Labs 7-9 – MATH 240 – Computational Statistics

Harrison Wolfe
Colgate University
Math Department
hwolfe@colgate.edu

4/3/2025

Abstract

In this lab we analyzed the beta distribution and it's real world application. We attempted to annalyze the unique attributes of the beta distribution and then display the results in graphs and tables to show how accurately we can use the beta distribution given a sample or how well we can estimate the population level properties like mean, variance and skewness using estimators.

Keywords: Beta distribution; Estimation; Parameters; Probability Distributions

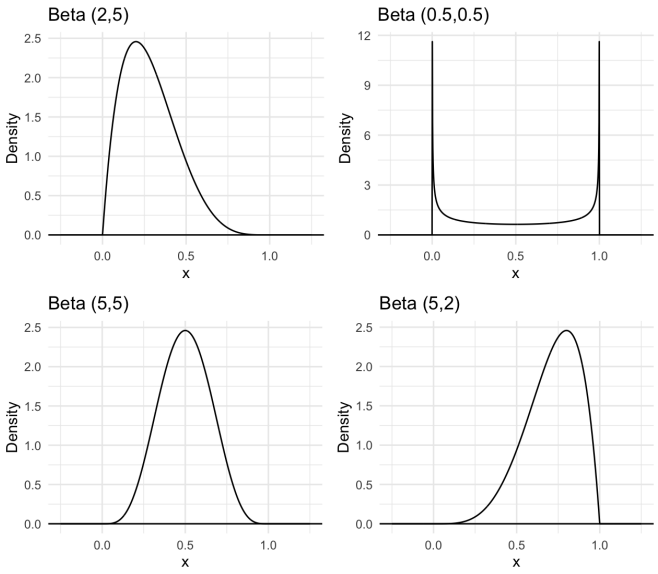

Figure 1: Beta Distributions Given Different Parameters

1 Introduction

The aim of this lab was to analyze several aspects of the beta distribution. We analyzed the parameters, alpha and beta, the properties of the distribution based on a population and sample. We also completed an example of a real world application of the beta distribution discussing death rates in various countries.

2 Density Functions and Parameters

The probability density function or PDF of the beta distribution is based on the parameters alpha and beta. The exact formula for the beta distribution is given below:

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt} \cdot 1_{(0<x<1)},$$

Some examples with density functions of different parameters can be seen below:

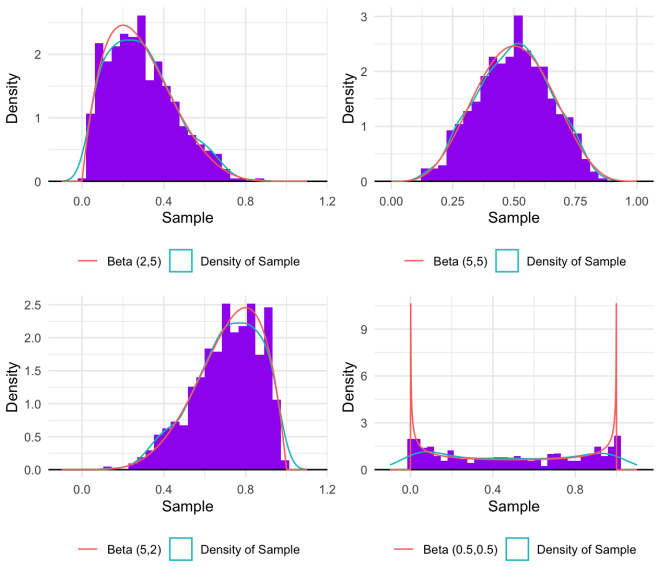

Figure 2: Samples of Beta Distributions

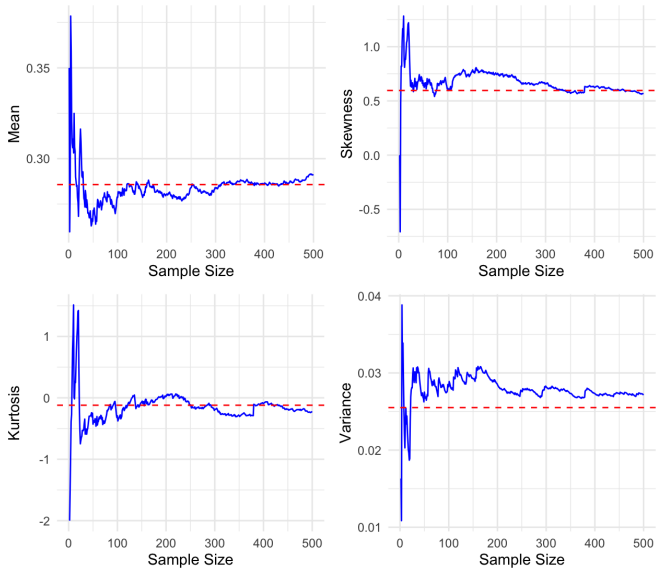

Figure 3: How Sample Size Effects Summary Stats

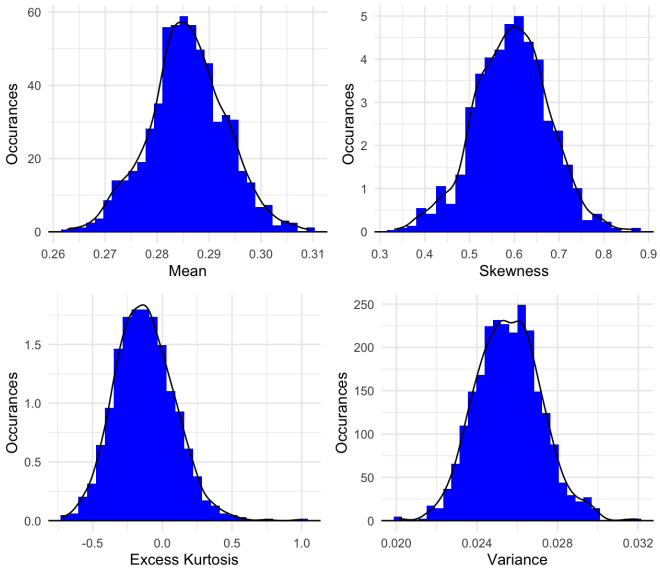

Figure 5: Sampling Distributions of Summary Statistics

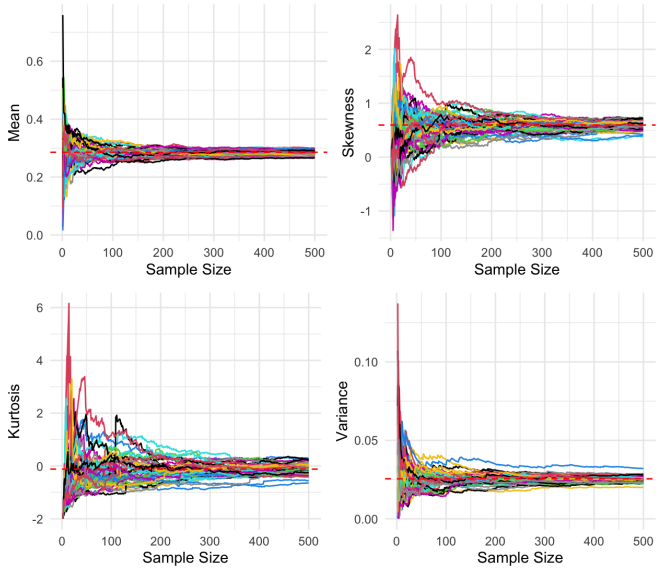

Figure 4: How Sample Size Effects Summary Stats (Simulated)

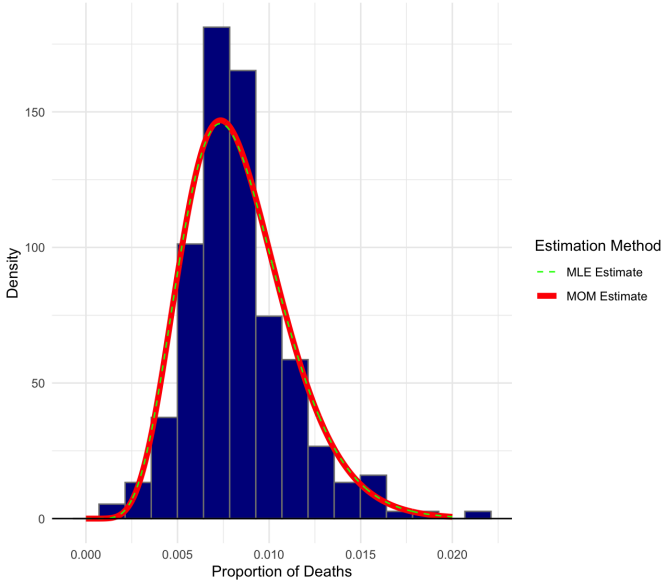

Figure 6: Superimposed Distributions with Estimates on Sample

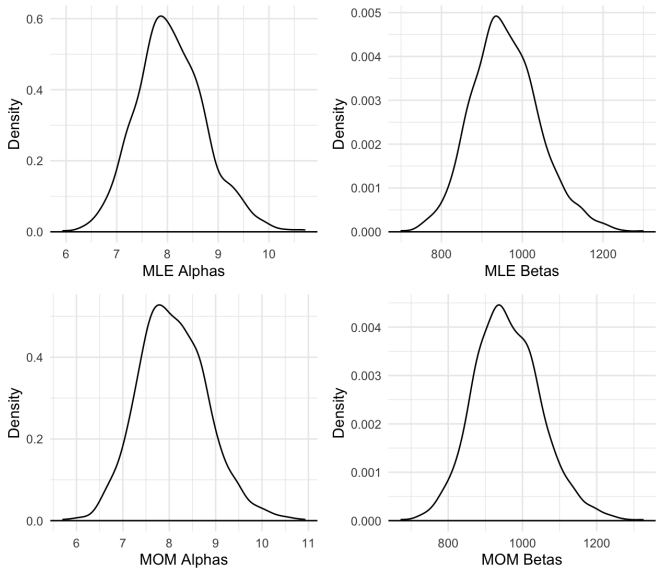

Figure 7: Simulated Estimators for Beta (8,950) Sample

2.1 Intro Subsection

You might need/want to discuss the topics in subsections. Or, you may have multiple questions.

3 Methods

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.

3.1 Methods Subsection

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.

4 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.

4.1 Results Subsection

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

5 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

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

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.

6 Appendix

These figures could not fit in the margins of the template above so they have been placed here. Each graph was created using `ggplot2` and organized using `patchwork` (Wickham, 2016) (Pedersen, 2024). Each table was made using `xtable` (Dahl et al., 2019).

	Alpha	Beta	Mean	Variance	Skewness	Excess Kurtosis	Method
1	2.00	5.00	0.29	0.03	0.60	-0.12	Formula
2	5.00	5.00	0.50	0.02	0.00	-0.46	Formula
3	5.00	2.00	0.71	0.03	-0.60	-0.12	Formula
4	0.50	0.50	0.50	0.12	0.00	-1.50	Formula
5	2.00	5.00	0.29	0.03	0.60	-0.12	Derived
6	5.00	5.00	0.50	0.02	-0.00	-0.46	Derived
7	5.00	2.00	0.71	0.03	-0.60	-0.12	Derived
8	0.50	0.50	0.50	0.12	-0.00	-1.50	Derived

Table 1: Summary Statistics Using Different Methods for the Beta Dist.

	Names	Bias	Precision	MSE
1	Alpha MLE	0.0720	2.1269	0.4754
2	Alpha MOM	0.0822	1.8286	0.5536
3	Beta MLE	9.1142	0.0001	7134.5577
4	Beta MOM	10.3424	0.0001	8290.0516

Table 2: Predictors to determine how good our estimators are