

Lab 7-8 – MATH 240 – Computational Statistics

Yuliia Heleveria
MATH 240 Lab A
Mathematics
yheleveria@colgate.edu

04/01/2025

Abstract

We explore the beta distribution, a flexible probability distribution, which is frequently used to model proportions and rates. We analyze beta's probability density function (PDF) and how beta's parameters α and β can influence its PDF's shape. Using various beta PDFs, we examine key properties of the beta distribution, such as its mean, variance, skewness, and excess kurtosis. We discuss parameter estimation for beta distribution using method of moments and maximum likelihood estimation. We then demonstrate a real-world application of beta distribution by showing its ability to model countries' death rates.

Keywords: beta distribution; probability density function; parameter estimators.

1 Introduction

The beta distribution is a continuous probability distribution that is defined on the interval $[0,1]$. It is used to model a random variable X . The beta distribution is used to model proportions, rates, and probabilities. The distribution is governed by two shape parameters: $\alpha > 0$ and $\beta > 0$, which determine the shape for different data distributions. The beta distribution is very flexible with its shape: it can be left-skewed, right-skewed, or symmetric.

Section 2 defines the probability density function (PDF) and its parameters for beta distribution. Section 3 covers statistical characteristics of the beta distribution such as mean, variance, skewness, and excess kurtosis. Section 4 discusses the methods for estimating α and β from sample data using the method of moments (MOM) and maximum likelihood estimation (MLE). Section 5 presents the example from the death rates data to demonstrate real world application of beta distribution.

2 Density Function and Parameters

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

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

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

The shape parameters α and β control the skewness and concentration of the distribution.

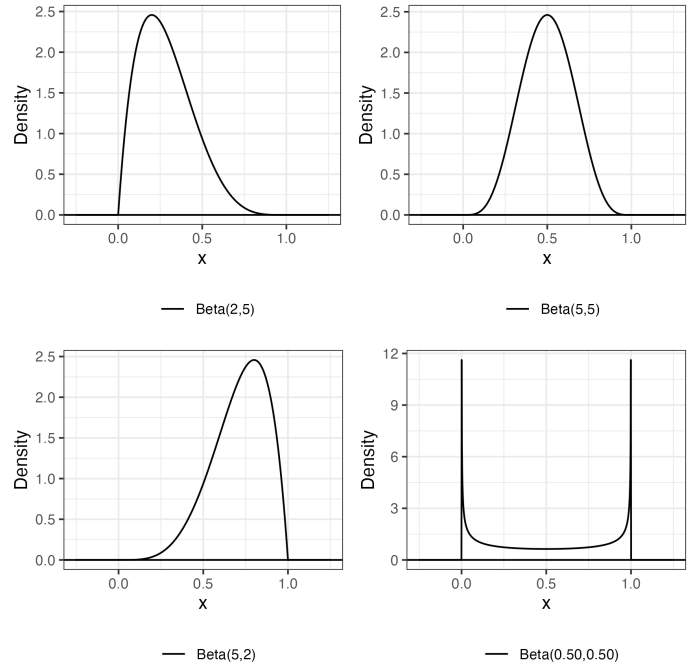


Figure 1: Beta distribution PDFs for different parameter values of α and β

The four plots in Figure (1) illustrate beta probability density function with various combinations of parameters α and β . We used `ggplot2` and `patchwork` packages in R to create and combine the graphs (Wickham, 2016; Pedersen, 2024). The top-left plot demonstrates a distribution skewed to the right (where $\beta > \alpha$), so there is a higher chance of observing values in the lower range of the interval $[0,1]$. The top-right plot demonstrates a symmetric distribution centered around 0.5 ($\alpha = \beta$). Here, the values in the middle of the interval are most likely to occur. The bottom-left plot demonstrates a left-skewed distribution (where $\alpha > \beta$). The values closer to the end of the interval are most likely to occur. The bottom-right plot illustrates a U-shaped distribution (both $\alpha, \beta < 1$). For these parameters, the density is highest at the endpoints of the interval and lowest around 0.5.

We can approximate the population distribution by using a sample and comparing its density to the true population's density. Figure (5) presents four samples and their estimates

for the beta population by comparison of density curves. By comparing the shape of the sample's histogram and its density, we can conclude that the sample estimates the population well. The sample density curve is very similar to the population's density curve.

3 Properties

The beta distribution has the following key properties:

- Mean:

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

- Variance:

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

- Skewness:

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

- Excess Kurtosis:

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

Based on Figure (1), we can see the values of the mean, variance, skewness, and excess kurtosis in Table (1). The mean of the distribution reflects the average value of the distribution. A low mean (such as for Beta(2, 5)) indicates a right-skewed distribution and a high mean (such as for Beta(5, 2)) indicates a left-skewed distribution. The mean when parameters $\alpha = \beta$ indicates a symmetrical distribution (Beta(5,5) and Beta(0.5,0.5)).

The variance indicates the spread of the distribution around the mean. Beta(5,5) has the lowest variance and Beta(0.5,0.5) has the highest variance.

The skewness describes the asymmetry of the distribution. Positive skewness indicates a right-skewed distribution (Beta(2,5)) and negative skewness indicates a left-skewed distribution (Beta(5,2)). If skewness is zero, the distribution is symmetric (Beta(0.5,0.5) and Beta(5,5)).

The excess kurtosis measures how peaked the distribution is compared to a normal distribution. If the excess kurtosis is equal to 0, the distribution is as peaked as the normal distribution is (mesokurtic). If the excess kurtosis is positive, the distribution has a sharp peak and flat tails (leptokurtic). If the excess kurtosis is negative, the distribution has a plat peak and thin tails (platykurtic). All four beta distributions analyzed are platykurtic because their excess kurtosis is less than 0.

Figure (2) illustrates how sample's statistics (mean, variance, skewness, and kurtosis) converge to the population values as the sample size increases. Each color line represents a different simulation, showing how the cumulative statistics evolve with increasing sample size for all of the trials. The cumulative statistics were calculated using `cumstats` package in R (Erdely and Castillo, 2017). The horizontal black line in each plot represents the true statistics value for the

population. The sample statistics for each trial converge to the population value as the sample size increases. At smaller sample sizes, there is a large fluctuation in values, but as the sample size increased, the fluctuation decreases.

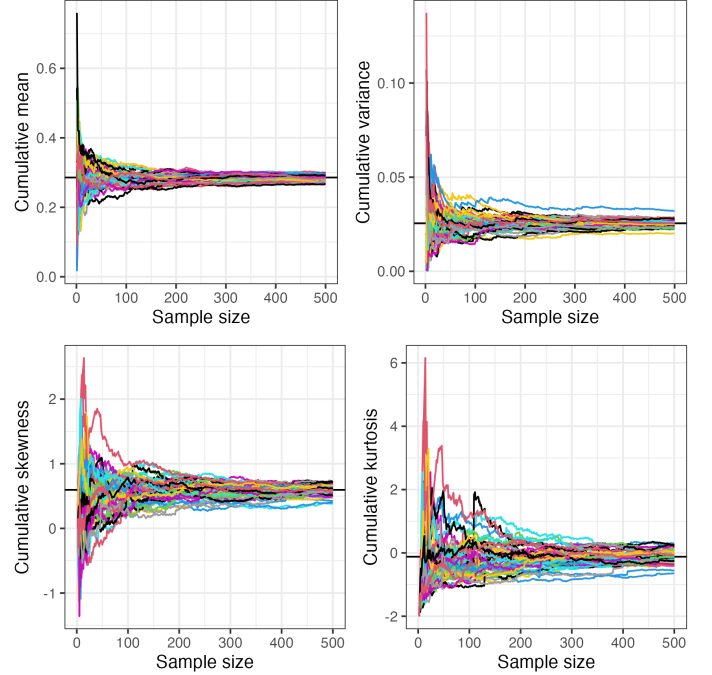


Figure 2: Histograms of cumulative statistics as the sample size increases

4 Estimators

Given a sample drawn from beta distribution, we can estimate parameters α and β using the method of moments (MOM) and maximum likelihood estimation (MLE).

To estimate α and β using MOM, we need to equate the first two sample moments ($E(X), E(X^2)$) to the population moments.

The mean (first moment) of the beta distribution is:

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

The variance (related to the second moment) of the beta distribution is:

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

To estimate α and β using by computing maximum likelihood estimation, we need to maximize the likelihood function based on the observed data. To simplify the optimization process, we can maximize the log-likelihood function which produces the same result as maximizing the likelihood function.

The log-likelihood function:

$$L(\alpha, \beta | X) = \prod_{i=1}^n f_X(x_i | \alpha, \beta)$$

In Section 5 we examine death rates data using beta distribution. The estimates for shape parameters α and β based on the death data from the World Bank were computed using

both MOM and MLE and we can see estimates' densities in Figure (3).

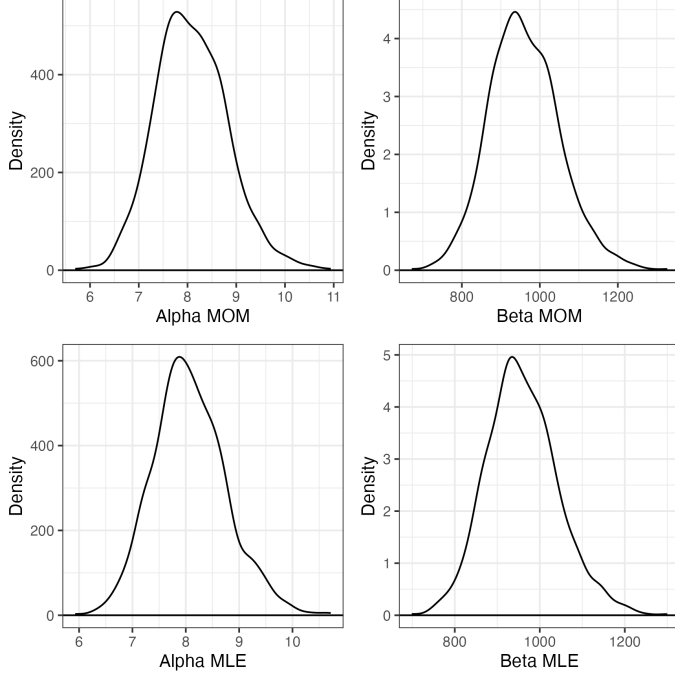


Figure 3: Density of estimates for α and β using MOM and MLE

We can see that the distributions of the estimated parameters are similar in shape. However, MLE has significantly lower variability compared to MOM estimates.

We also computed bias, precision, and mean squared error for estimates in Figure (3). The numerical estimates are provided in Table (2). It is evident that MLE provides lower bias, higher precision, and lower mean squared error for both α and β estimates.

In our example, MLE has a smaller variance, higher precision, and lower bias compared to MOM, so MLE should be the preferred estimator.

5 Example

As an example of practical application of the beta distribution, consider country-level death rates. We use the data set from the World Bank and we convert the raw data from the number of death per 1000 citizens to a rate to align with the beta distribution's requirements of values within the interval $X \in [0, 1]$. We used `tidyverse` package in R (Wickham et al., 2019) to convert the raw data into the rate to fit beta distribution.

Figure (4) presents the histogram of the transformed death rate data with fitted beta distribution obtained using both method of moments and maximum likelihood estimates.

The histogram presents the distribution that is slightly right-skewed. Most countries' death rates are concentrated around 0.007. Both superimposed MOM and MLE provide reasonable fits for the data as they closely follow the histogram's shape. However, the peak is slightly better captured by MOM estimation.

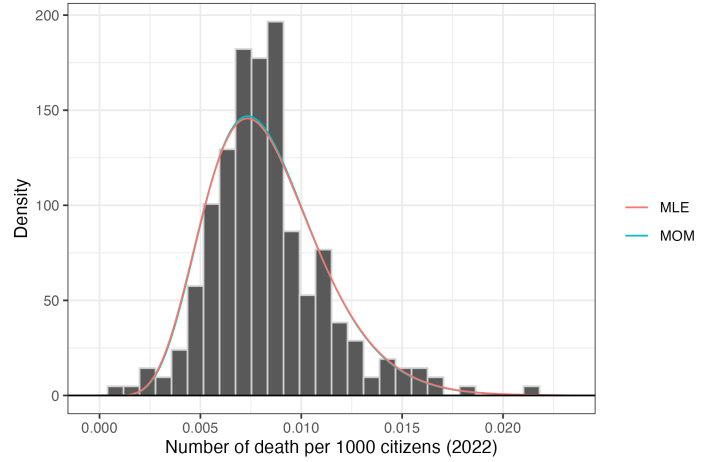


Figure 4: Histogram of death rates with MOM and MLE distributions superimposed

Beta distribution is applicable to real-world data, such as death rated in a population. Fitted distribution obtained from MLE and MOM helps us to analyze and understand the patterns of mortality.

6 Discussion

The beta distribution is a flexible probability distribution commonly used for rate and proportion modeling. The shape parameters α and β control skewness and spread of the distribution, allowing it to fit a wide range of data patterns.

We used both MOM and MLE to estimate the parameters and fit the distribution for the given data. Our results demonstrate that MLE is often less biased and more precise method for parameter estimation, especially when working with large datasets.

The beta distribution has a wide real-world application, such as for modeling death rates. Its ability to represent probabilities and proportions makes it especially useful in fields where uncertainty and prior knowledge play a critical role.

References

- Erdelyi, 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. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemond, 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.

7 Appendix

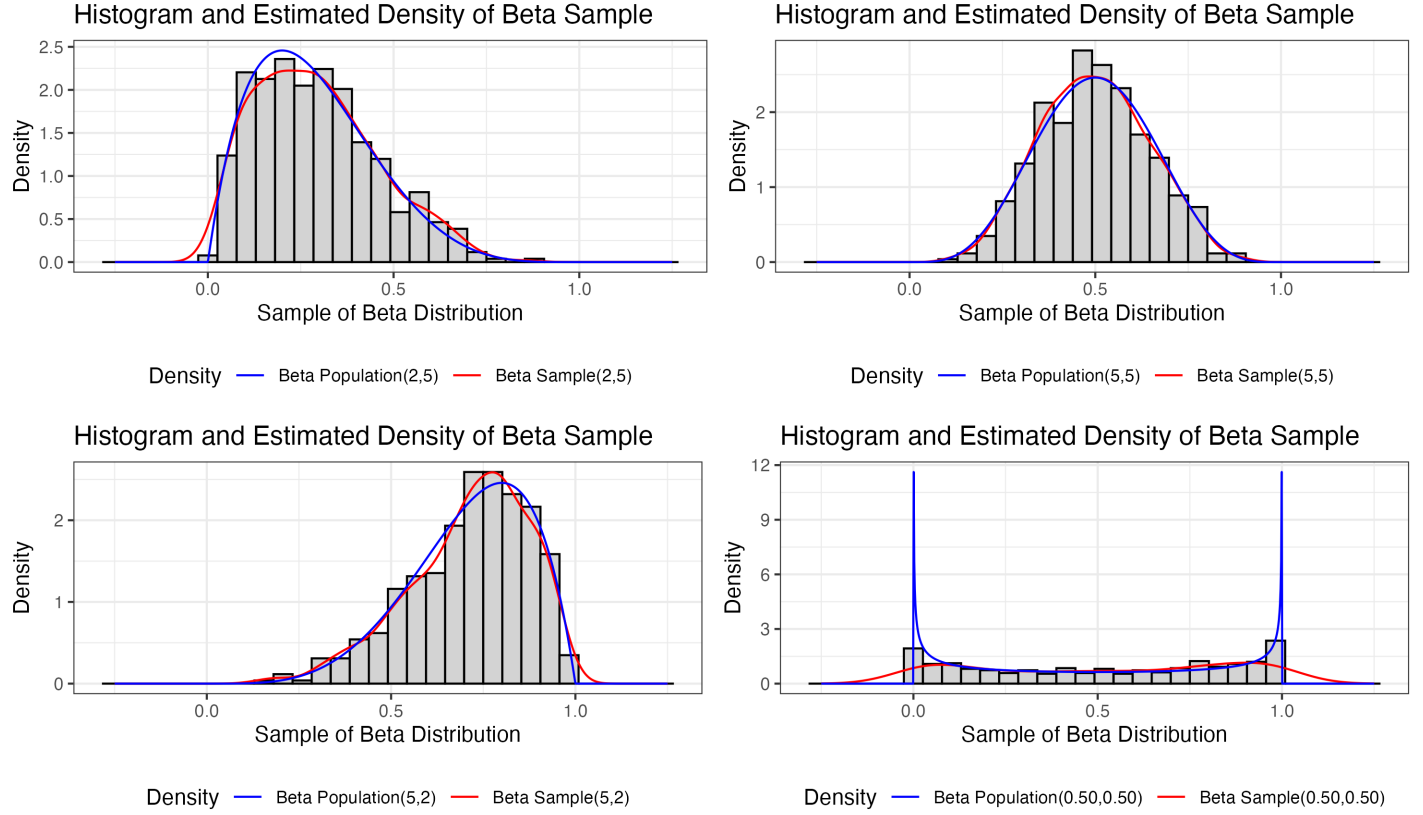


Figure 5: Histogram of estimated densities of beta sample and beta population

Distribution	Mean	Variance	Skewness	Excess.Kurtosis
Beta(2,5)	0.29	0.03	0.60	-0.12
Beta(5,5)	0.50	0.02	0.00	-0.46
Beta(5,2)	0.71	0.03	-0.60	-0.12
Beta(0.5,0.5)	0.50	0.12	0.00	-1.50

Table 1: Statistical properties for various parameteres of beta distribution

Estimate	Bias	Precision	Mean.Squared.Error
Alpha MOM	0.08	1.83	0.55
Beta MOM	10.57	0.00	8303.00
Alpha MLE	0.07	2.13	0.47
Beta MLE	9.18	0.00	7118.73

Table 2: Bias, precision, and mean squared error for estimates