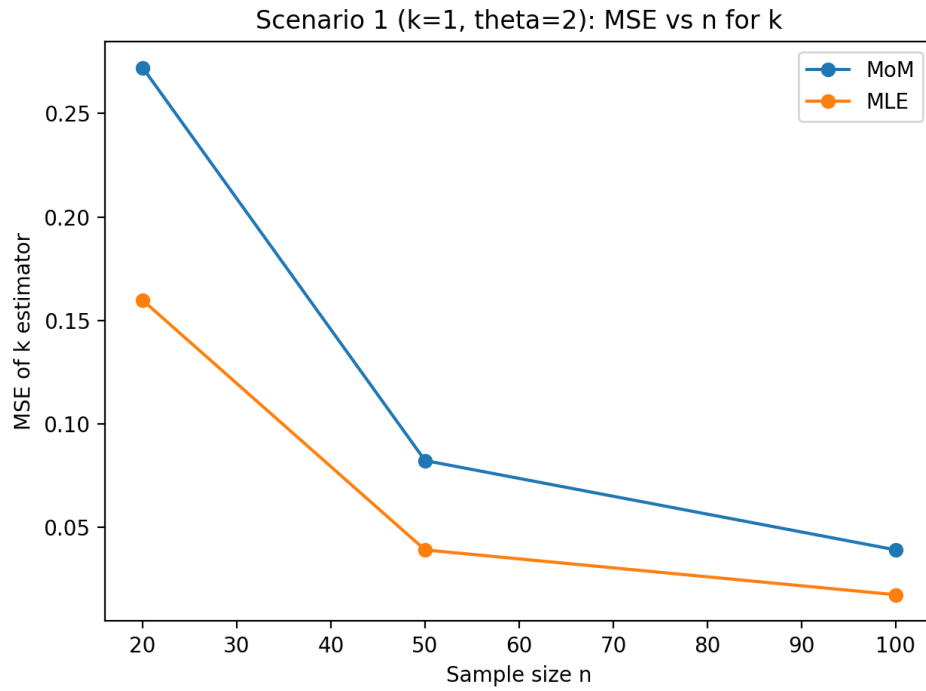
3.3 MSE plots

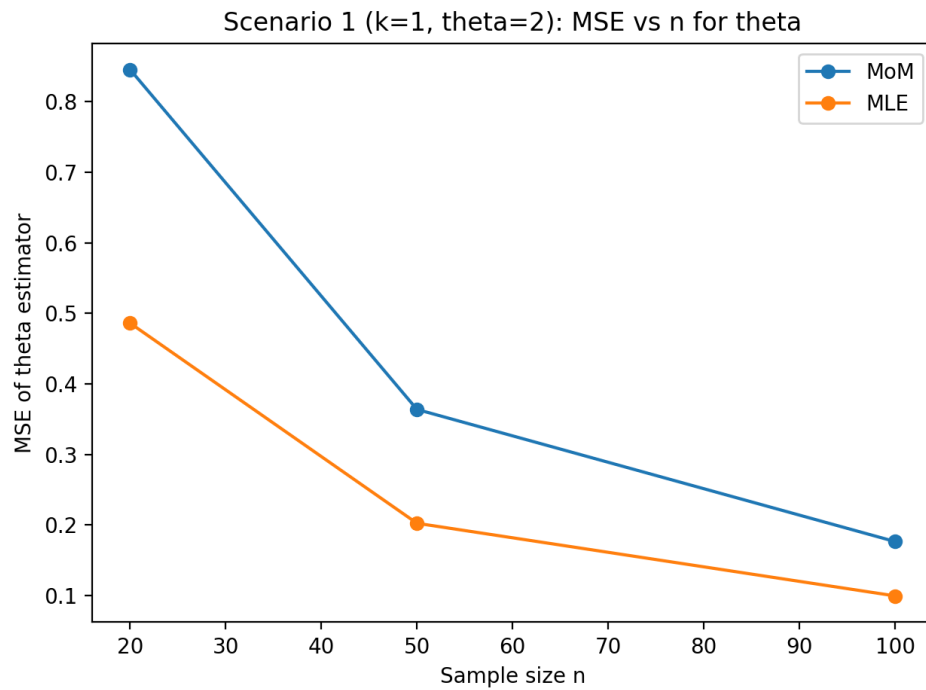Figure 2: Case 1: MSE and sample size for the shape parameter k.

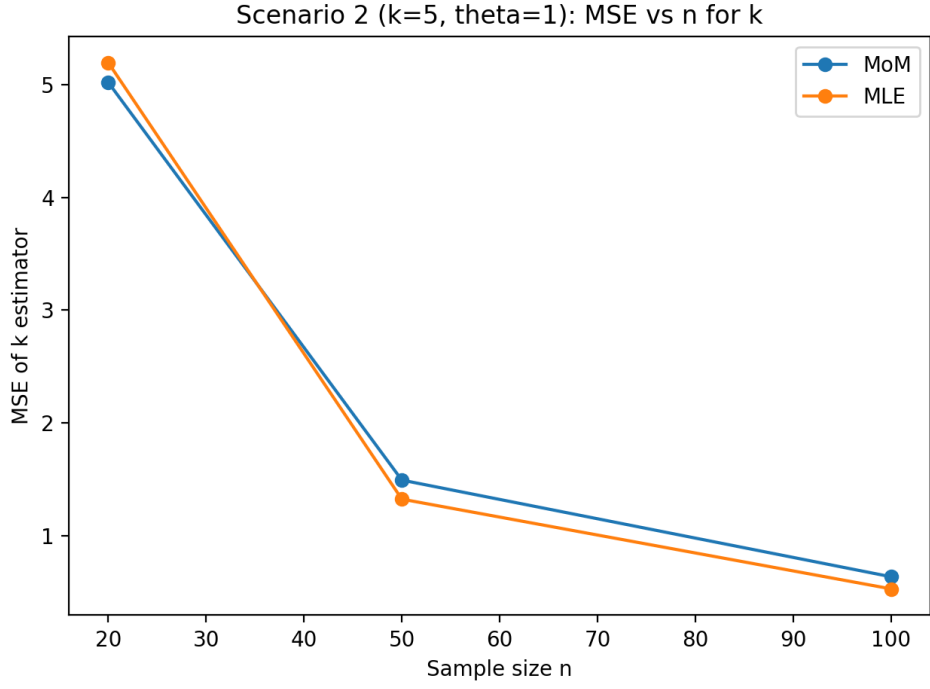Figure 3: Case 1: MSE and sample size for the scale parameter θ.

Figure 4: Case 2: MSE and sample size for the shape parameter k.
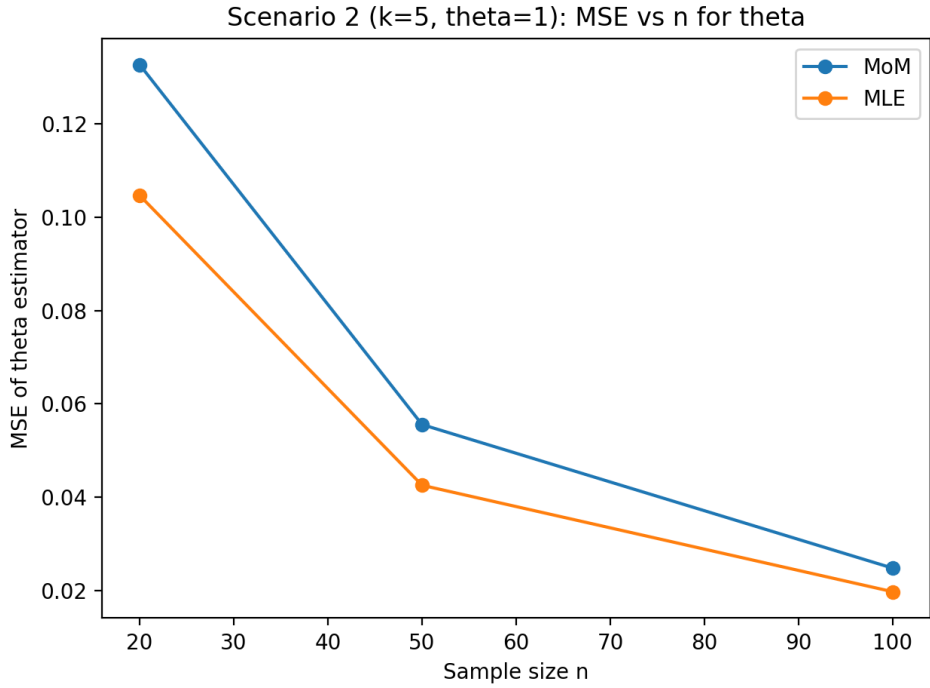

Figure 5: Case 2: MSE and sample size for the scale parameter θ.

The plots visually confirm the same conclusion as the tables. MSE decreases with n for both parameters and both estimators. The MoM and MLE gap is most visible in Case 1. Skewness is high and finite-sample instability is stronger.

3.4 Discussion

Effect of sample size: In both cases and for both parameters, MSE decreases as n increases. This is expected because larger samples reduce sampling variability, and both estimators become more stable. The tables also shows that for each parameter, the variance component becomes smaller as n grows, which lowers the MSE.

3.5 Effect of skewness:

Case 1 is more right skewed, so estimation is harder at small n. A small sample from a highly skewed distribution can contain a few unusually large values. It can inflate the sample variance and also change the likelihood shape. This increases both variance and MSE, specially for θ. In Case 2, the distribution is less skewed and the estimators stabilize faster.

3.6 MoM vs MLE:

Also, its nice to remember that $\text{MSE}(\hat{\eta}) = \text{Var}(\hat{\eta}) + \{\text{Bias}(\hat{\eta})\}^2$. So when n increases variance dropping fast and MSE goes down. In our tables, the primary source of improvement as n increases is the decreasing variance term. So, in Case 1 with $n = 20$, the MSE for estimating k is 0.1598 MLE vs 0.2721 MoM, and the MSE for estimating θ is 0.4861 MLE vs 0.8454 MoM. This gap shrinks by $n = 100$, where the MSE become 0.0175 MLE vs 0.0392 MoM for k, and 0.0996 MLE vs 0.1766 MoM for θ. So, MLE tends to produce smaller MSE than MoM, particularly in the more skewed case and at smaller sample sizes. As n increases, the difference between the methods becomes smaller (learned as asymptotic theory). However, MoM can still be appealing because of its simple closed form solution. In the other hand, MLE requires numerical methods to determine k. In Case 2, the two methods get closer fastly. Because the sample moments are more stable and the likelihood surface is less extreme. So the MoM and MLE gap can look smaller there. A small difference is normal because MoM has no optimization error. While MLE depends on a numerical solve for k in each replication. Finally, I would prefer MLE when the sample is small or the data is highly skewed.

3.7 Practical recommendations

Based on the simulation results, if the dataset is highly right skewed with small k and the sample size is small, MLE is the safer choice because it tends to have lower MSE and better stability. If the sample size is large, the performance gap between MoM and MLE becomes smaller, so MoM can be acceptable when a fast closed form estimate is needed. In applications where computation is not a limitation, I would generally prefer MLE since it uses more information from the likelihood and is typically more efficient.

4 Conclusion

This project examined and contrasted method of moments and maximum likelihood estimators for Gamma distribution parameters. On the observed dataset, the two methods produced different (k, θ) pairs while fitting the same sample mean. So the main differences come from how each method represents skewness and tail behavior. In the simulation study, MSE decreased with sample size. Also, MLE generally outperformed MoM. The largest advantage appearing in the more skewed case and for smaller samples. Overall, the simulations suggest that if the data are highly skewed or the sample is small, MLE is a safer choice in terms of MSE. MoM remains useful as a quick baseline estimator and as a practical starting point for numerical MLE routines.

5   Codes

All codes are submitted as a zip file in the "Efe303ProjectCodes" upload.