Lab XX – MATH 240 – Computational Statistics

First Author
Affiliation
Department
`email@domain`

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 validated numerically. Parameter estimates were generated using the Method of Moments and Maximum Likelihood Estimation. These methods were evaluated through simulations and applied to global death rate data from the World Bank (2022) to assess the estimator performance.

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

1 Introduction

The Beta distribution is a powerful tool for modeling continuous variables that are bounded between 0 and 1. This makes it particularly useful in applications such as proportions, probabilities, and rates. The distribution is defined by two shape parameters α and β, which control the shape of the density function. By varying these parameters, the Beta distribution can take many forms—left-skewed, right-skewed, symetric, or U-shaped.

Building on this, this lab investigates the Beta distribtuion by deriving its key statistics using formulas for the mean, variance, skewness, and kurtosis. These properties are computed directly and explored across different parameterizations to understand how shape parameters influence distribution behavior. The lab also examines how sample-based summaries behave in relation to the population characteristics, and it compares the Method of Moments and Maximum Likelihood Estimation using both simulated and real-world data. The goal is to develop a comprehensive understanding of the Beta distribution's features.

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 entirely governed by α and β. Four common forms include:

- 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 can be derived from its PDF:

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

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

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

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

These expressions were confirmed by comparing theoretical values to those computed using `R` functions written for each formula. Additional tools such as the `cumstats` and `tidyverse` packages were used to summarize the behavior of these properties in samples.

4 Estimators

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

- **Method of Moments (MOM):** Matches sample mean and variance to theoretical expressions to solve for parameters.

- **Maximum Likelihood Estimation (MLE):** Uses numerical optimization to maximize the log-likelihood function derived from the PDF.

5 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.

5.1 Results Subsection

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

6 Discussion

You should objectively evaluate the evidence you found in the data. Do not embellish or wish-terpet (my made-up phase 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.

References

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

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.

7 Appendix

If you have anything extra, you can add it here in the appendix. This can include images or tables that don't work well in the two-page setup, code snippets you might want to share, etc.