

# Lab 07 – MATH 240 – Computational Statistics

Ben Horner  
Colgate University  
Math Department  
email@domain

March 13, 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:** Beta Distribution; Probability Density; Moments; Sample Size; Variation

## 1 Introduction

The beta distribution is a continuous distribution that is used to model a random variable  $X$  that ranges from 0 to 1. This makes it useful for modeling proportions, probabilities, or rates. The beta distribution is also known for being remarkably flexible with regards to its shape — it can be left-skewed, right-skewed, or symmetric depending on the value of the parameters that define its shape:  $\alpha > 0$  and  $\beta > 0$ . We can use these parameters alone to not only define the distribution, but calculate the mean, variance, skewness, and excess kurtosis. Additionally, using the centered and uncentered moments of the beta distribution is another way to calculate these population-level characteristics.

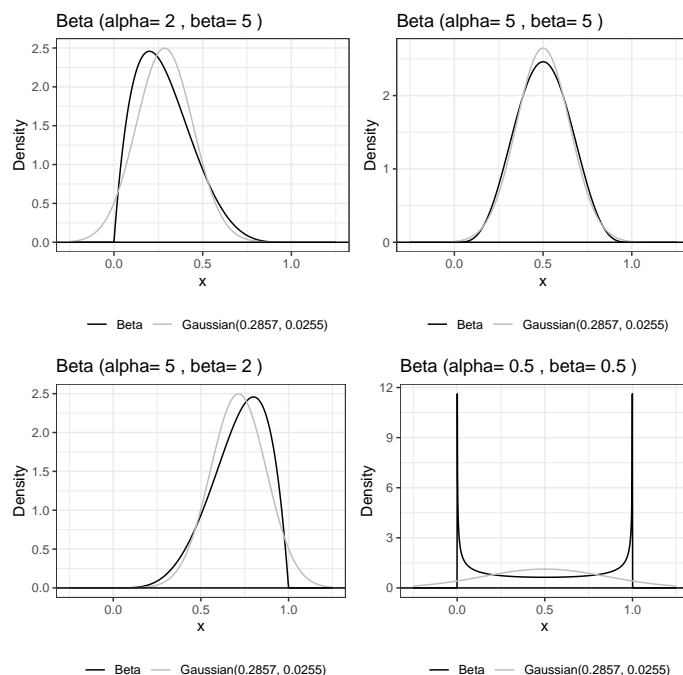
However, often times we may not have access to the beta, alpha, or moment of the distribution, and thus need to approximate it based off of the data. As such, we will examine whether data summaries help, is the sample size important, and also model the variation of the data summaries when compared to the population-level values.

## 2 Methods

The mean, variance, skewness, and excess kurtosis of a distribution are key summary characteristics that tell us more about the data we are looking at. With a beta distribution, we can calculate the population-level values using either the  $\alpha$  and  $\beta$  or the moments. Throughout this process, we load and use the following libraries: `ggplot2` (Wickham, 2016) and `patchwork` (Pedersen, 2024) for plotting and visualizing data, `tidyverse` (Wickham et al., 2019) to manipulate and summarize data, and `e1071` (Meyer et al., 2024) and `cumstats` (Erdely and Castillo, 2017) for additional statistical functions and analysis.

## 2.1 Describing and Summarizing The Population

As the beta distribution’s probability density function is defined in terms of the parameters  $\alpha$  and  $\beta$ , the population-level characteristics are similarly defined by them. Using a function we created to summarize the mean, variance, skewness, and excess kurtosis (hereafter referred to as key characteristics) and a function to plot a beta distribution, compared to a Gaussian, we considered four cases:  $\text{Beta}(\alpha = 2, \beta = 5)$ ,  $\text{Beta}(\alpha = 5, \beta = 5)$ ,  $\text{Beta}(\alpha = 5, \beta = 2)$ , and  $\text{Beta}(\alpha = 0.50, \beta = 0.50)$ . To confirm our function works, we also computed the population-level key characteristics via the moments of the beta distribution, which match those using alpha and beta.

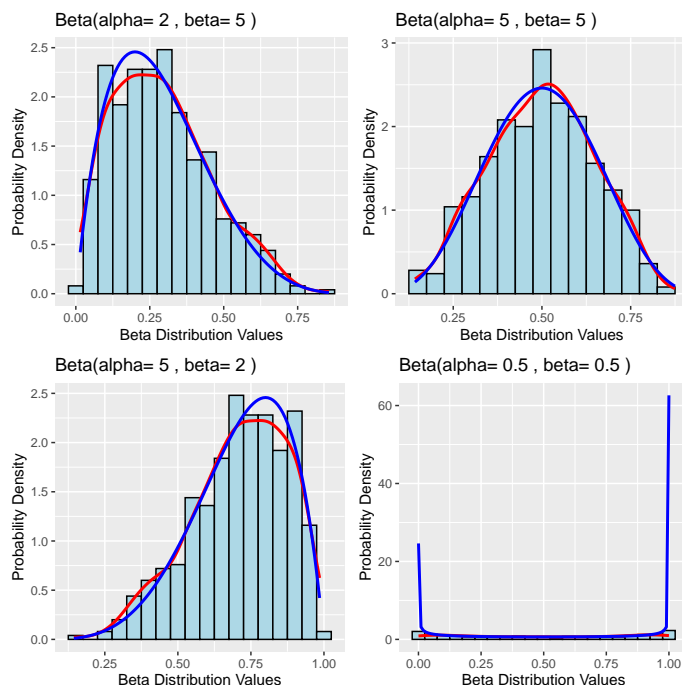


	beta_distribution	mean	variance	skewness	excess_kurtosis
1	2,5	0.29	0.03	0.60	-0.12
2	5,5	0.50	0.02	0.00	-0.46
3	5,2	0.71	0.03	-0.60	-0.12
4	0.5,0.5	0.50	0.12	0.00	-1.50

Table 1: Summary of Beta Distribution Parameters

## 2.2 Summarizing Random Data

When summarizing data, our goal is to approximate what the population distribution might be. We test this by generating a sample of  $n = 500$  from each of the considered beta distributions using `set.seed(7272)` to ensure consistency across samples and attempts. We summarize the data using `summarize()` from `dplyr` (Wickham et al., 2019).



## 3 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 an-

swering 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.

## 3.1 Results Subsection

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

## 4 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.

## References

- Erdelyi, A. and Castillo, I. (2017). *cumstats: Cumulative Descriptive Statistics*. R package version 1.0.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. (2024). *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*, TU Wien. R package version 1.7-16.
- 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.

## 5 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.