Lab 7 and 8 – MATH 240 – Computational Statistics

Camilo Granada Cossio
Colgate University
Mathematics Department
cgranadacossio@colgate.edu

Abstract

This lab explored the statistical properties and applications of the Beta distribution using `R`. Derivations and simulations were conducted to describe the shape, behavior, and use cases of the distribution. Using `tidyverse` (Wickham et al., 2019) and `cumstats` (Erdely and Castillo, 2017), key properties such as mean, variance, skewness, and kurtosis were derived and compared to sdample-based summaries. Parameter estimates were generated using the Method of Moments and Maximum Likelihood Estimation. These estimation methods were evaluated through simulations and applied to global death rate data from the World Bank (2022).

Keywords: Beta distribution; Parameter estimation; Simulation; Method of Moments; Maximum Likelihood Estimation

1 Introduction

The Beta distribution is a flexible model for continuous variables bounded between 0 and 1, often used for proportions, probabilities, and rates. It is defined by two shape parameters α and β, which determine the distributions form. Depending on these parameters, the distribution can appear left-skewed, right-skewed, symetric, or U-shaped.

Building on this, this lab investigates the Beta distribution by deriving its key statistics using closed-form expressions for the mean, variance, skewness, and excess kurtosis. These statistics are explored under different parameter values to understand how α and β influence distribution behavior. The lab also examines how sample-based summaries compare to population characteristics and evaluates the Method of Moments and Maximum Likelihood Estimation through simulations and real-world data.

2 Density Functions and Parameters

The probability density function (PDF) of the Beta distributions is:

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for } x \in [0,1],$$

where $\alpha, \beta > 0$ and $\Gamma(\cdot)$ is the gamma function.

The shape of the distribution is fully determined by α and β. The following examples illustrate its flexibility:

- Beta(2, 5): Right-skewed
- Beta(5, 5): Symmetric
- Beta(5, 2): Left-skewed
- Beta(0.5, 0.5): U-shaped

3 Properties

Several population-level characteristics of the Beta distribution are derived from its PDF:

$$\text{Mean: } \mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$$

$$\text{Variance: } \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$\text{Skewness: } \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}}$$

$$\text{Excess Kurtosis: } \frac{6[(\alpha-\beta)^2(\alpha+\beta+1)-\alpha\beta(\alpha+\beta+2)]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$$

These theoretical values were compared to those computed using `R` functions for each formula. The `cumstats` and `tidyverse` packages were used to summarize the behavior of these statistics from samples.

4 Estimators

To estimate α and β from the data, two approaches were used:

- **Method of Moments (MOM):** Matches sample moments to theoretical moments
- **Maximum Likelihood Estimation (MLE):** Optimizes the log-likelihood based on the PDF.

Simulations with $n = 1000$ showed that MLE offers improved performance over MOM, including higher precision and lower error. The quality of the estimators was assessed using:

$$\text{Bias: } \mathbb{E}(\hat{\theta}) - \theta$$

$$\text{Precision: } \frac{1}{\mathrm{Var}(\hat{\theta})}$$

$$\text{MSE: } \mathrm{Var}(\hat{\theta}) + \text{Bias}^2$$

5 Example: Death Rates Data

Estimation procedures were applied to country-level death rates per 1000 people (World Bank, 2022). Both MOM and MLE were used to fit Beta distributions.

To further examine estimator performance, 1000 samples of size $n = 266$ were simulated from a Beta(8, 950) distribution. Results:

Estimator	Bias	Precision	MSE
MOM α	0.0827	1.8281	0.5539
MOM β	10.4071	0.0001222	8294.1232
MLE α	0.0720	2.1273	0.4753
MLE β	9.1136	0.0001418	7133.5690

These results support the conclusion that MLE generally produces lower error and greater precision than MOM, particularly when estimating α. However, both estimators struggled with bias when estimating β

(Pedersen, 2024)

References

Erdely, A. and Castillo, I. (2017). *cumstats: Cumulative Descriptive Statistics.* R package version 1.0.

Pedersen, T. L. (2024). *patchwork: The Composer of Plots.* R package version 1.3.0.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

6 Appendix

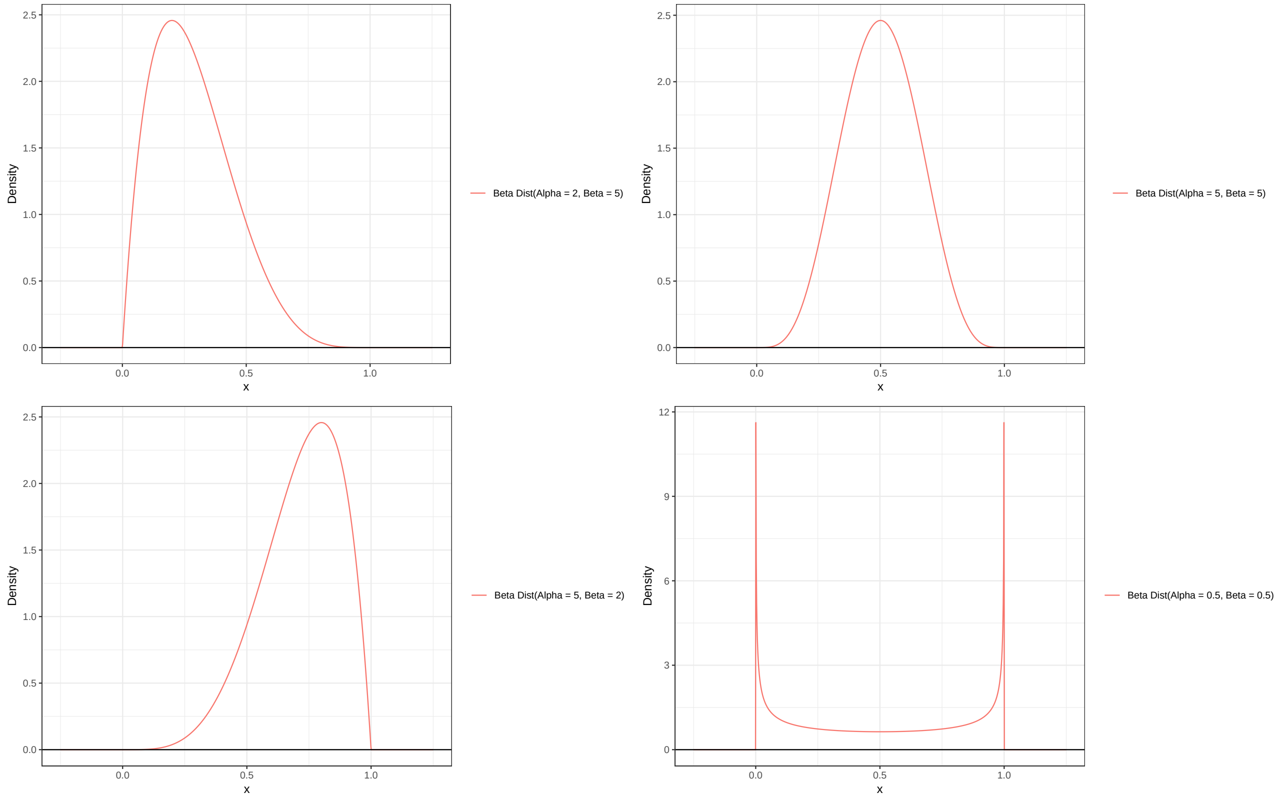

Figure 1: Probability Density Functions for Different Beta Distributions

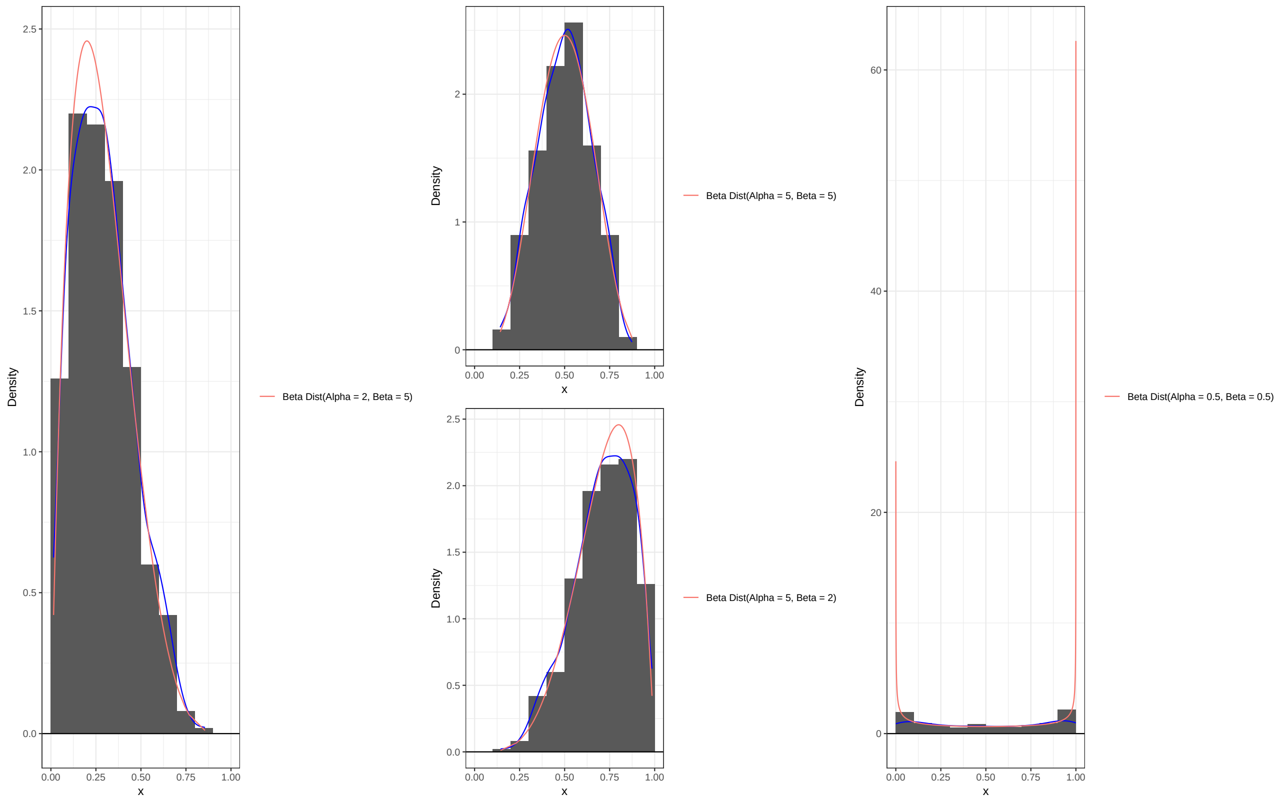

Figure 2: Histograms and Estimated Densities from Sample Data Compared to True PDFs

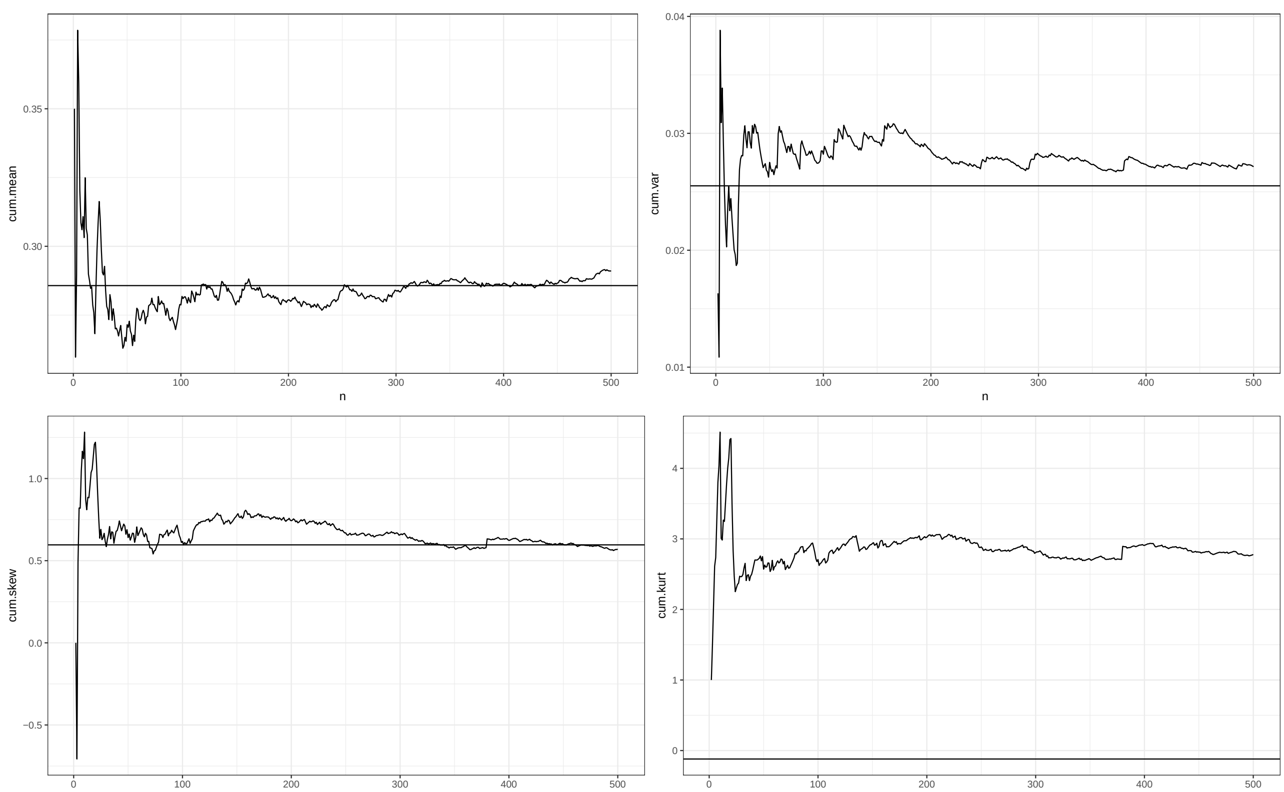

Figure 3: Cumulative Statistics for Beta(2, 5) Sample

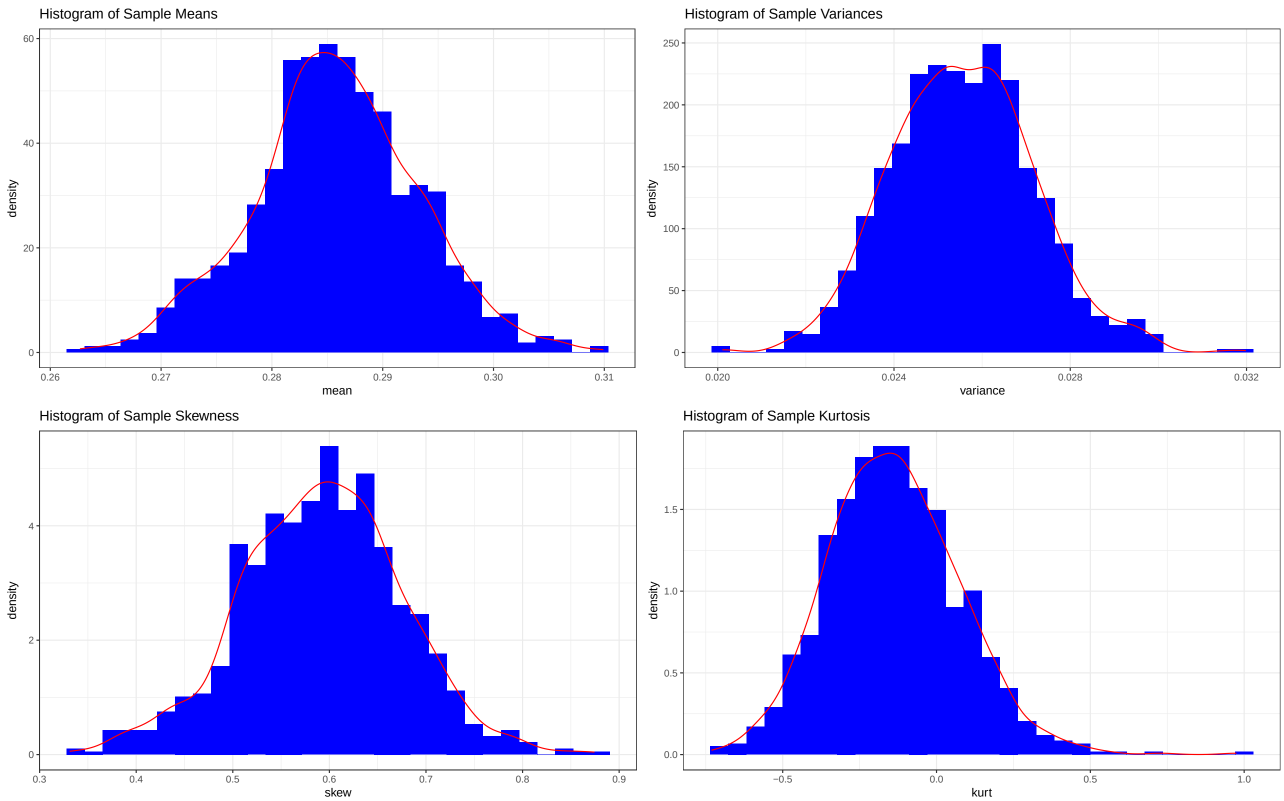

Figure 4: Sampling Distributions from 1000 Simulated Samples

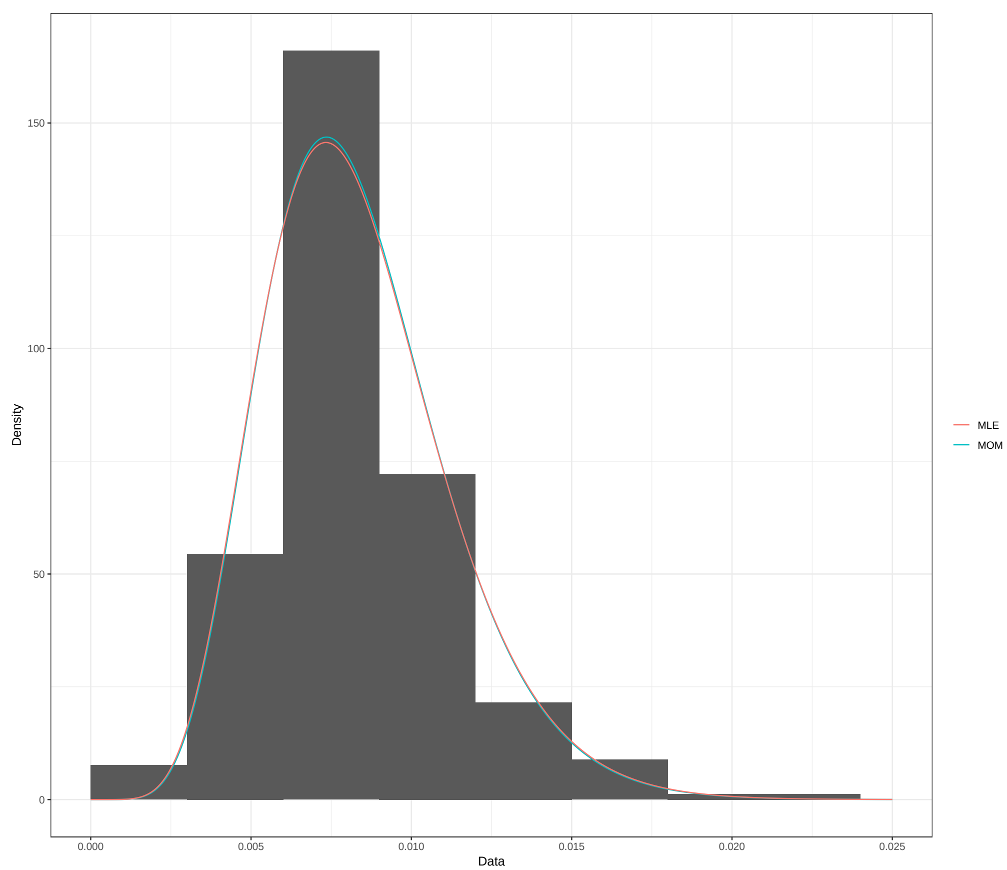

Figure 5: Histogram of Real Death Rate Data with MOM and MLE Fits

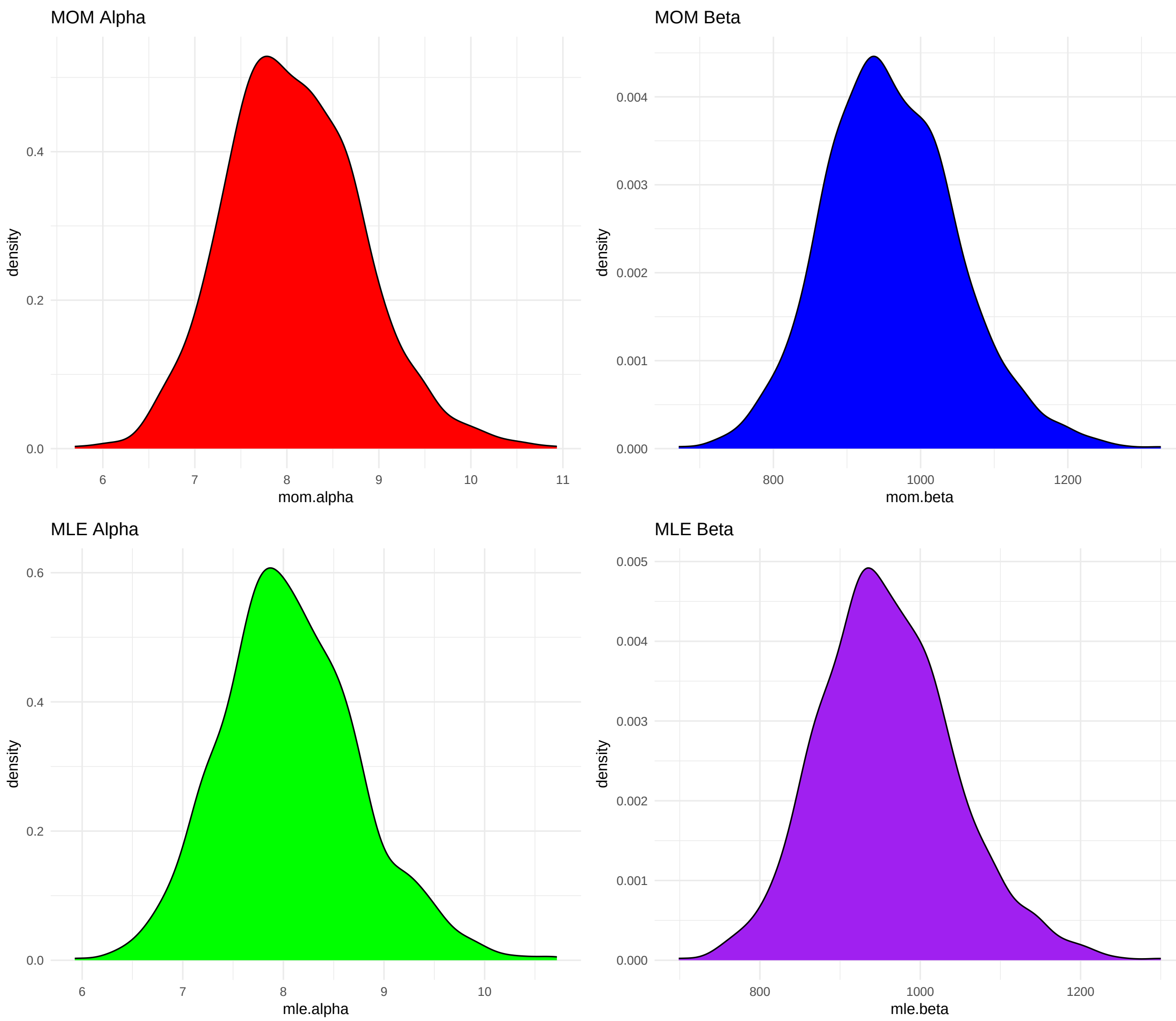

Figure 6: Distributions of Estimated α and β Values from MOM and MLE