

# Lab 8 – MATH 240 – Computational Statistics

Henry Sun  
Colgate University  
Department of Mathematics

04/01/2025

## Abstract

In this lab, we explored the beta distribution in great detail. We discussed probability functions, parameters, and properties of the beta distribution. We then use two point techniques, the method of moments (MOM) and the maximum likelihood (MLE) to help us model country death rates. Overall, we found that the estimated parameters from MLE were more accurate than MOM compared to actual parameters suggested.

**Keywords:** Point estimation; continuous probability distributions; parameters

## 1 Introduction

The beta distribution is a continuous distribution used to model a random variable  $X$  ranging from 0 to 1 (the distribution's support), making it particularly useful for modeling proportions, probabilities, and rates. For some data, like the example we will look at later in the lab, data must be transformed to fit the support. The beta distribution is remarkably flexible in its shape; it can be right-skewed, left-skewed, or symmetric.

Section 2 provides a more in-depth overview of the beta distribution, including the Probability Density Function, support, and parameters. Section 3 discusses the properties of the beta distribution using four specific cases. Sections 4 and 5 contextualize point estimators (Method of Moments and Maximum Likelihood) and explores the beta distribution further with an example of real-life data.

## 2 Density Functions and Parameters

As the beta distribution is a continuous distribution, it has a probability density function (PDF) and a cumulative density function (CDF). It takes two shape parameters,  $\alpha$  and  $\beta$ , which are both positive numbers. In R, the parameters for  $\alpha$  and  $\beta$  are `shape1` and `shape2`. The beta distribution's probability density function is given as:

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

For the beta distribution,  $I(x \in [0, 1]) = 1$  when  $x \in [0, 1]$  and 0 otherwise, as the support of a random variable  $X$  is  $[0, 1]$  in the beta distribution.

## 3 Properties and Statistics

Various properties of the beta distribution are based on the shape parameters of the beta distribution. The beta distribution is very flexible with regards to its shape. The various statistics, such as the mean, variance, skewness, and excess kurtosis can be written in terms of  $\alpha$  and  $\beta$ :

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

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

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

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

It's key to note that in R, `kurt()` calculates the excess kurtosis and not the kurtosis. Instead,  $\text{kurtosis} = \text{excess kurtosis} + 3$ , or `kurt(X) + 3`.

However, by computing the moments of the beta function, we write each statistic as a combination of various  $k$ th centered and uncentered moments.

$$\mu_X = E(X) \quad (\text{The Mean})$$

$$\text{var}(X) = \sigma_X^2 = E[(X - \mu_X)]^2 \quad (\text{The Variance})$$

$$\text{skew}(X) = \frac{E[(X - \mu_X)^3]}{E[(X - \mu_X)^2]^{3/2}} \quad (\text{The Skewness})$$

$$\text{kurt}(X) = \frac{E[(X - \mu_X)^4]}{E[(X - \mu_X)^2]^2} - 3 \quad (\text{The Excess Kurtosis})$$

For the beta distribution, the  $k$ th uncentered and centered moments are

$$E(X^k) = \int_{\mathcal{X}} x^k f_X(x) dx$$

and

$$E((X - \mu_X)^k) = \int_{\mathcal{X}} (x - \mu_X)^k f_X(x) dx$$

respectively. For the purposes of this lab, we wrote a function `beta.moment(alpha, beta, k, centered)` to help us compute the population-level characteristics using this moments.

Due to the beta distribution’s flexibility, it can take on numerous different shapes. Below is a table (Table 1) showing various different summary statistics for four different cases of the beta distribution. Additionally, the actual shape of these different cases can be found in the appendix (Figure 1).

alpha	beta	mean	variance	skewness	e.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: Table showing various summary statistics for different shape parameters,  $\alpha$  and  $\beta$ . Notice the large variation in the summary statistics for different shape parameters. **e.kurtosis** means excess kurtosis.

As with many other distributions, as the sample size (samples were generated from each known beta distribution) increases, summary statistics calculated on each sample converges towards the true population-level characteristics. A plot of samples (Figure 2) generated from sizes 0 to 500 from the beta( $\alpha = 2$ ,  $\beta = 5$ ) distribution show this property. The cumulative statistics (Erdely and Castillo, 2017) of samples were chaotic in smaller sample sizes, but stabilized at sample size was increased.

The sampling distributions for summary statistics from the original beta distribution also show an approximately Gaussian distribution (Figure 3), with the mean of the sampling distributions being approximately the population-level summary statistics.

## 4 Estimators

In class, we discussed two point estimators that help us find the approximate estimates for unknown parameters,  $\theta$  in various distributions, the Method of Moments (MOM) and Maximum Likelihood (MLE). In the beta distribution,  $\theta$  are the shape parameters  $\alpha$  and  $\beta$ .

In many cases, the distribution a set of data follows is unknown, and point estimators can help us come up with a good approximation of the parameters for a distribution that fits the data.

The Method of Moments works by equating the first  $k$  uncentered population moments with the first  $k$  sample moments (assuming that the sample size is large enough where they are approximately equal), setting up an approximate systems of equation to solve for each parameter. In R, we accomplished this by using the `nleqslv()` (Hasselmann, 2023) to obtain  $\alpha$  and  $\beta$ . For the beta distribution, the first two uncentered population moments are:

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

$$E(X^2) = \frac{(\alpha + 1)\alpha}{(\alpha + \beta + 1)(\alpha + \beta)}$$

Maximum Likelihood works by first defining the likelihood function:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f_X(x_i|\theta)$$

The likelihood function is asking the question: What is the likelihood of observing  $\mathbf{x}$ ? However, in most situations, it is

more practical and useful to work with the log likelihood function. For Maximum Likelihood, we are essentially trying to find which parameters maximize the probability of observing  $\mathbf{x}$ . This basically becomes an optimization problem that can be solved with calculus. In R, this is accomplished by using the `optim()` function.

We will apply both MOM and MLE in the Example section, using actual data.

## 5 Example: World Death Rates

Fatih (2022) suggests that country death rates worldwide can be modeled with a beta distribution, specifically with  $\alpha = 8$  and  $\beta = 950$ . Both MOM and MLE will provide us with slightly different parameters, so we must analyze bias, precision, and mean squared error (MSE) in order to determine which one gives us a closer estimate to the actual parameters. The data also had to be modified, changing the death rate so it would fit the support of the beta distribution.

$$bias = E(\hat{\theta}) - \theta$$

$$precision = \frac{1}{var(\hat{\theta})}$$

$$MSE = var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2$$

The summaries of these calculations for each method and shape parameter can be found in the table below (Table 2). Overall, it appears that MLE seems to be the better estimator for this data than MOM. The bias and MSE for the MLE parameters is smaller than MOM, and the precision is higher. However, it’s still quite hard to notice the difference between the actual PDF, MLE PDF, and MOM PDF, as shown in Figure 5. Still, the MLE PDF seems to be marginally better fit for the data.

parameters	method	bias	precision	mse
Alpha	MOM	0.08	1.83	0.63
Alpha	MLE	0.07	2.13	0.54
Beta	MOM	10.29	0.00	8192.93
Beta	MLE	9.11	0.00	7058.82

Table 2: The bias, precision, and MSE for  $\alpha$  and  $\beta$  with MOM and MLE. Due to the exceedingly small values for  $\beta$  precision, it appears as 0 on the table.

Just like how the sampling distribution for the summary statistics is approximately Gaussian, the sampling distribution of the shape parameters  $\alpha$  and  $\beta$ , follow an approximately Gaussian distribution (Figure 4), with the actual parameter (shown as a red vertical line), regardless if we are using MOM and MLE. Samples were generated from the actual parameters for the beta distribution fitting the death rates,  $\alpha = 8$  and  $\beta = 950$ . Both estimations are not exact, but can be considered close enough.

## 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.  
 Hasselman, B. (2023). *nleqslv: Solve Systems of Nonlinear Equations*. R package version 3.3.5.

## 6 Appendix

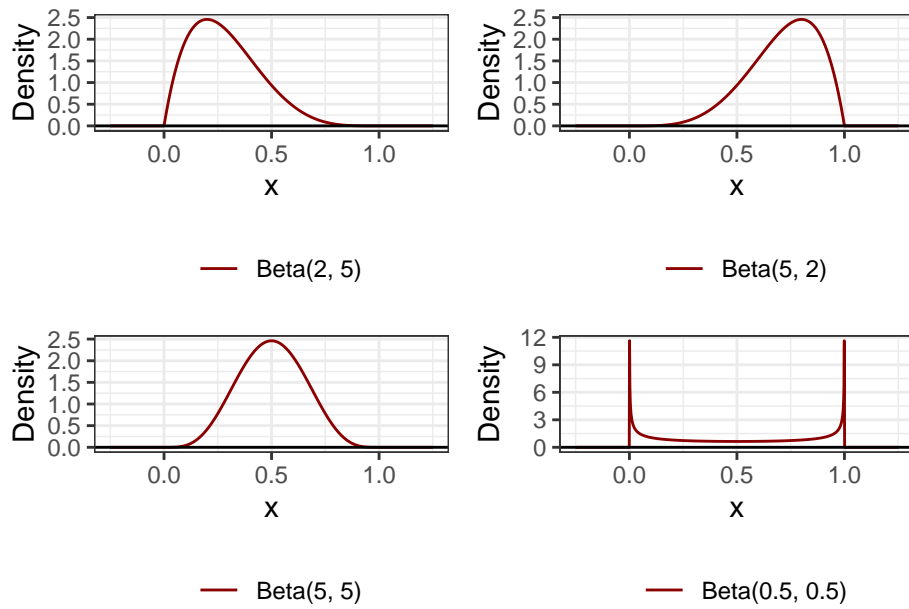


Figure 1: Plotting of various different beta PDFs.

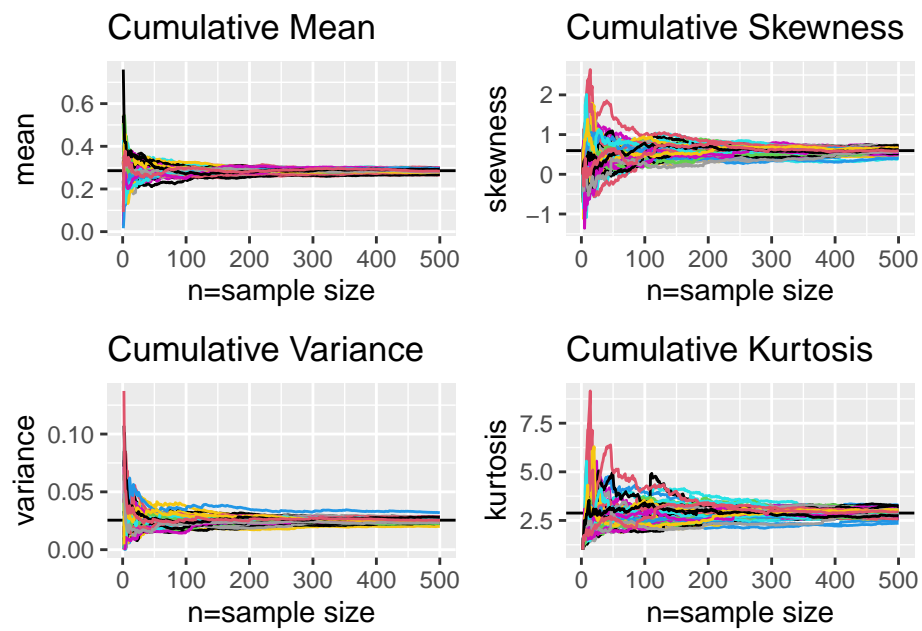


Figure 2: Plots showing the cumulative statistics, showing how statistics change as sample size increases from 0 to 500.

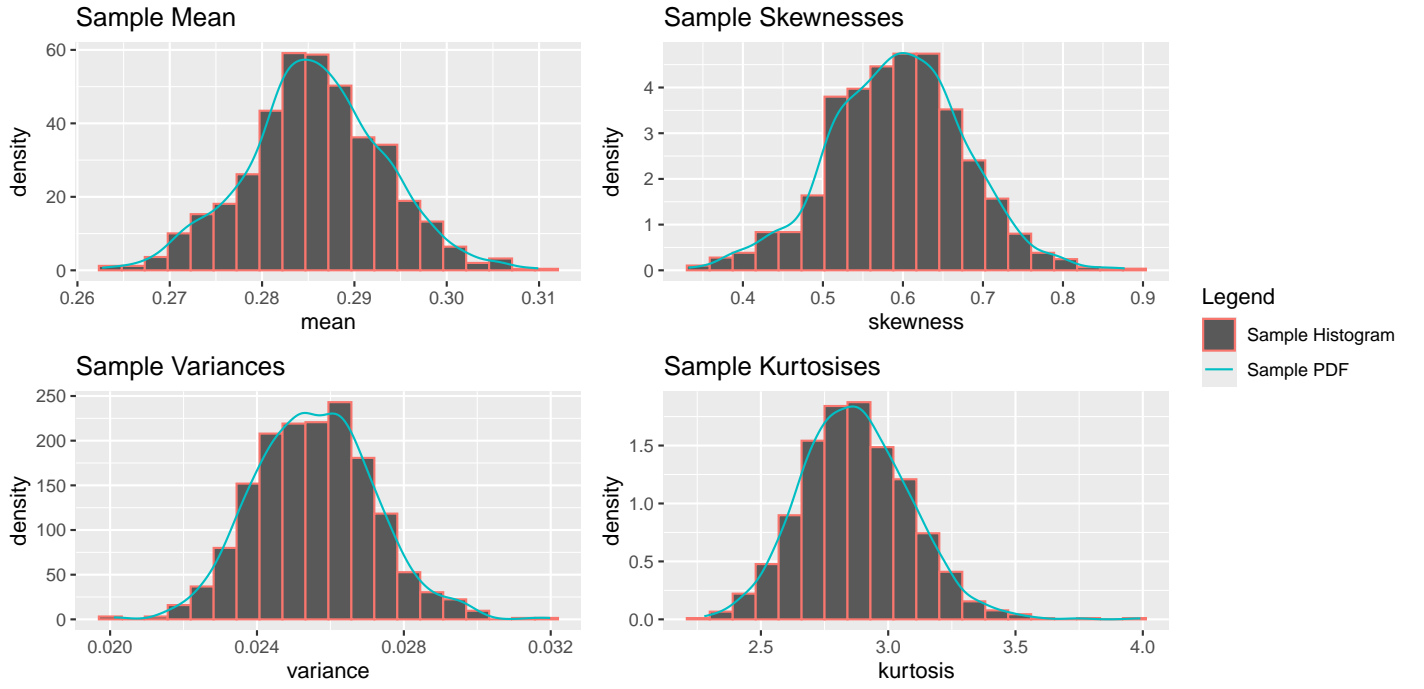


Figure 3: Sampling Distributions for various summary statistics. The samples were generated from the  $\text{beta}(\alpha = 2, \beta = 5)$  distribution.

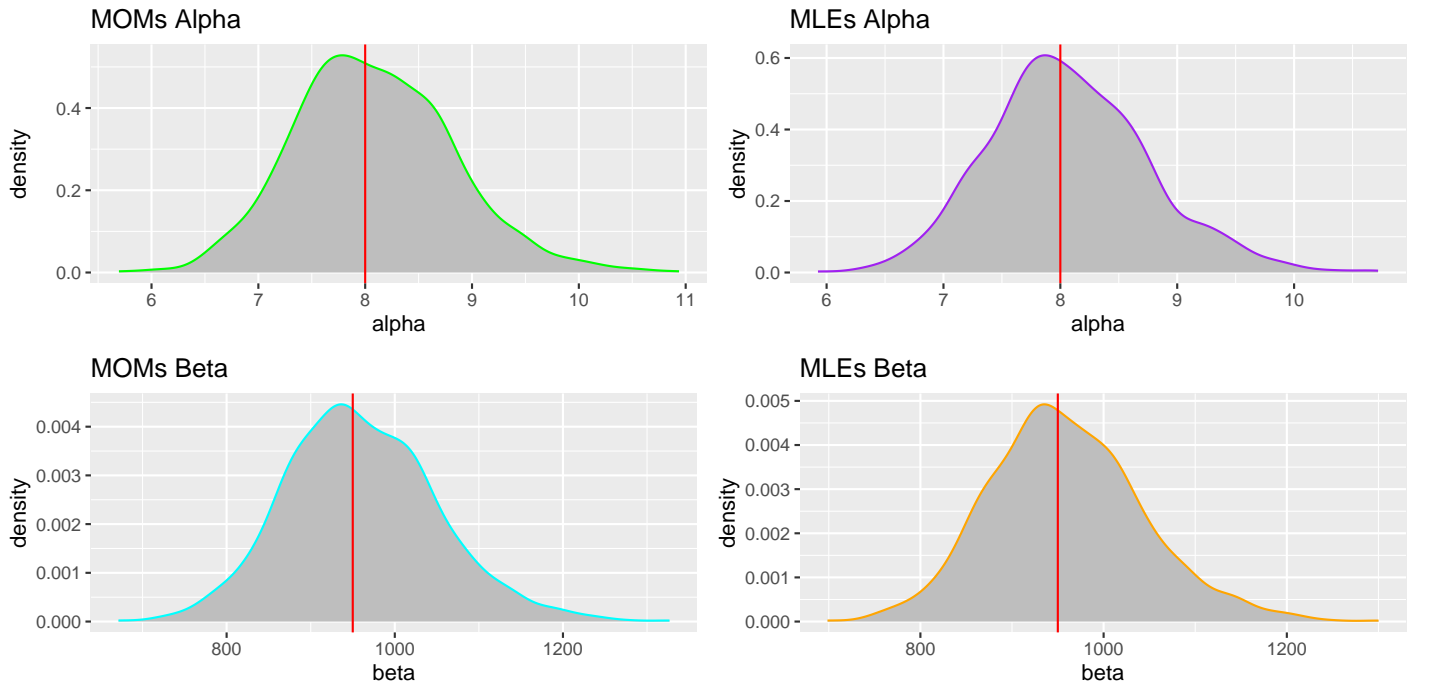


Figure 4: Sampling Distributions of the shape parameters  $\alpha$  and  $\beta$ , with resamples generated from the  $\text{beta}(\alpha = 8, \beta = 950)$  distribution.

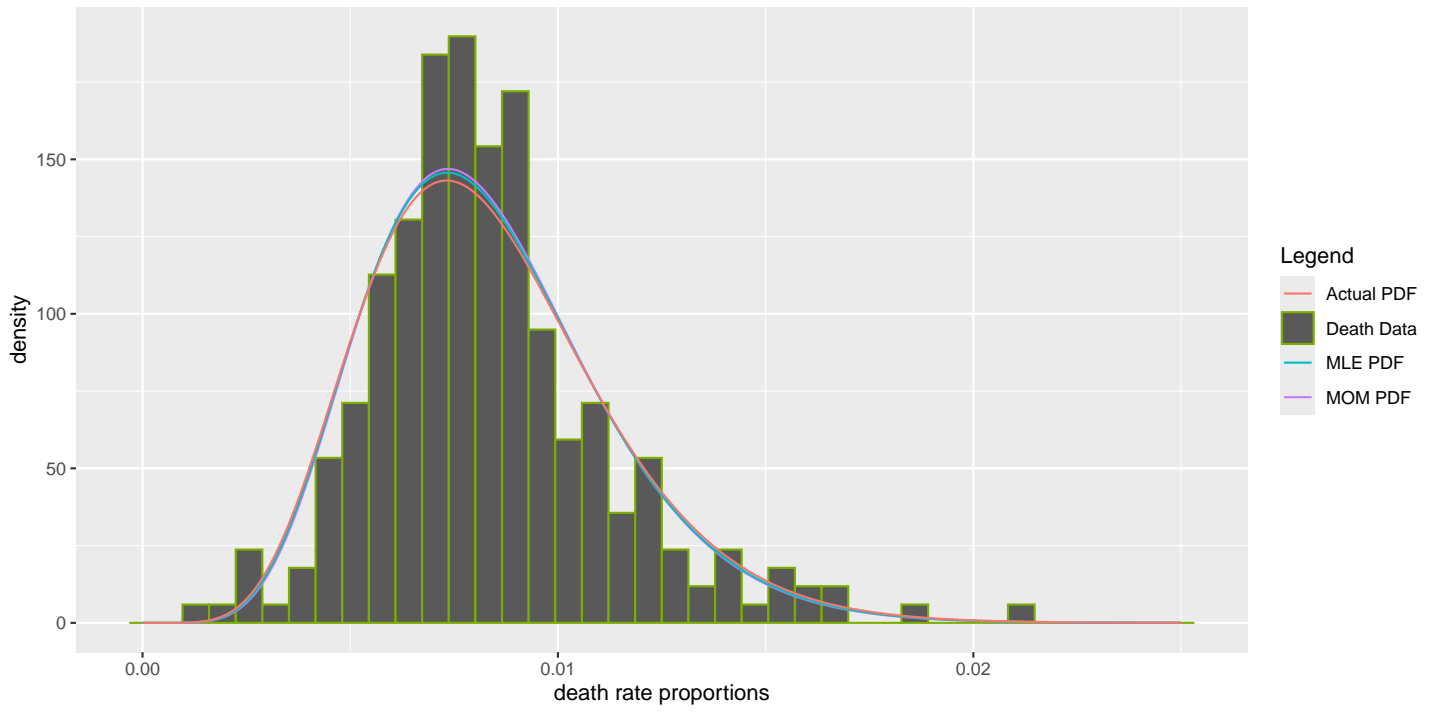


Figure 5: Death Rates data plotted alongside the actual PDF proposed, and the PDFs from the estimated shape parameters obtained from MOM and MLE. The vertical shows the actual proposed shape parameters.