

# Lab 7/8 – MATH 240 – Computational Statistics

Sankalp Ojha  
Colgate University  
Mathematics  
sojha@colgate.edu

04/03/2025

## Abstract

In this lab, we explored the Beta distribution. We explored the parameters and properties of the Beta distribution. We also explored the use of two point estimators, Method of Moments (MOM) and Maximum Likelihood Estimator (MLE), and how to check their accuracy. The point estimation methods were tested with real world data. We concluded that the MLE is better at finding the appropriate  $\alpha$  and  $\beta$  values compared to MOM.

**Keywords:** Beta distribution, MOM, MLE,  $\alpha$ , and  $\beta$

## 1 Introduction To Beta Distributions

The Beta distribution is a continuous distribution. The Beta distribution models the behavior of a random variable  $X$  within the support of 0 to 1. The Beta distribution is widely known for its flexibility, which makes it useful to model various datasets, probabilities, etc. Depending on the values of the parameters  $\alpha$  and  $\beta$ , which will be discussed furthered in section 2, the Beta distribution can take on left-skewed, right-skewed, U-shaped, or symmetric shapes, as seen in [Figure 1](#). Section 2 provides a greater description for the parameters for the Beta distribution. Section 3 discusses the properties and statistics of the Beta distribution. Sections 4 and 5 discusses point estimators and three metrics which we can use to check how accurate the estimators are, followed by a real world example of the use case for the estimators and checking metrics.

## 2 Density Functions and Parameters

A continuous distribution is one which has probability density function (PDF) and a cumulative density function (CDF) to represent its probability. The Beta distribution has two parameters,  $\alpha$  and  $\beta$ , which dictate on the distribution looks. Both  $\alpha$  and  $\beta$  are positive numbers and correspond to `shape1` and `shape2`, respectively, when write code for the Beta distribution in R. The Beta distribution PDF is given as:

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

In the beta distribution,  $I(x \in [0, 1]) = 1$  when  $x \in [0, 1]$  and 0 whenever  $x$  does not hold a number in support on the Beta distribution. The support of a random variable  $X$  in the Beta distribution is  $[0, 1]$ .

## 3 Properties and Key Statistics

The Beta distribution is a very versatile and flexible function when it comes to the shapes it can take on based on the various values of the  $\alpha$  and  $\beta$  parameters. The various forms of the Beta distribution can be shown through statistics such as mean, variance, skewness, and excess kurtosis. Something to note about the `kurt()` function in R is that it computes the excess kurtosis. So, if we are interested in finding kurtosis of a distribution, we can simply add three to the excess kurtosis found through the `kurt()` function. The equations which can be used to calculate the statistics for a Beta distribution are written in terms of  $\alpha$  and  $\beta$ , as seen below:

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad (\text{The Mean})$$

$$\text{var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (\text{The Variance})$$

$$\text{skew}(X) = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}} \quad (\text{The Skewness})$$

$$\text{kurt}(X) = \frac{6[(\alpha - \beta)^2(\alpha + \beta + 1) - \alpha\beta(\alpha + \beta + 2)]}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)}$$

(The Excess Kurtosis)

**Table 2** tables all the summary statistics for the various Beta distributions seen in [Figure 1](#) and illustrates how the  $\alpha$  and  $\beta$  dictate the shape of the Beta distribution.

Another way to compute these statistics is through their centered and uncentered  $k$ th moments. We can compute various combinations of centered and uncentered  $k$ th moments of the Beta distribution to calculate a certain statistic, as seen below:

$$\begin{aligned}
\mu_X &= E(X) && \text{(The Mean)} \\
\text{var}(X) &= \sigma_X^2 = E[(X - \mu_X)]^2 && \text{(The Variance)} \\
\text{skew}(X) &= \frac{E[(X - \mu_X)^3]}{E[(X - \mu_X)^2]^{3/2}} && \text{(The Skewness)} \\
\text{kurt}(X) &= \frac{E[(X - \mu_X)^4]}{E[(X - \mu_X)^2]^2} - 3 && \text{(The Excess Kurtosis)}
\end{aligned}$$

To generalize these equations to calculate any centered or uncentered  $k$ th moment of the Beta distribution, we can use the equations

$$\begin{aligned}
E(X^k) &= \int_{\chi} x^k f_x(x) dx \\
&\text{and} \\
E((X - \mu_X)^k) &= \int_{\chi} (x - \mu_X)^k f_x(x) dx
\end{aligned}$$

where the first equation corresponds to a centered  $k$ th moment and the second equation corresponds to a uncentered  $k$ th moment.

For the purposes of this lab, a function `beta.moment(alpha, beta, k, centered)` was made to calculate the various statistics of the Beta distribution for various values of  $\alpha$  and  $\beta$ .

### 3.1 Power of Larger Sample Sizes

By generating a large random sample from a fixed Beta distribution with parameters  $\alpha = 2$  and  $\beta = 5$ , we were able to prove that larger random sample sizes converge closer to the exact distribution. Using the `cumstats` package (Erdely and Castillo, 2017), I made [Figure 2](#), which illustrates how the Weak Law Of Large Numbers works. The Weak Law Of Large Numbers suggests that as the sample size increases so will the precision increase towards the total population distribution.

## 4 Estimators And Precision

When dealing with data in real world applications, we don't necessarily know what the values of the parameters  $\alpha$  and  $\beta$  will be for the given data set. This is where we use point estimators to calculate the precise values of  $\alpha$  and  $\beta$  which will best fit the data set. The two methods to estimate include: Method of Moments (MOM) and Maximum Likelihood Estimation (MLE). The MOM approach equates the first sample  $k$ th moments to the first population  $k$ th moments. The goal is to find which values of  $\alpha$  and  $\beta$  will match the data set moments best with the population moments. The MLE approach utilizes a likelihood function which aims to find the values of  $\alpha$  and  $\beta$  which will maximize the likelihood of matching the shape given data set. In R, to compute the values of  $\alpha$  and  $\beta$  using MOM, we first make a function which returns the first population  $k$ th moments minus the first sample  $k$ th

moments. Then we used the `nleqslv` package (Hasselmann, 2023) and the function `nleqslv()` to solve for the values of  $\alpha$  and  $\beta$  which make first population  $k$ th moments minus the first sample  $k$ th moments equal to zero. The MLE is implemented through a function which is passed into `optim()`. The `optim()` function maximizes the negative log-likelihood function. As both of these methods are estimates, we would like to know how precise the estimates are, which we can do through calculating Bias, Precision, and Mean Squared Error (MSE). Bias calculates the average difference from the true value, precision calculates the consistency of the estimates, and MSE combines both prior metrics to calculate the error. The equations to calculate each measure are:

$$\begin{aligned}
\text{Bias} &= \mathbb{E}(\hat{\theta}) - \theta \\
\text{Precision} &= \frac{1}{\text{Var}(\hat{\theta})} \\
\text{MSE} &= \text{Var}(\hat{\theta}) + (\mathbb{E}(\hat{\theta}) - \theta)^2
\end{aligned}$$

where  $\hat{\theta}$  is a given vector of estimates and  $\theta$  is the true value.

## 5 Example: World Death Rates

The study, Fatih (2022), suggests the worldwide death rates can be modeled by a Beta distribution with the specific parameters,  $\alpha = 8$  and  $\beta = 950$ . Using the MOM and MLE estimator methods, we calculated the  $\alpha$  and  $\beta$  which would best fit the dataset for 1000 random samples of size  $n = 266$  for world death rates. To check which estimator is more accurate, we can use Bias, Precision, and MSE. Below is a table information about the accuracy of each estimator:

parameters	method	bias	precision	mse
Alpha	MOM	0.08	1.83	0.55
Alpha	MLE	0.07	2.13	0.48
Beta	MOM	10.29	0.00	8288.46
Beta	MLE	9.11	0.00	7132.70

Table 1: MOM and MLE bias, precision, and MSE for  $\alpha$  and  $\beta$ .

As can be seen in the table, the MLE estimator is the more precise estimator. In the Bias, Precision, and MSE categories, MLE had lower numbers than MOM. This indicates that MLE is less bias, more precise, and overall a more accurate point estimator tool. It should be noted that the precision calculations for MOM and MLE  $\beta$  were extremely small and are represented as zero as a result. While both methods are not fully accurate, MLE proves to be the superior method.

## References

- Erdely, 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.  
Hasselmann, B. (2023). *nleqslv: Solve Systems of Nonlinear Equations*. R package version 3.3.5.

## 6 Appendix

Table 2: Population Summary Of 4 Beta Distribution Parameters with varying  $\alpha$  and  $\beta$  values.

	Alpha	Beta	Mean	Variance	Skewness	Excess Kurtosis
1	2.00	5.00	0.29	0.03	0.60	-0.12
2	5.00	5.00	0.50	0.02	0.00	-0.46
3	5.00	2.00	0.71	0.03	-0.60	-0.12
4	0.50	0.50	0.50	0.12	0.00	-1.50

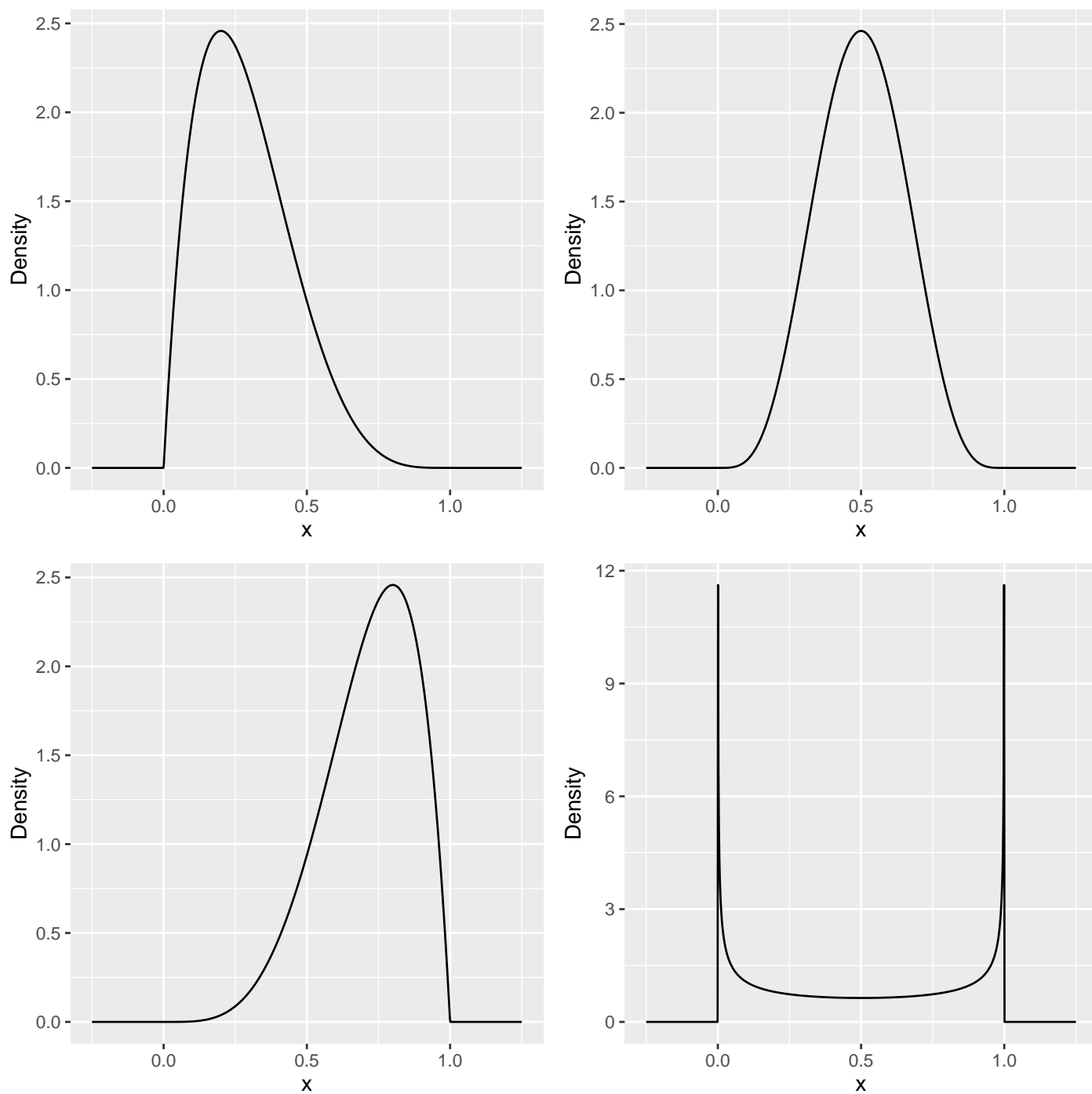


Figure 1: Plotting various beta PDFs with varying  $\alpha$  and  $\beta$  values.

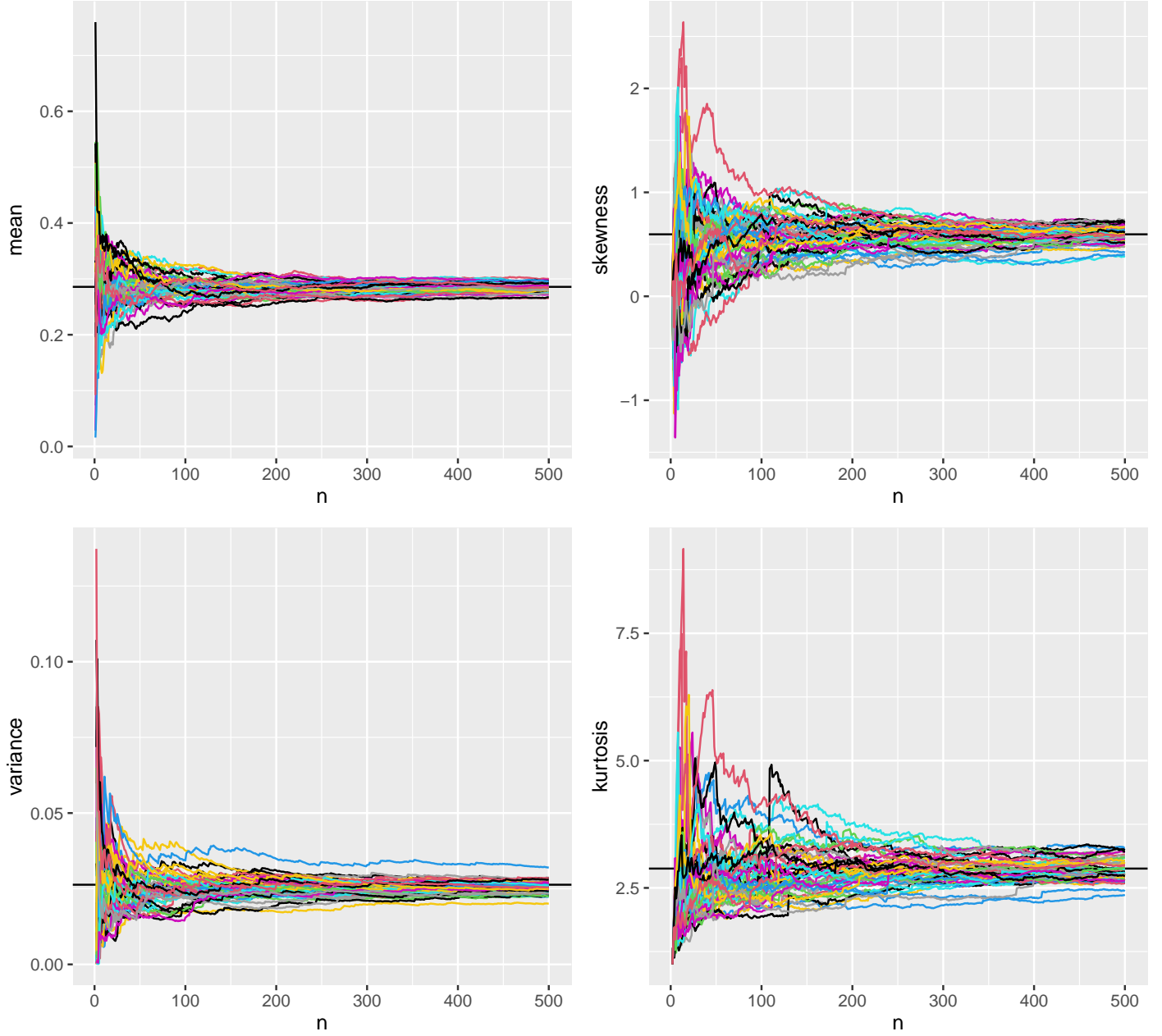


Figure 2:  $\alpha=2$  and  $\beta=5$  convergence simulation to show the weak law of large numbers.

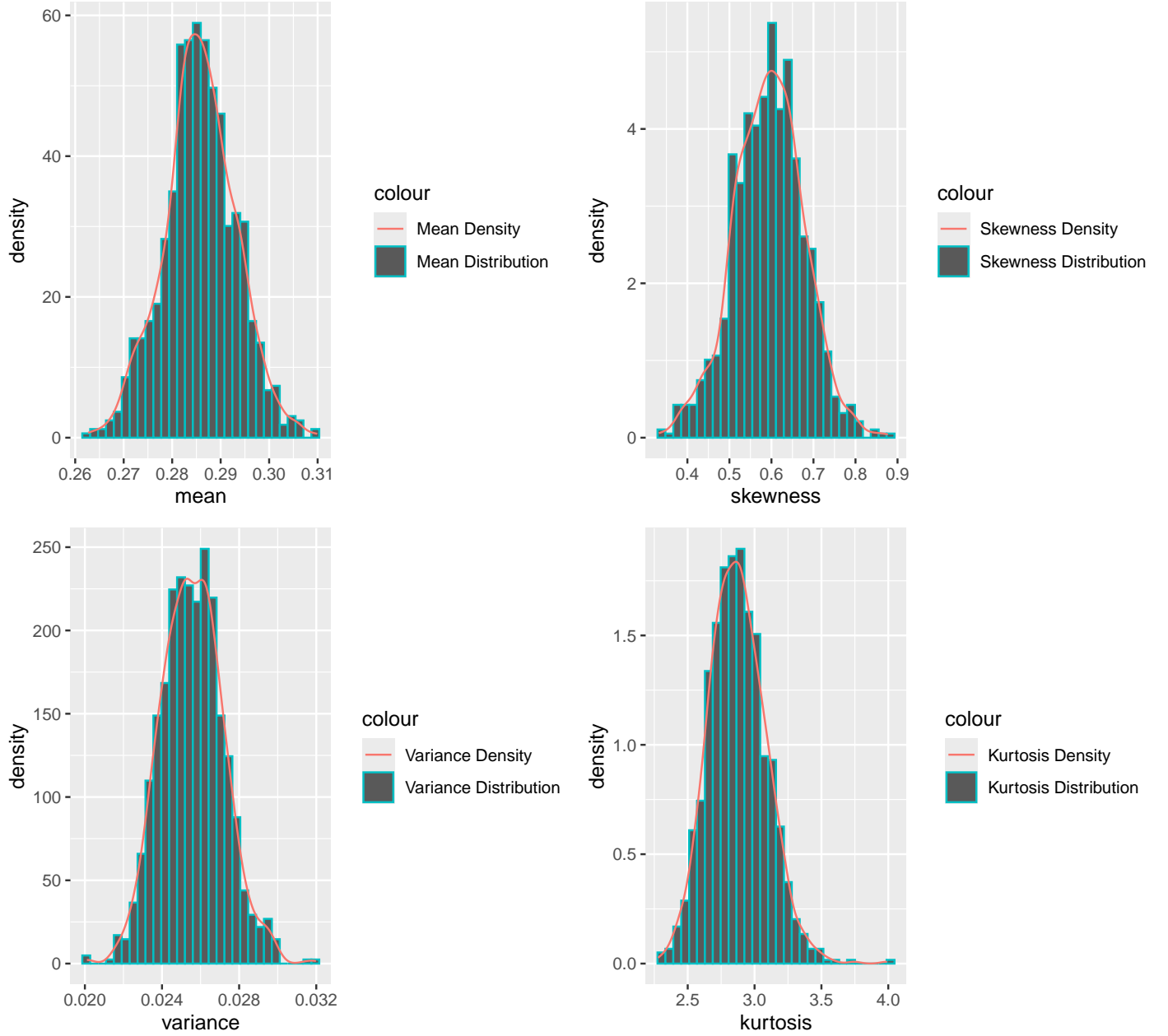


Figure 3: Beta( $\alpha=2$  and  $\beta=5$ ) sampling distributions to illustrate various summary statistics.

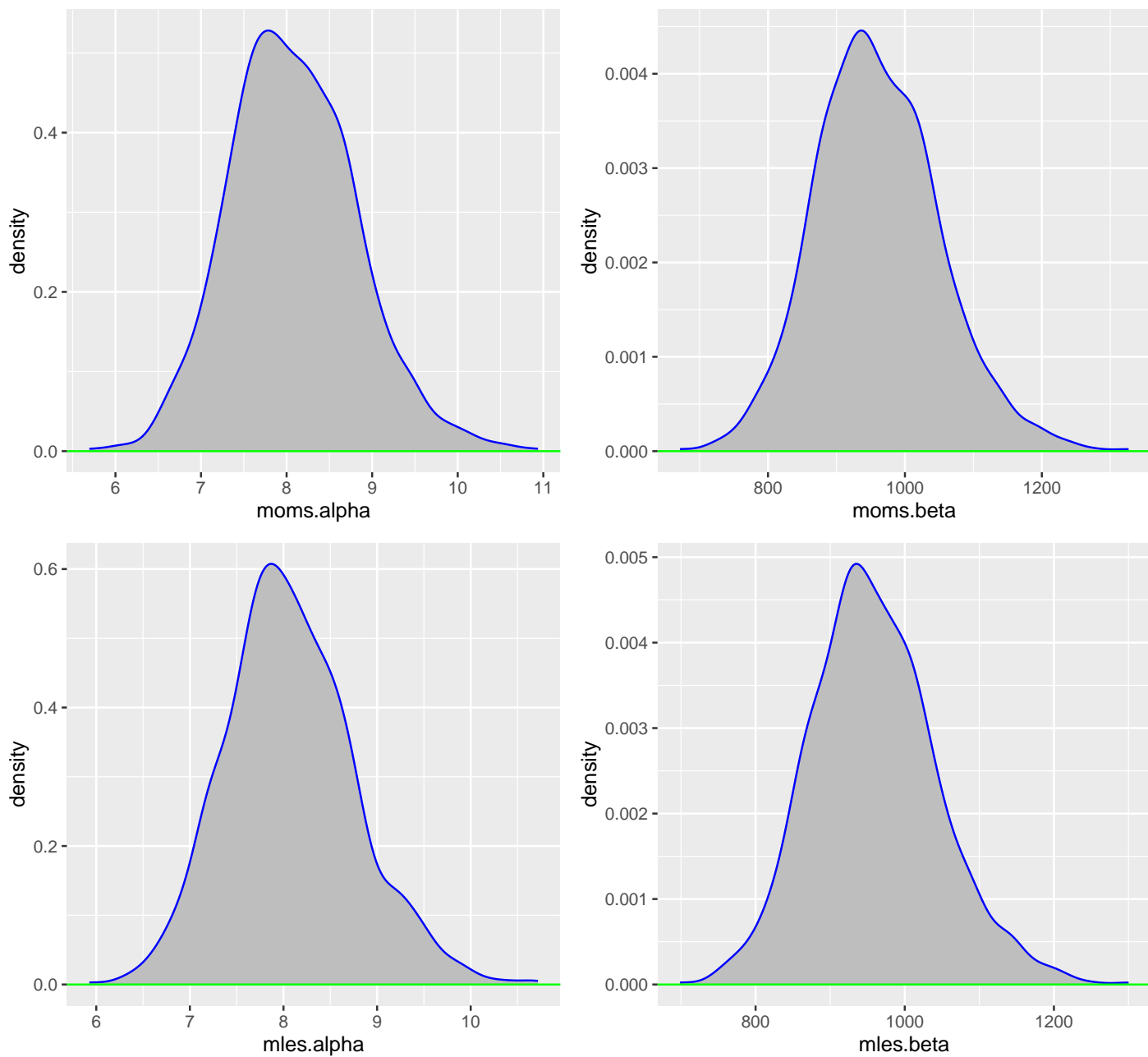


Figure 4: Resampled  $\text{beta}(\alpha=8 \text{ and } \beta=950)$  sampling distributions to illustrate the shape parameters behavior.