

# Lab 8 – MATH 240 – Computational Statistics

Avery Johnson  
Colgate University  
Department of Mathematics  
aqjohnson@colgate.edu

## Abstract

The Beta distribution is a flexible probability distribution on  $[0, 1]$ , commonly used for modeling proportions and probabilities. This lab examines how graphical and numerical summaries of Beta-distributed samples connect to the underlying population distribution, particularly as sample size increases. Two estimation methods, the Method of Moments (MOM) and Maximum Likelihood Estimation (MLE), are compared using death rate data. Results show that MLE provides more precise estimates for  $\alpha$ , demonstrating a slight advantage over MOM, although both yield similar results.

**Keywords:** Beta distribution, moments, mean, variance, skewness, kurtosis, MOM estimator, MLE estimator.

## 1 Introduction

The Beta distribution is a continuous probability distribution defined on  $[0, 1]$ , commonly used to model proportions, probabilities, and rates. Controlled by two parameters,  $\alpha$  and  $\beta$ , its shape ranges from highly symmetric to highly skewed, making it flexible for various statistical applications. This lab investigates the properties of the Beta distribution, compares population and sample-based estimates, and applies parameter estimation techniques to real-world data. Additionally, the effectiveness of MOM and MLE in estimating  $\alpha$  and  $\beta$  is assessed through simulation.

## 2 Density Functions and Parameters

The probability density function (PDF) of a beta distribution is given by:

$$f_X(x | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma\alpha\Gamma\beta} x^{\alpha-1} (1-x)^{\beta-1} I(x \in [0, 1])$$

where  $I(x \in [0, 1]) = 1$  when  $x \in [0, 1]$  and 0 otherwise.

To visualize the impact of the parameters  $\alpha$  and  $\beta$  on the distribution, four cases were examined: **Beta(2,5)**, **Beta(5,5)**, **Beta(5,2)**, and **Beta(0.5,0.5)**. The first distribution is right-skewed, the second is symmetric, the third is left-skewed, and the fourth is U-shaped. The plots in Figure 1 demonstrate this, illustrating that increasing  $\alpha$  shifts the distribution towards 1, while increasing  $\beta$  shifts it towards zero. When  $\alpha = \beta$ , the distribution is symmetric.

## 3 Properties

Several important properties of the Beta distribution include the mean, variance, skewness, and excess kurtosis. These values were calculated for each Beta distribution using established mathematical formulas (Table 1). To verify accuracy, numerical integration was performed to compute moments directly from the PDF using a custom function, **beta.moment()**, which calculated both the centered and uncentered moments. The results aligned with theoretical expectations.

Random samples of size  $n = 500$  were generated from each Beta distribution. Histograms were plotted alongside the true PDFs in Figure 2, and numerical summaries were computed (Table 2). Comparing Table 2 with Table 1, along with analyzing Figure 2, confirms that sample estimates closely approximate theoretical values.

To analyze sample size effects, cumulative statistics were plotted for the **Beta(2,5)** distribution using the **cumstats** package in R (Erdely and Castillo, 2017). A single sample was first examined, followed by a simulation using a **for()** loop to generate data sets of size 500, adding new cumulative statistics lines at each iteration. The results (Figure 3 in the Appendix) show that variability is high for small samples but decreases as  $n$  increases, illustrating convergence to population values. Results indicate the importance of sample size.

alpha	beta	mean	var	skew	kurtosis
2.00	5.00	0.29	0.03	0.60	-0.12
5.00	5.00	0.50	0.02	0.00	-0.46
5.00	2.00	0.71	0.03	-0.60	-0.12
0.50	0.50	0.50	0.12	0.00	-1.50

Table 1: Population-Level Summary

alpha	beta	mean	var	skew	kurtosis
2.00	5.00	0.29	0.03	0.57	-0.23
5.00	5.00	0.50	0.02	-0.06	-0.58
5.00	2.00	0.71	0.03	-0.57	-0.23
0.50	0.50	0.51	0.13	-0.02	-1.51

Table 2: Sample Summary

## 4 Estimators

Since the true value of  $\alpha$  and  $\beta$  are unknown in real-world applications, parameter estimation methods are essential. Two widely used techniques are the Method of Moments (MOM) and Maximum Likelihood Estimation (MLE). MOM estimates parameters by equating sample moments to their corresponding population moments. Using the `nleqslv` package in R, parameters were chosen to match sample and population moments. MLE is a more general and widely used technique for parameter estimation. It finds the values of  $\alpha$  and  $\beta$  that maximize the likelihood of the observed data. The likelihood function is a product of the Beta PDF evaluated at each data point, but computationally, it is easier to work with the log-likelihood function. The `optim()` function in R was used to maximize the log-likelihood. MLE is generally regarded as more efficient since it often provides more precise estimates.

## 5 Death Rates Example

The flexibility of the Beta distribution was tested by modeling real-world death rate data from the world bank (Fatih, 2022). The data, representing deaths per 1,000 citizens, was transformed into a rate and fitted to a Beta distribution using both MOM and MLE. Parameter estimates were obtained by solving moment equations for MOM and optimizing the likelihood function for MLE (Table 3). While the estimates for  $\alpha$  were similar across methods, the estimates for  $\beta$  differed slightly.

Method	Alpha	Beta
MOM	8.43	1003.46
MLE	8.27	985.05

Table 3: MOM and MLE Estimates

To visualize the results, histograms of the empirical data were plotted with the estimated Beta distributions superimposed (Figure 5). Both methods produced reasonable fits. To determine which estimators we should use, then, we computed bias, precision, and mean squared error (Table 4). Density plots (Figure 6) further helped us visualize estimator differences. The results indicate that while MLE provided more narrow (precise) estimates for  $\alpha$ , both methods performed similarly overall.

## 6 Conclusion

This study illustrates the flexibility of the Beta distribution and its usefulness in modeling proportion-based data. Graphical and numerical summaries show that sample estimates converge to population values as sample size increases. Additionally, the comparison of MOM and MLE estimators reveals that while both methods yield accurate and comparable results, MLE generally provides more precise parameter estimates. These findings reinforce the importance of selecting appropriate estimation techniques depending on the application and data characteristics.

## References

- Erdelyi, A. and Castillo, I. (2017). *cumstats: Cumulative Descriptive Statistics*. R package version 1.0.  
 Fatih, C. (2022). Determinants of mortality rates from covid-19: A macro level analysis by extended-beta regression model. *Revista de Salud Pública*, 24(2):1.

## 7 Appendix

### 7.1 Density Function and Parameters

#### 7.1.1 Population Distribution Figures

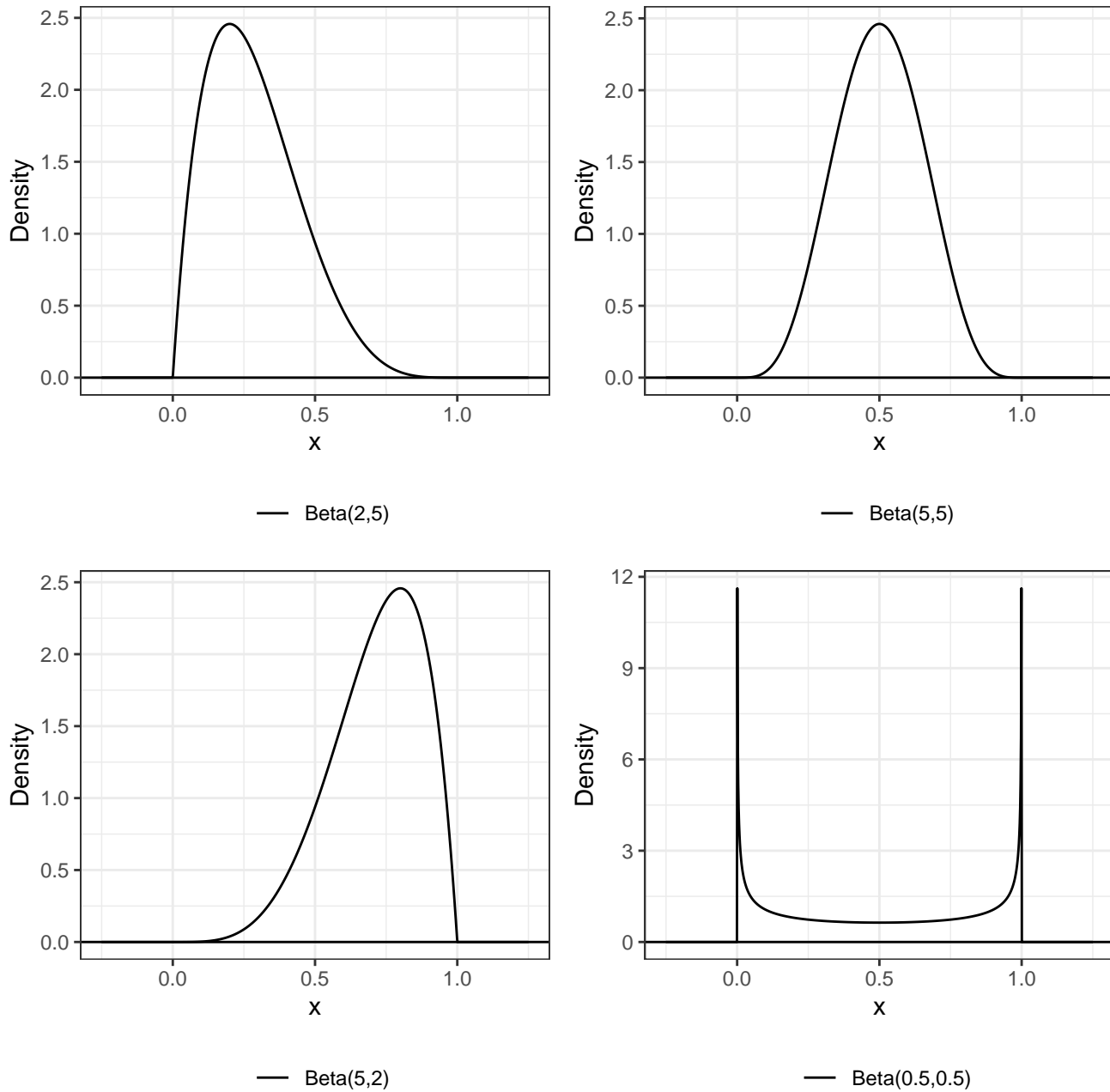


Figure 1: Population Figures for Various Beta Distributions

## 7.2 Properties

### 7.2.1 Sample Distribution Figures

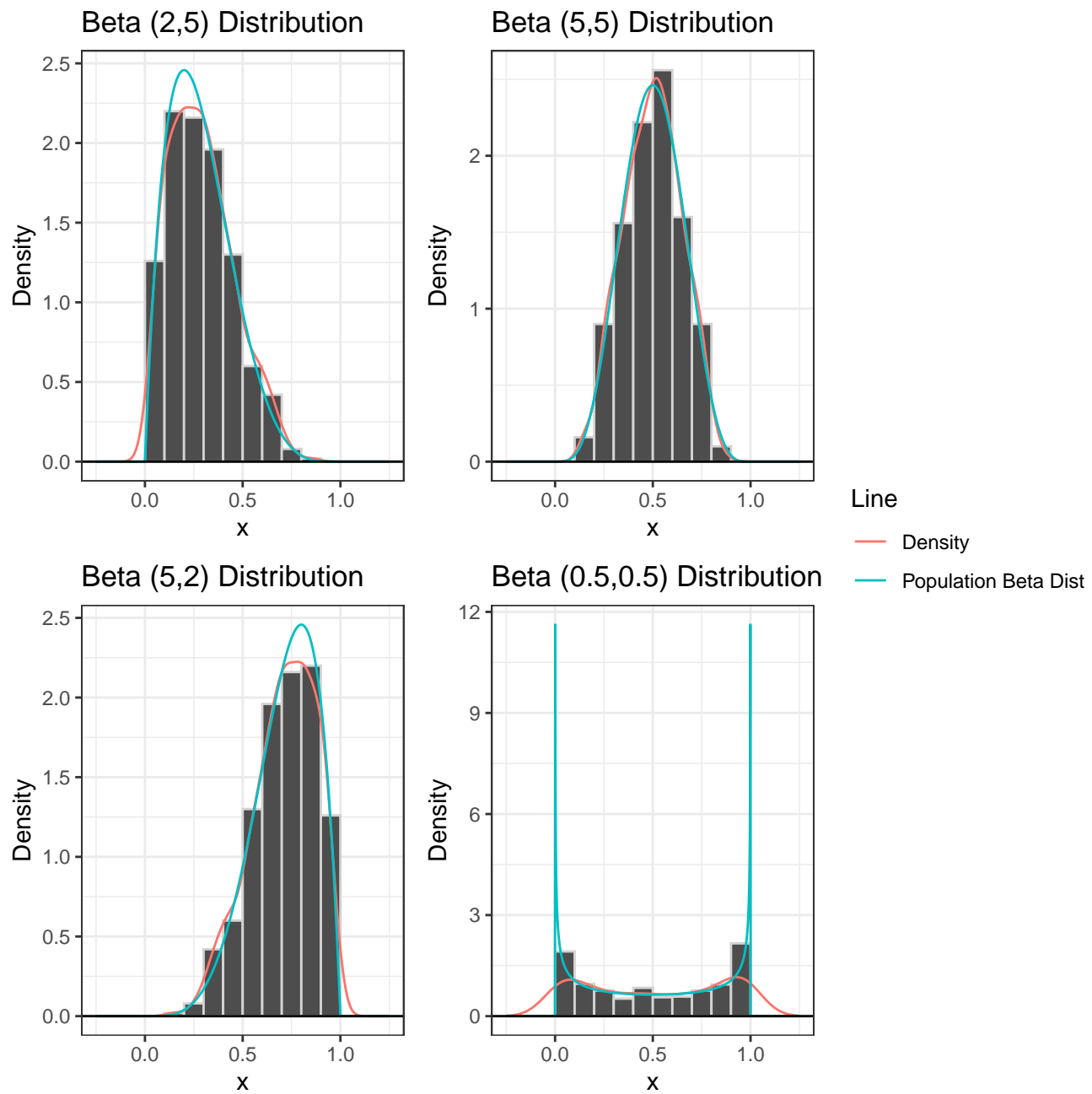


Figure 2: Sample Figures for Various Beta Distributions

### 7.2.2 Cumulative Statistics: Does Sample Size Matter?

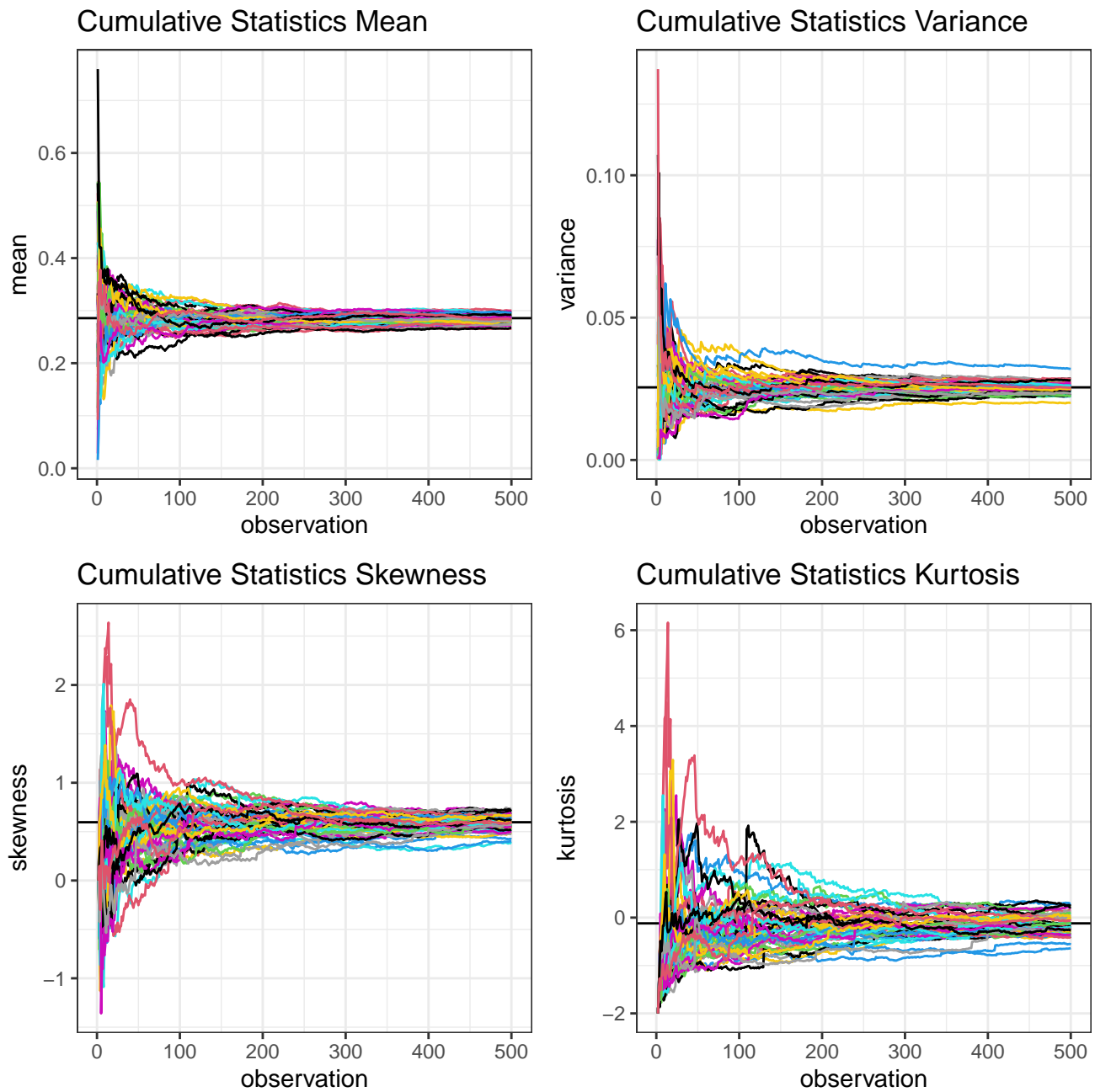


Figure 3: Cumulative Statistics as Sample Size Increases

### 7.2.3 Modeling the Variation

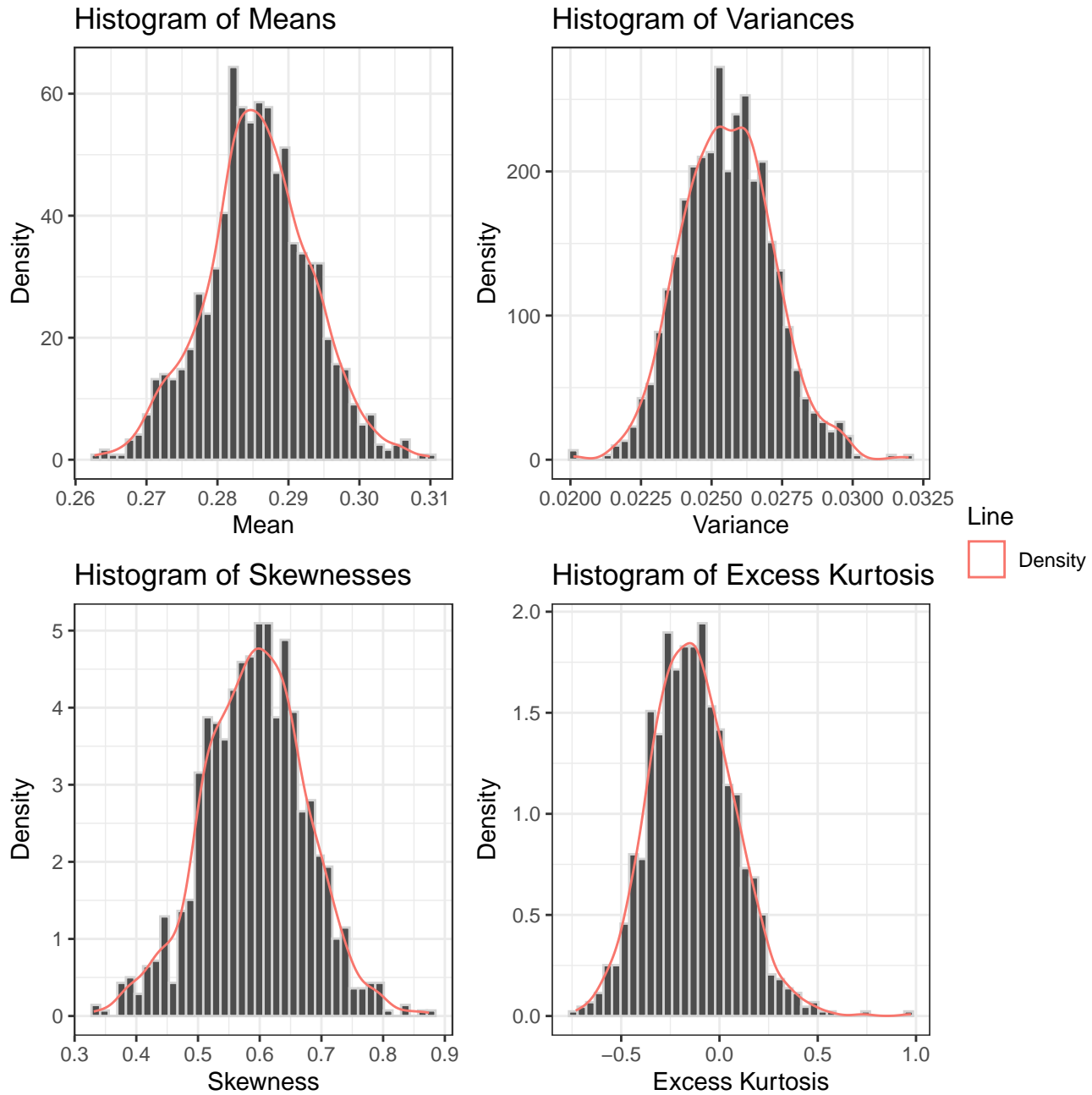


Figure 4: Modeling the Variation of Statistics

### 7.3 Estimators

### 7.4 Example

#### 7.4.1 MOM vs. MLE Estimates

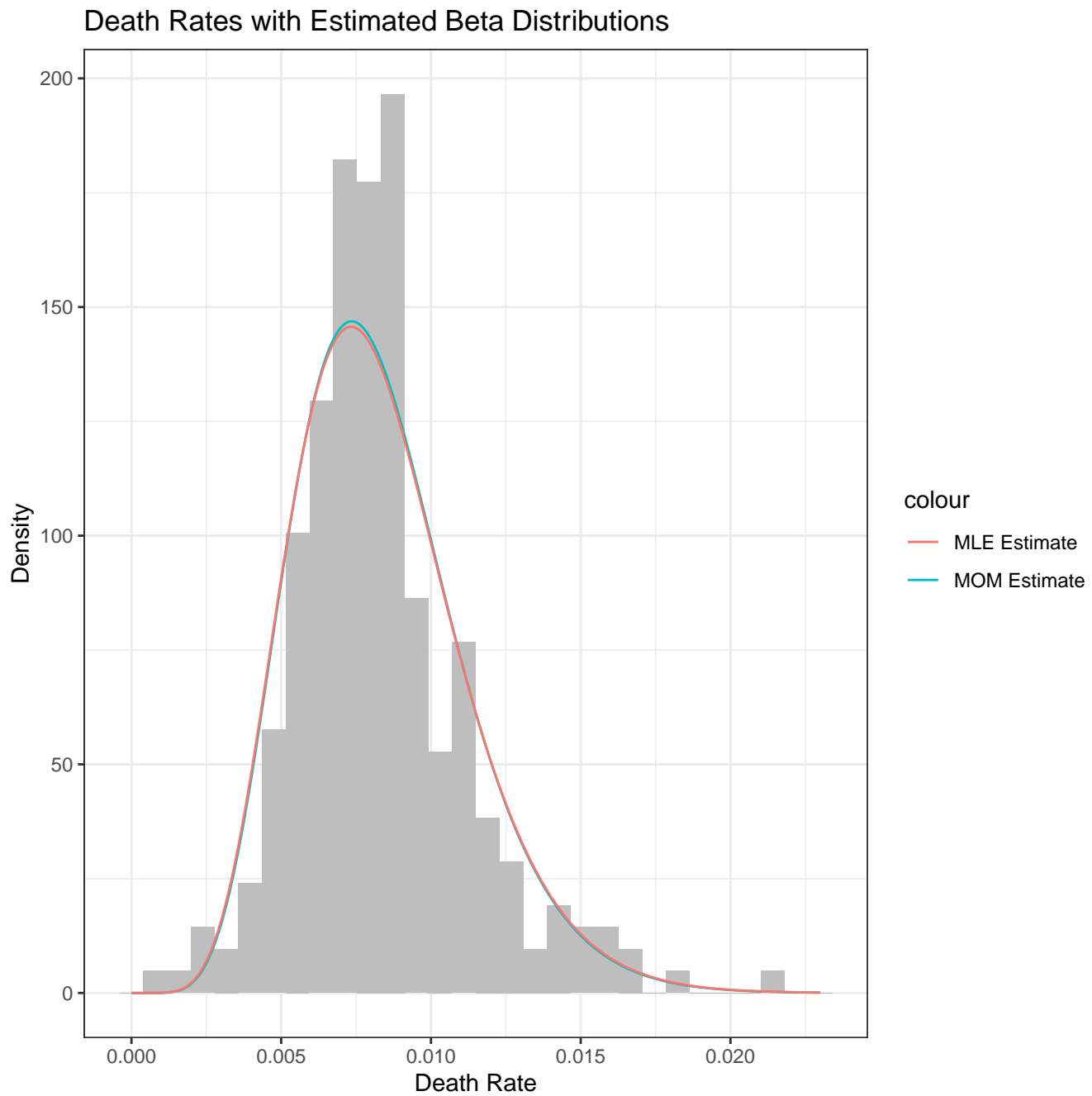


Figure 5: Estimated Beta Distributions

### 7.4.2 Bias, Precision, and Mean Squared Error of Estimates

Parameter	Bias_MOM	Bias_MLE	Precision_MOM	Precision_MLE	MSE_MOM	MSE_MLE
Alpha	0.08	0.07	1.83	2.13	0.55	0.48
Beta	10.29	9.11	0.00	0.00	8288.46	7132.70

Table 4: Bias, Precision, and Mean Squared Error



### 7.4.3 Estimated Density for MOM and MLE

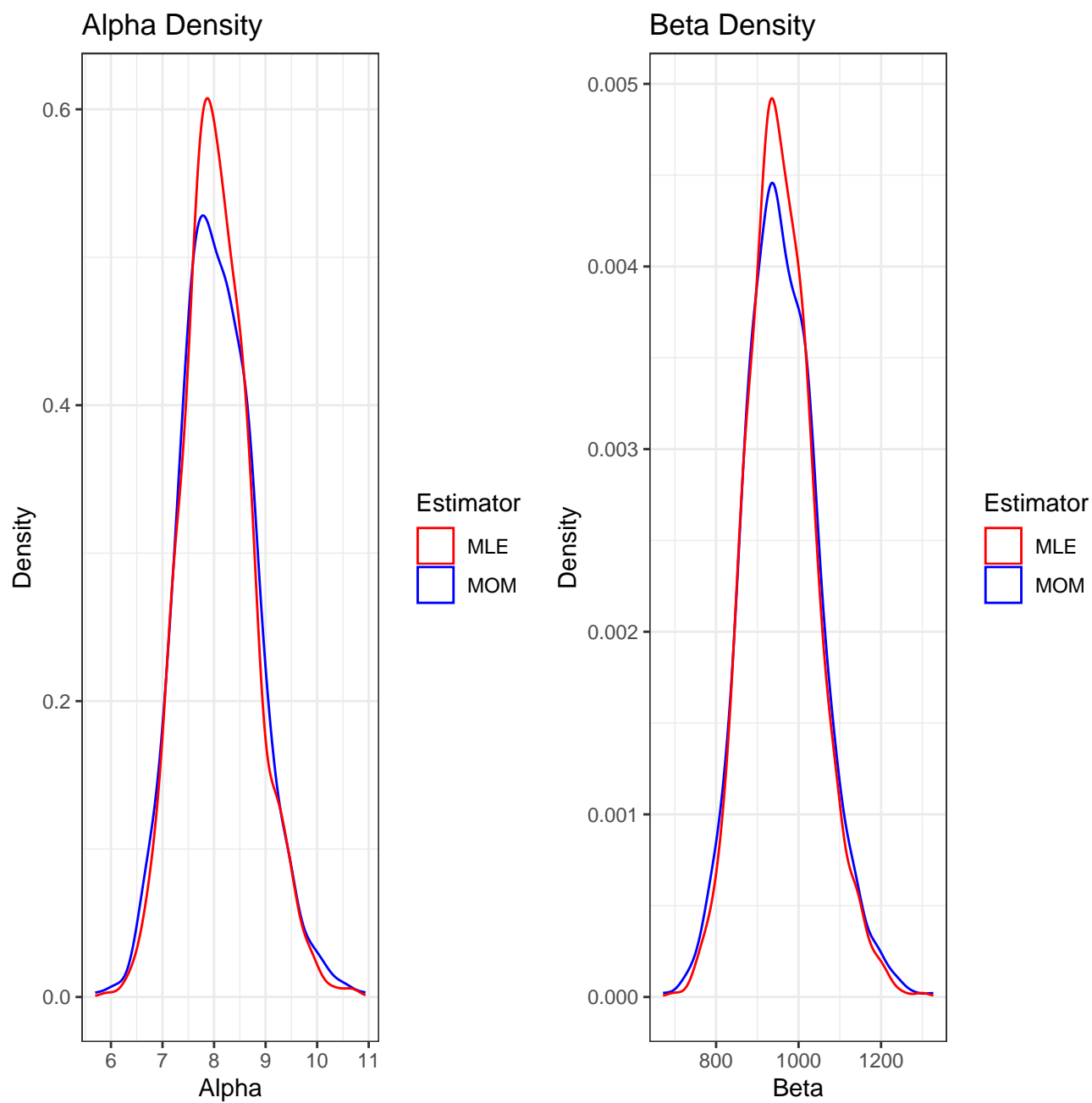


Figure 6: Estimated Densities for MOM and MLE