STAT 303 – Mathematical Statistics I

Term Project: Point Estimation of Gamma Distribution Parameters

Rubar Akyıldız Melih Akay

January 21, 2026

1 Objective of the Project

The aim of this project is to:

- Apply point estimation methods covered in class,
- Derive and compute Method of Moments (MoM) and Maximum Likelihood (MLE) estimators for the Gamma distribution,
- Compare estimators using finite-sample performance measures (Bias, Variance, MSE),
- Gain experience with numerical likelihood optimization and simulation-based inference.

This project focuses exclusively on point estimation; confidence intervals are not considered.

2 Statistical Model

We assume that the data $X_1, \dots, X_n$ are independent and identically distributed with a **Gamma distribution**:

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

with probability density function

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

Here:

- k: shape parameter
- θ: scale parameter

Remark: The exponential distribution is a special case when $k = 1$.

3 Data

For this project, we use a **simulated dataset** generated from a Gamma distribution. The true parameter values are:

- Shape parameter: $k = 3$
- Scale parameter: $\theta = 2$
- Sample size: $n = 100$

The data were generated using R with `set.seed(303)` for reproducibility.

4 Required Tasks

4.1 Descriptive Statistics

In this section, we analyze the simulated dataset to assess the suitability of the Gamma model.

4.1.1 Data Visualization

Figure 1 shows the histogram of the simulated data with an overlaid kernel density estimate.

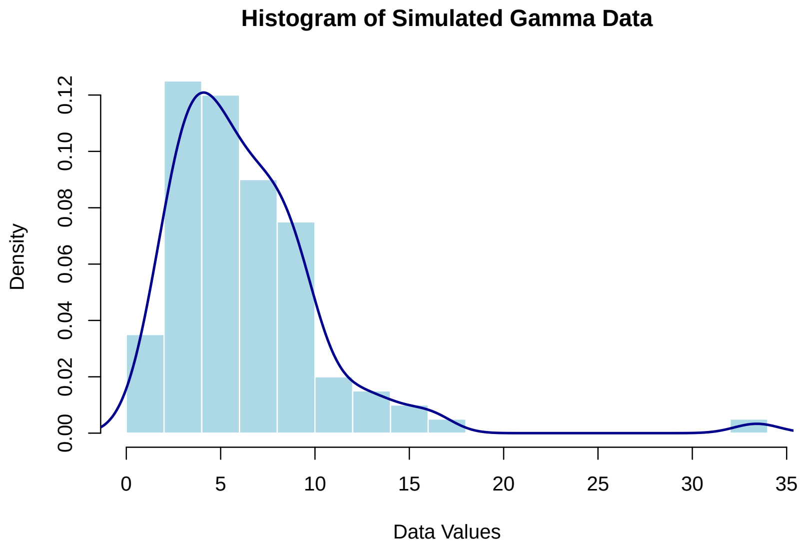

Figure 1: Histogram of simulated Gamma data ($n = 100$, $k = 3$, $\theta = 2$) with kernel density estimate.

4.1.2 Sample Statistics

The computed sample statistics are:

- Sample Mean: $\bar{X} = 6.3157$
- Sample Variance: $S^2 = 18.1508$

For comparison, the theoretical values are $E[X] = k\theta = 6$ and $\text{Var}(X) = k\theta^2 = 12$.

4.1.3 Remarks on the Distribution

The data exhibit **positive skewness** (skewed to the right), which is consistent with the characteristics of the Gamma family of distributions. The **Gamma model is appropriate** for this dataset because:

1. The variable of interest is continuous and strictly positive ($x > 0$).
2. The observed right-skewness matches the theoretical shape of Gamma distributions.
3. The sample statistics are reasonably close to the theoretical moments.

4.2 Point Estimation

4.2.1 (A) Method of Moments (MoM)

Let $X_1, \ldots, X_n$ be i.i.d. random variables from a Gamma distribution with parameters k (shape) and θ (scale). The probability density function is:

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

The first two population moments of the Gamma distribution are:

$$E[X] = k\theta \tag{1}$$

$$\mathrm{Var}(X) = k\theta^2 \tag{2}$$

Derivation of MoM Estimators:

The Method of Moments estimators are obtained by equating the population moments to the sample moments ($\bar{X}$ and S^2).

1. From the expectation equation (1):

$$\mu = k\theta \implies \theta = \frac{\mu}{k}$$

2. Substituting θ into the variance equation (2):

$$\sigma^2 = k\left(\frac{\mu}{k}\right)^2 = \frac{\mu^2}{k}$$

3. Solving for k:

$$k = \frac{\mu^2}{\sigma^2}$$

4. Solving for θ:

$$\theta = \frac{\sigma^2}{\mu}$$

Replacing population moments (μ, σ^2) with sample moments $(\bar{X}, S^2)$, we obtain the **MoM estimators**:

$$\boxed{\hat{k}_{\text{MM}} = \frac{\bar{X}^2}{S^2}} \tag{3}$$

$$\boxed{\hat{\theta}_{\text{MM}} = \frac{S^2}{\bar{X}}} \tag{4}$$

4.2.2 (B) Maximum Likelihood Estimation (MLE)

The likelihood function for the random sample $X_1, \ldots, X_n$ is:

$$L(k, \theta) = \prod_{i=1}^{n} \frac{1}{\Gamma(k)\theta^k} x_i^{k-1} e^{-x_i/\theta}$$

This can be rewritten as:

$$L(k, \theta) = \left[\Gamma(k)\theta^k\right]^{-n} \left(\prod_{i=1}^{n} x_i\right)^{k-1} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n} x_i\right)$$

The **log-likelihood function** $\ell(k, \theta) = \ln L(k, \theta)$ is:

$$\ell(k, \theta) = -n \ln \Gamma(k) - nk \ln \theta + (k-1) \sum_{i=1}^{n} \ln x_i - \frac{1}{\theta} \sum_{i=1}^{n} x_i \tag{5}$$

To find the MLEs, we compute the partial derivatives (score functions) and set them to zero.
1. Derivative with respect to θ:

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

Multiplying by θ^2:

$$-nk\theta + \sum_{i=1}^{n} x_i = 0 \implies \hat{\theta} = \frac{\sum x_i}{nk} = \frac{\bar{X}}{k}$$

Thus:

$$\boxed{\hat{\theta}_{\text{MLE}} = \frac{\bar{X}}{\hat{k}_{\text{MLE}}}} \tag{6}$$

2. Derivative with respect to k:

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

Here, $\psi(k) = \frac{\Gamma'(k)}{\Gamma(k)}$ is the **digamma function**. Substituting $\hat{\theta} = \frac{\bar{X}}{\hat{k}}$:

$$-n\psi(\hat{k}) - n \ln \left(\frac{\bar{X}}{\hat{k}} \right) + \sum_{i=1}^{n} \ln x_i = 0$$

Expanding and rearranging:

$$-n\psi(\hat{k}) - n(\ln \bar{X} - \ln \hat{k}) + \sum_{i=1}^{n} \ln x_i = 0$$

Dividing by n:

$$\boxed{\ln(\hat{k}) - \psi(\hat{k}) = \ln(\bar{X}) - \frac{1}{n} \sum_{i=1}^{n} \ln x_i} \tag{7}$$

Why $\hat{k}_{\text{MLE}}$ has no closed-form solution:
Unlike the MoM estimators, the MLE for k does **not have a closed-form solution**. This is because the parameter k appears nonlinearly inside both the logarithmic function $\ln(k)$ and the digamma function $\psi(k)$. Since $\psi(k)$ is a transcendental function (defined as the derivative of $\ln \Gamma(k)$), it is algebraically impossible to isolate $\hat{k}$ on one side of Equation (7).

Therefore, $\hat{k}_{\text{MLE}}$ must be computed using **numerical optimization methods** (e.g., Newton-Raphson, bisection, or root-finding algorithms such as `uniroot` in R). Once $\hat{k}_{\text{MLE}}$ is found numerically, $\hat{\theta}_{\text{MLE}}$ can be calculated directly using Equation (6).

4.3 Comparison on Observed Data

We apply both MoM and MLE methods to our simulated dataset and compare the results.

4.3.1 Numerical Comparison

Table 1 presents the parameter estimates from both methods alongside the true values.

Table 1: Comparison of Parameter Estimates

Method	**Shape** ($\hat{k}$)	**Scale** ($\hat{\theta}$)
True Values	3.0000	2.0000
MoM	2.1976	2.8739
MLE	2.8685	2.2017

4.3.2 Visual Comparison

Figure 2 shows the histogram of the data with fitted Gamma densities from both estimation methods.

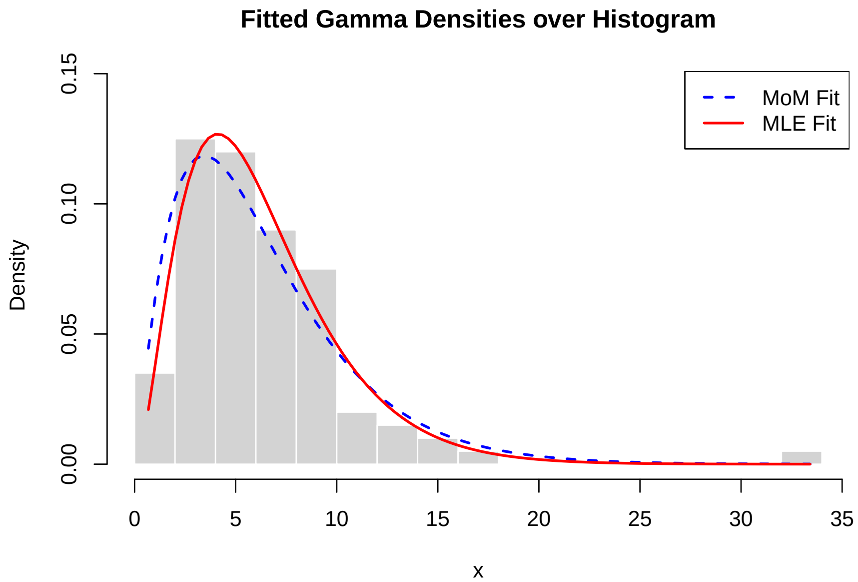

Figure 2: Histogram with fitted Gamma densities using MoM (blue dashed) and MLE (red solid) estimates.

4.3.3 Interpretation

When comparing the fitted densities:

- The **MLE fit** (red solid line) captures the peak and right-skewed tail of the histogram more accurately.
- The **MoM fit** (blue dashed line) underestimates the peak height and does not represent the central tendency as well.
- The MLE estimate for k ($\hat{k}_{\text{MLE}} \approx 2.87$) is closer to the true value ($k = 3$) than the MoM estimate ($\hat{k}_{\text{MoM}} \approx 2.20$).

This demonstrates the effectiveness of MLE in estimating parameters for Gamma distributions, particularly due to its use of the full likelihood function rather than just the first two moments.

5 Simulation Study

In this section, we conduct a Monte Carlo simulation to evaluate the finite-sample performance of MoM and MLE estimators under different scenarios.

5.1 Simulation Design

We consider two parameter settings with different levels of skewness:

- **Scenario 1 (High Skewness):** $k = 1$, $\theta = 2$
- **Scenario 2 (Moderate Skewness):** $k = 5$, $\theta = 1$

For each scenario:

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

5.2 Performance Measures

For each estimator $\hat{\theta}$ and each sample size, we compute:

- **Bias:** $\text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$
- **Variance:** $\text{Var}(\hat{\theta})$
- **Mean Squared Error (MSE):** $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Bias}(\hat{\theta})^2$

5.3 Results

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

Table 2: Simulation Results for Scenario 1 (High Skewness)

Parameter	**Method**	n	**Bias**	**Variance**	**MSE**
k	MoM	20	0.2152	0.2241	0.2705
	MLE	20	0.1446	0.1364	0.1573
	MoM	50	0.0938	0.0789	0.0877
	MLE	50	0.0484	0.0380	0.0403
	MoM	100	0.0435	0.0403	0.0422
	MLE	100	0.0254	0.0175	0.0182
θ	MoM	20	-0.1022	0.7749	0.7853
	MLE	20	-0.0973	0.4686	0.4781
	MoM	50	-0.0527	0.3528	0.3556
	MLE	50	-0.0356	0.2026	0.2039
	MoM	100	-0.0092	0.1982	0.1983
	MLE	100	-0.0176	0.0980	0.0983

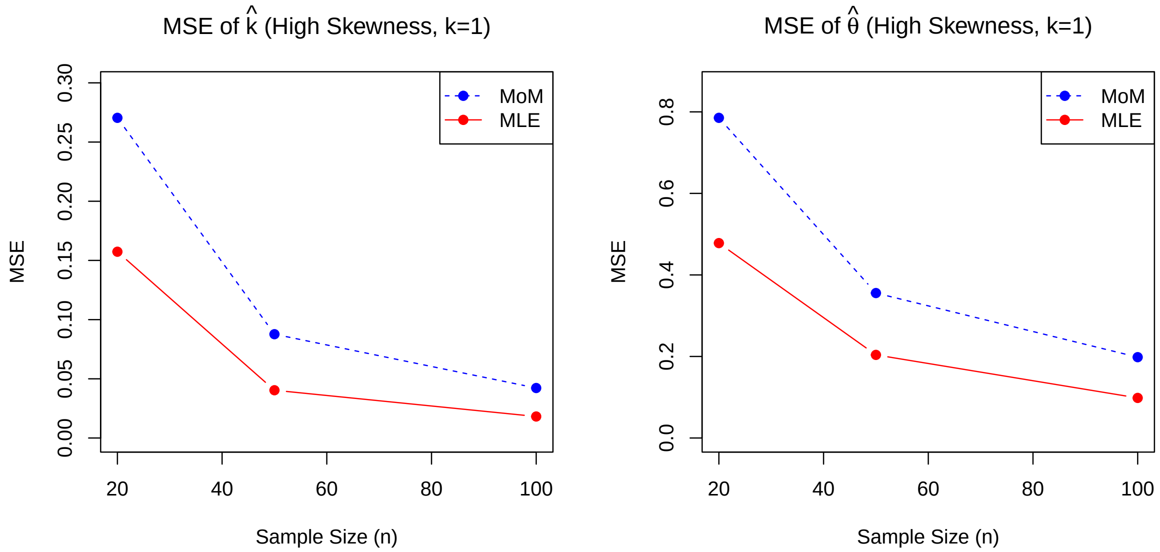

Figure 3: MSE versus sample size for Scenario 1 (High Skewness, $k = 1$, $\theta = 2$).

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

Table 3: Simulation Results for Scenario 2 (Moderate Skewness)

Parameter	Method	n	Bias	Variance	MSE
k	MoM	20	0.6558	4.6534	5.0834
	MLE	20	0.8394	4.6231	5.3277
	MoM	50	0.2162	1.2721	1.3188
	MLE	50	0.2609	1.0813	1.1494
	MoM	100	0.1186	0.6520	0.6661
	MLE	100	0.1375	0.5350	0.5539
θ	MoM	20	0.0017	0.1370	0.1370
	MLE	20	-0.0437	0.1045	0.1064
	MoM	50	0.0012	0.0487	0.0487
	MLE	50	-0.0149	0.0392	0.0394
	MoM	100	0.0011	0.0260	0.0260
	MLE	100	-0.0073	0.0206	0.0206

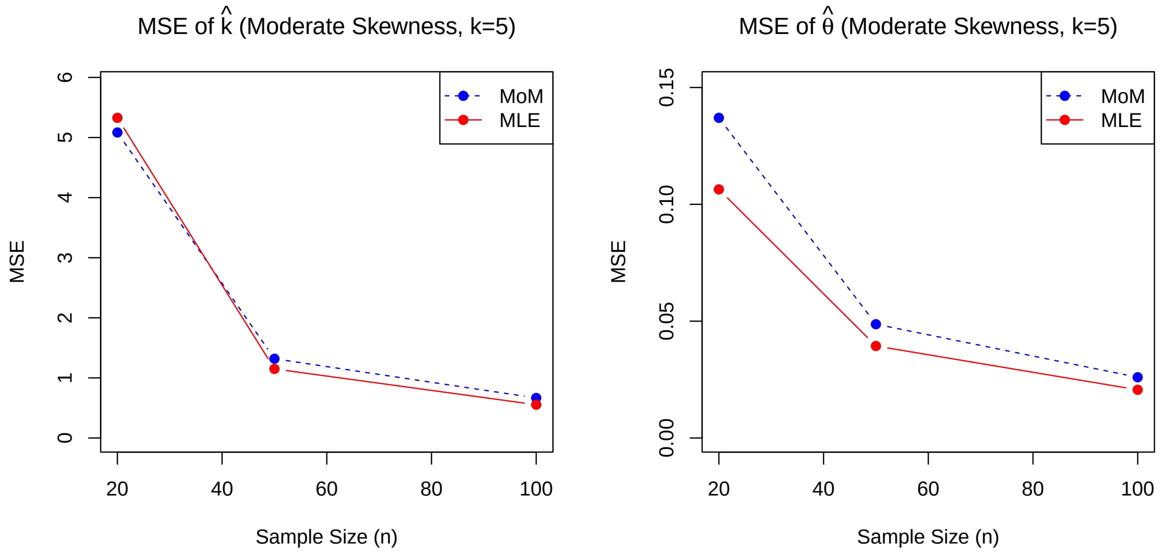

Figure 4: MSE versus sample size for Scenario 2 (Moderate Skewness, $k = 5$, $\theta = 1$).

6 Discussion and Conclusions

Based on the simulation results presented in Tables 2–3 and Figures 3–4, we draw the following conclusions:

6.1 Effect of Sample Size

The simulation results clearly support the **consistency property** for both estimators. As the sample size increases from $n = 20$ to $n = 100$:

- Bias decreases toward zero for both MoM and MLE.
- Variance decreases substantially.
- MSE decreases, indicating improved estimation accuracy.

This demonstrates that both approaches converge to the true parameter values as more data become available.

6.2 Comparison: MoM vs. MLE

The comparison between MoM and MLE reveals nuanced results that depend on the scenario:

Scenario 1 (High Skewness, $k = 1$):

- MLE **consistently outperforms** MoM across all sample sizes.
- At $n = 20$, the MSE for $\hat{k}_{\text{MLE}}$ (0.1573) is substantially smaller than for $\hat{k}_{\text{MoM}}$ (0.2705).
- MLE achieves lower variance, which compensates for any slight differences in bias.

Scenario 2 (Moderate Skewness, $k = 5$):

- The results are more nuanced. For the shape parameter k, MoM actually shows slightly lower MSE than MLE at small sample sizes ($n = 20, 50$).
- For the scale parameter θ, MLE maintains an advantage across all sample sizes.

- As sample size increases, MLE's variance advantage becomes more apparent.

The superior performance of MLE in highly skewed distributions can be attributed to its use of the **full likelihood function**, which extracts more information from the data compared to MoM, which only uses the first two moments. However, in less skewed distributions with small samples, MoM can be competitive.

6.3 Effect of Skewness

The degree of skewness significantly affects estimation difficulty and the relative performance of the estimators:

- In **Scenario 1** ($k = 1$, highly skewed exponential-like distribution), MLE performed notably better than MoM across all sample sizes. The high skewness means that higher-order information captured by the likelihood is valuable.
- In **Scenario 2** ($k = 5$, more symmetric), the situation is more complex. For small samples, MoM can actually outperform MLE for estimating k, though MLE maintains an advantage for θ.
- This suggests that when the distribution is closer to symmetric, the first two moments capture most of the relevant information, making MoM more competitive.

The key insight is that MLE's advantage is most pronounced when the underlying distribution is highly skewed.

6.4 Practical Recommendations

For practitioners working with Gamma-distributed data:

1. **For highly skewed data ($k \approx 1$):** Use MLE, as it provides substantially better estimates across all sample sizes.
2. **For moderately skewed data ($k \geq 5$):** Both methods perform reasonably well. MoM may be preferred for simplicity when sample sizes are small, while MLE shows advantages for the scale parameter.
3. **MoM as a starting point:** The closed-form MoM estimators can serve as excellent initial values for numerical optimization algorithms when computing MLEs.
4. **Sample size considerations:** For small samples ($n < 30$), both methods show noticeable bias. Larger samples ($n \geq 100$) yield substantially more reliable estimates regardless of the method chosen.
5. **Software implementation:** Modern statistical software (R, Python) provides efficient numerical routines for computing MLEs, making the computational overhead of MLE negligible in practice.

A R Code

The complete R code used for this project is provided below. This code generates all figures and performs the simulation study. It can also be found in the separate file `figures.R`.

A.1 Setup and Data Generation (Section 4.1)

```
# Setting seed for reproducibility
set.seed(303)

# Create figures directory if it doesn't exist
if (!dir.exists("figures")) {
  dir.create("figures")
}

# True parameters for descriptive statistics
true_k <- 3
true_theta <- 2
n <- 100

# Generate data
data_sample <- rgamma(n, shape = true_k, scale = true_theta)

# Calculate sample statistics
sample_mean <- mean(data_sample)
sample_var <- var(data_sample)

cat("Sample Mean (X_bar):", round(sample_mean, 4), "\n")
cat("Sample Variance (S^2):", round(sample_var, 4), "\n")

# Save histogram
pdf("figures/histogram_descriptive.pdf", width = 7, height = 5)
hist(data_sample, breaks = 15, probability = TRUE,
     main = "Histogram of Simulated Gamma Data",
     xlab = "Data Values", col = "lightblue", border = "white")
lines(density(data_sample), col = "darkblue", lwd = 2)
dev.off()
```

A.2 Method of Moments Estimation (Section 4.2A)

```
# MoM Formulas: k = mean^2 / var, theta = var / mean
k_mom <- (mean(data_sample)^2) / var(data_sample)
theta_mom <- var(data_sample) / mean(data_sample)
```

A.3 Maximum Likelihood Estimation (Section 4.2B)

```
# MLE: Solve log(k) - psi(k) = log(x_bar) - mean(log(x))
target_val <- log(mean(data_sample)) - mean(log(data_sample))

mle_eqn <- function(k) {
  log(k) - digamma(k) - target_val
}
```

```
mle_root <- uniroot(mle_eqn, interval = c(0.1, 20), extendInt = "yes")
k_mle <- mle_root$root
theta_mle <- mean(data_sample) / k_mle
```

A.4 Fitted Densities Plot (Section 4.3)

```
pdf("figures/fitted_densities.pdf", width = 7, height = 5)
hist(data_sample, probability = TRUE, breaks = 15,
     main = "Fitted Gamma Densities over Histogram",
     xlab = "x", col = "lightgray", border = "white",
     ylim = c(0, max(density(data_sample)$y) * 1.2))

x_vals <- seq(min(data_sample), max(data_sample), length.out = 100)

lines(x_vals, dgamma(x_vals, shape = k_mom, scale = theta_mom),
      col = "blue", lwd = 2, lty = 2)
lines(x_vals, dgamma(x_vals, shape = k_mle, scale = theta_mle),
      col = "red", lwd = 2)

legend("topright", legend = c("MoM Fit", "MLE Fit"),
       col = c("blue", "red"), lty = c(2, 1), lwd = 2)
dev.off()
```

A.5 Simulation Study (Section 5)

```
R <- 2000     # Number of replications
sample_sizes <- c(20, 50, 100)

scenarios <- list(
  "Scenario 1" = c(k = 1, theta = 2),  # High Skewness
  "Scenario 2" = c(k = 5, theta = 1)   # Moderate Skewness
)

results_df <- data.frame()

for (scen_name in names(scenarios)) {
  true_k <- scenarios[[scen_name]]["k"]
  true_theta <- scenarios[[scen_name]]["theta"]

  for (n in sample_sizes) {
    k_mom_est <- numeric(R); t_mom_est <- numeric(R)
    k_mle_est <- numeric(R); t_mle_est <- numeric(R)

    for (i in 1:R) {
      # Generate Data
      x <- rgamma(n, shape = true_k, scale = true_theta)

      # MoM Estimation
      x_bar <- mean(x)
      s2 <- var(x)
      k_hat_mom <- (x_bar^2) / s2
```

```
      t_hat_mom <- s2 / x_bar
      k_mom_est[i] <- k_hat_mom
      t_mom_est[i] <- t_hat_mom

      # MLE Estimation
      target <- log(x_bar) - mean(log(x))
      try({
        mle_sol <- uniroot(function(k) log(k) - digamma(k) - target,
                           interval = c(1e-5, 100), extendInt = "yes")
        k_hat_mle <- mle_sol$root
        t_hat_mle <- x_bar / k_hat_mle
        k_mle_est[i] <- k_hat_mle
        t_mle_est[i] <- t_hat_mle
      }, silent = TRUE)
    }

    # Performance Metrics Calculation
    calc_metrics <- function(est, true_val) {
      bias <- mean(est, na.rm=TRUE) - true_val
      variance <- var(est, na.rm=TRUE)
      mse <- variance + bias^2
      return(c(bias, variance, mse))
    }

    mom_k <- calc_metrics(k_mom_est, true_k)
    mle_k <- calc_metrics(k_mle_est, true_k)
    mom_t <- calc_metrics(t_mom_est, true_theta)
    mle_t <- calc_metrics(t_mle_est, true_theta)

    temp_res <- data.frame(
      Scenario = scen_name, n = n,
      Method = rep(c("MoM", "MLE"), 2),
      Parameter = rep(c("k", "theta"), each = 2),
      Bias = c(mom_k[1], mle_k[1], mom_t[1], mle_t[1]),
      Variance = c(mom_k[2], mle_k[2], mom_t[2], mle_t[2]),
      MSE = c(mom_k[3], mle_k[3], mom_t[3], mle_t[3])
    )
    results_df <- rbind(results_df, temp_res)
  }
}
```

A.6 MSE Plots (Section 5)

```
# MSE Plots for Scenario 1 (High Skewness)
pdf("figures/mse_scenario1.pdf", width = 10, height = 5)
par(mfrow = c(1, 2))

plot_data <- subset(results_df, Scenario == "Scenario 1" & Parameter == "k")
mom_data <- subset(plot_data, Method == "MoM")
mle_data <- subset(plot_data, Method == "MLE")

plot(mom_data$n, mom_data$MSE, type = "b", col = "blue", pch = 19, lty = 2,
```

```
     ylim = c(0, max(plot_data$MSE)*1.1),
     xlab = "Sample Size (n)", ylab = "MSE",
     main = expression(paste("MSE of ", hat(k), " (High Skewness)")))
lines(mle_data$n, mle_data$MSE, type = "b", col = "red", pch = 19, lty = 1)
legend("topright", legend = c("MoM", "MLE"),
       col = c("blue", "red"), lty = c(2, 1), pch = 19)

plot_data_t <- subset(results_df, Scenario == "Scenario 1" & Parameter == "theta")
mom_data_t <- subset(plot_data_t, Method == "MoM")
mle_data_t <- subset(plot_data_t, Method == "MLE")

plot(mom_data_t$n, mom_data_t$MSE, type = "b", col = "blue", pch = 19, lty = 2,
     ylim = c(0, max(plot_data_t$MSE)*1.1),
     xlab = "Sample Size (n)", ylab = "MSE",
     main = expression(paste("MSE of ", hat(theta), " (High Skewness)")))
lines(mle_data_t$n, mle_data_t$MSE, type = "b", col = "red", pch = 19, lty = 1)
legend("topright", legend = c("MoM", "MLE"),
       col = c("blue", "red"), lty = c(2, 1), pch = 19)
dev.off()

# Similar code for Scenario 2 (Moderate Skewness)
# ... (same structure, filtering for "Scenario 2")
```