

Lab 7-8 – MATH 240 – Computational Statistics

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

04/01/2025

Abstract

This document provides a basic template for the 2-page labs we will complete each week. Here, briefly summarize what you did and why it might be helpful. Provide all the top-line conclusions, but avoid providing *all* the details. Results should be limited to “we show X, Y, and Z.”

Keywords: What topics does the lab cover concerning class? List 3-4 key terms here, separated by semicolons.

1 Introduction

The Beta distribution is a continuous 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 parameter estimation.

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.

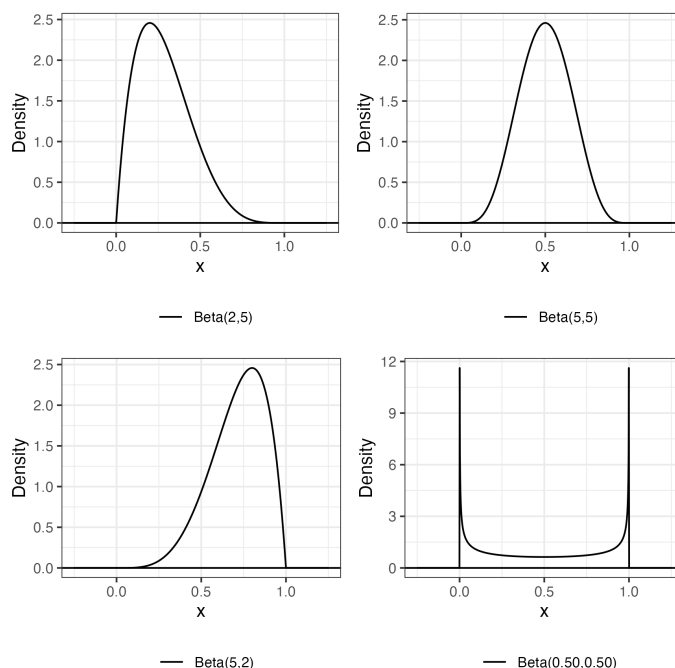


Figure 1: Beta distribution PDFs for different parameter values of α and β

The four plots in Figure (1) illustrate Beta probability density function with various combinations of parameters α and β . 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 $\alpha, \beta < 1$). The density is highest at the endpoints of the interval and lowest around 0.5.

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)).

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(5,5)). 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 it is equal to 0, the distribution is as peaked as the normal distribution is (mesokurtic). If it is positive, the distribution has a sharp peak and flat tails (leptokurtic). If excess kurtosis is negative, the distribution has a plat peak and thin tails (platykurtic). All four Beta distributions analyzed are platykurtic.

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.

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 (2).

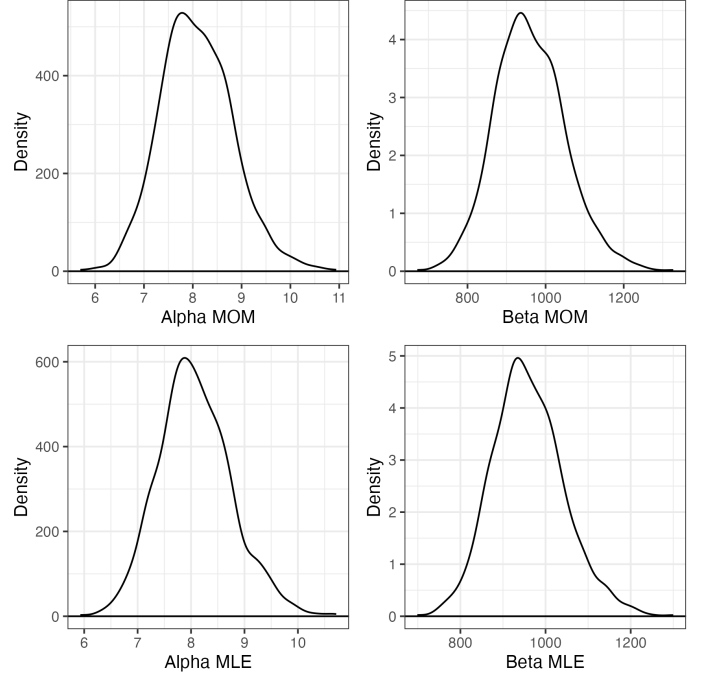


Figure 2: Density of estimates for α and β using MOM and MLE

We can see that the distributions of the estimated parameters are similar in shape. However, there is a slight difference in peak locations for MOM and MLE estimates. MLE has slightly lower means and significantly lower variability compared to MOM estimates.

We also computed bias, precision, and mean squared error for estimates in Figure (2). 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 application for the Beta distribution, consider country death rates. We draw the data set from the World Bank and convert the data to a rate of number of deaths per 1000 citizens to satisfy the interval $X \in [0, 1]$.

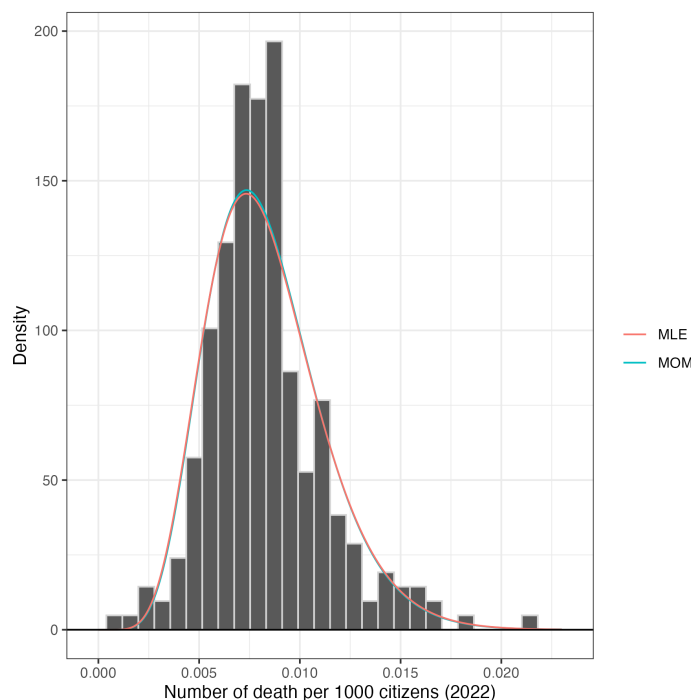


Figure 3: Histogram of death rates with MOM and MLE distributions superimposed

5.1 Methods Subsection

Describe the data you are working with, if applicable. Describe the specific process you will follow to answer the question at hand. This does not mean you should write something like this.

I did this and then I did that and then I did this other thing and then..., and then..., and then...

Instead, it should provide a clear and concise narrative that flows from the problem specification in the Introduction to how you will approach answering it. This is where I would expect to see some citations for R packages you will use to conduct the statistical analysis reported in the Results section. Much like the Introduction, subsections can be helpful for the

Methods section. For example, you might describe data collection and the statistical analyses of the collected data in different subsections. Or, you may have different questions that require distinct methods.

6 Results

Tie together the Introduction – where you introduce the problem at hand – and the methods – what you propose to do to answer the question. Present your data, the results of your analyses, and how each reported aspect contributes to answering the question. This section should include table(s), statistic(s), and graphical displays. Make sure to put the results in a sensible order and that each result contributes a logical and developed solution. It should not just be a list. Avoid being repetitive.

6.1 Results Subsection

Subsections can be helpful for the Results section, too. This can be particularly helpful if you have different questions to answer.

7 Discussion

You should objectively evaluate the evidence you found in the data. Do not embellish or wish-terpet (my made-up phrase for making an interpretation you, or the researcher, wants to be true without the data *actually* supporting it). Connect your findings to the existing information you provided in the Introduction.

Finally, provide some concluding remarks that tie together the entire paper. Think of the last part of the results as abstract-like. Tell the reader what they just consumed – what’s the takeaway message?

Bibliography: Note that when you add citations to your bib.bib file *and* you cite them in your document, the bibliography section will automatically populate here.

8 Appendix

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