

Lab 7+8 – MATH 240 – Computational Statistics

Henry Sun
Colgate University
Mathematics

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

In this lab, we explored the beta distribution in great detail. The beta distribution is a continuous distribution, modeling the Provide an overarching summary of what you’re talking about. In this section, you introduce the idea to the reader, and your goal is to pull them in. What’s the mystery you aim to solve?

You want to provide enough background to understand the context of the work. Specifically, what is the question you are addressing? If it applies, describe what information currently exists about this problem, including citations, and explain how the question you’re answering complements this work.

Provide a roadmap of the structure of the paper.

2 Density Functions and Parameters

As the beta distribution is a continuous distribution, it has a probability density function (PDF) and a cumulative density function (CDF). It takes two shape parameters, α and β , which are both positive numbers. In R, the parameters for α and β are **shape1** and **shape2**. The beta distribution’s probability density function is given as:

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

For the beta distribution, $I(x \in [0, 1]) = 1$ when $x \in [0, 1]$ and 0 otherwise, as the support of a random variable X is $[0, 1]$ in the beta distribution. For some data, like the example we will look at later in the lab, data must be transformed to fit the support.

3 Properties and Statistics

Various properties of the beta distribution are based on the shape parameters of the beta distribution. The beta distribution is very flexible with regards to its shape. The various statistics, such as the mean, variance, skewness, and excess kurtosis can be written in terms of α and β :

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad (\text{The Mean})$$

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

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

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

It’s key to note that in R, **kurt()** calculates the excess kurtosis and not the kurtosis. Instead, $\text{kurtosis} = \text{excess kurtosis} + 3$, or **kurt(X) + 3**.

However, by computing the moments of the beta function, we write each statistic as a combination of various k th centered and uncentered moments.

$$\mu_X = E(X) \quad (\text{The Mean})$$

$$\text{var}(X) = \sigma_X^2 = E[(X - \mu_X)]^2 \quad (\text{The Variance})$$

$$\text{skew}(X) = \frac{E[(X - \mu_X)^3]}{E[(X - \mu_X)^2]^{3/2}} \quad (\text{The Skewness})$$

$$\text{kurt}(X) = \frac{E[(X - \mu_X)^4]}{E[(X - \mu_X)^2]^2} - 3 \quad (\text{The Excess Kurtosis})$$

For the beta distribution, the k th uncentered and centered moments are

$$E(X^k) = \int_{\mathcal{X}} x^k f_X(x) dx$$

and

$$E((X - \mu_X)^k) = \int_{\mathcal{X}} (x - \mu_X)^k f_X(x) dx$$

respectively. For the purposes of this lab, we wrote a function **beta.moment(alpha, beta, k, centered)** to help us compute the population-level characteristics using this moments.

maybe add code snippet here

Due to the beta distribution's flexibility, it can take on numerous different shapes, can be right-skewed, platykurtic, etc. Below is a data summary of this. **reference a figure showing this.**

3.1 Intro Subsection

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

4 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.

4.1 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.

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 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.

7 Appendix

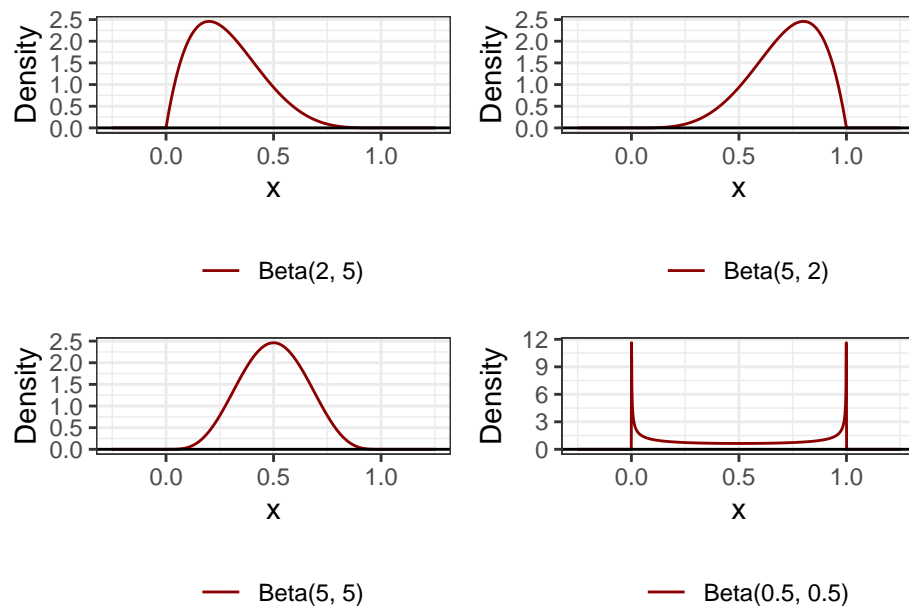


Figure 1: Plotting of various different beta PDFs.

alpha	beta	mean	variance	skewness	e.kurtosis
2.00	5.00	0.29	0.03	0.60	-0.12
5.00	5.00	0.50	0.02	0.00	-0.46
5.00	2.00	0.71	0.03	-0.60	-0.12
0.50	0.50	0.50	0.12	0.00	-1.50

Table 1: task 1 table

alpha	beta	mean	variance	skewness	e.kurtosis
2.00	5.00	0.29	0.03	0.57	-0.22
5.00	5.00	0.50	0.02	-0.06	-0.57
5.00	2.00	0.71	0.03	-0.57	-0.22
0.50	0.50	0.51	0.13	-0.02	-1.50

Table 2: task 3 table

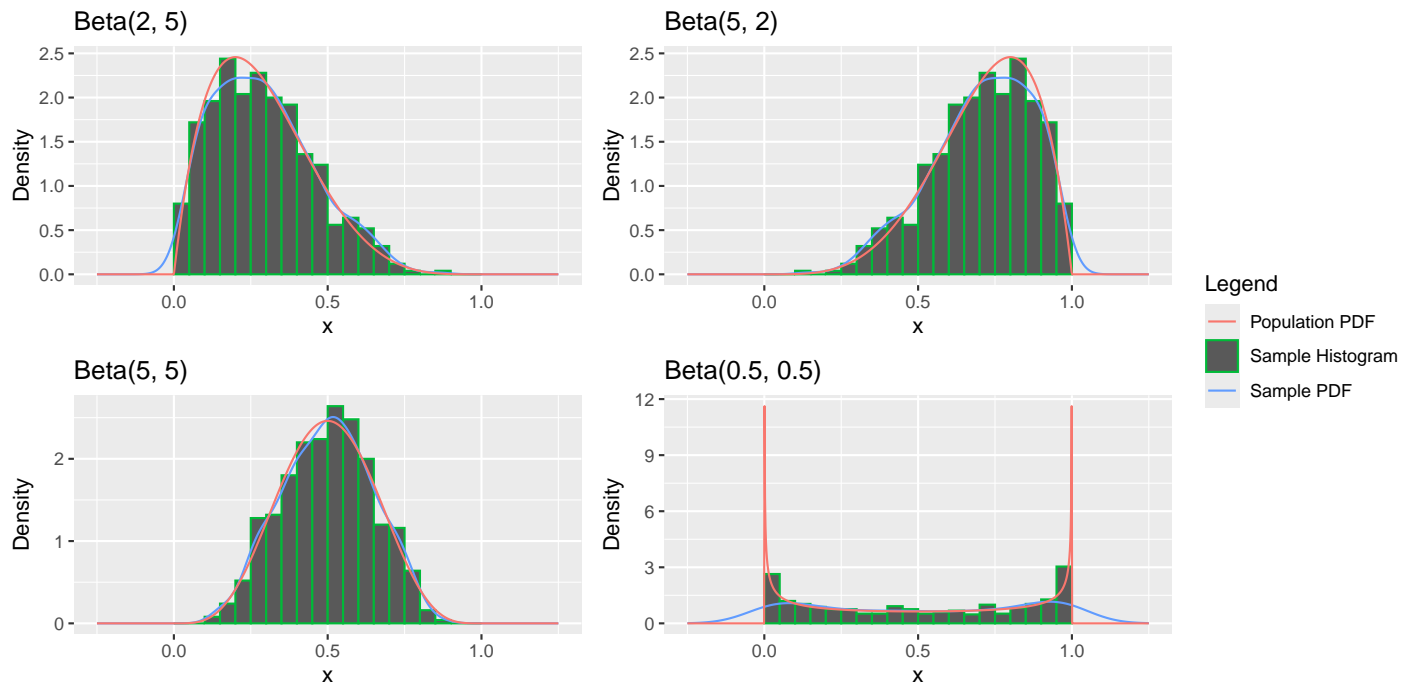


Figure 2: task 3 histograms

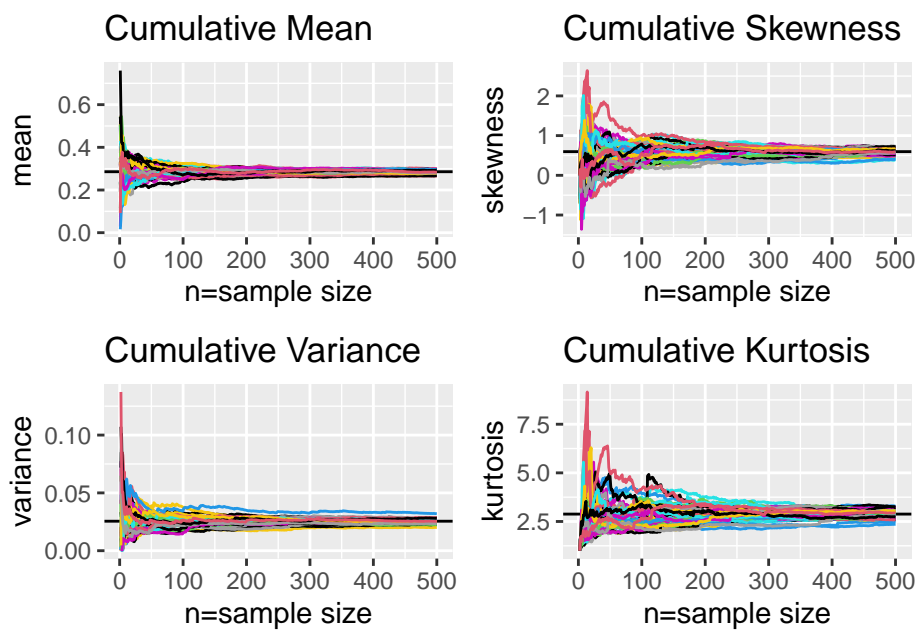


Figure 3: cumstats plots

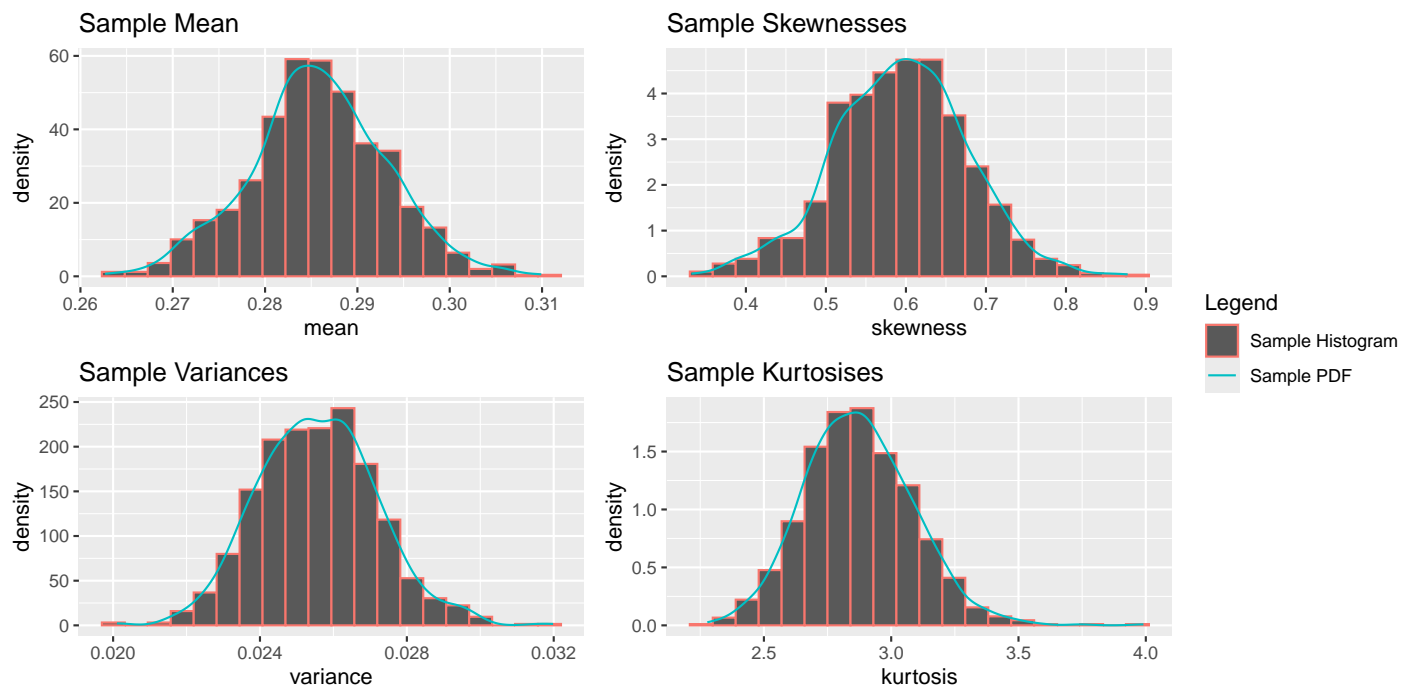


Figure 4: stats samplings dist plots

parameters	method	bias	precision	mse
Alpha	MOM	0.08	1.83	0.63
Alpha	MLE	0.07	2.13	0.54
Beta	MOM	10.29	0.00	8192.93
Beta	MLE	9.11	0.00	7058.82

Table 3: mles moms

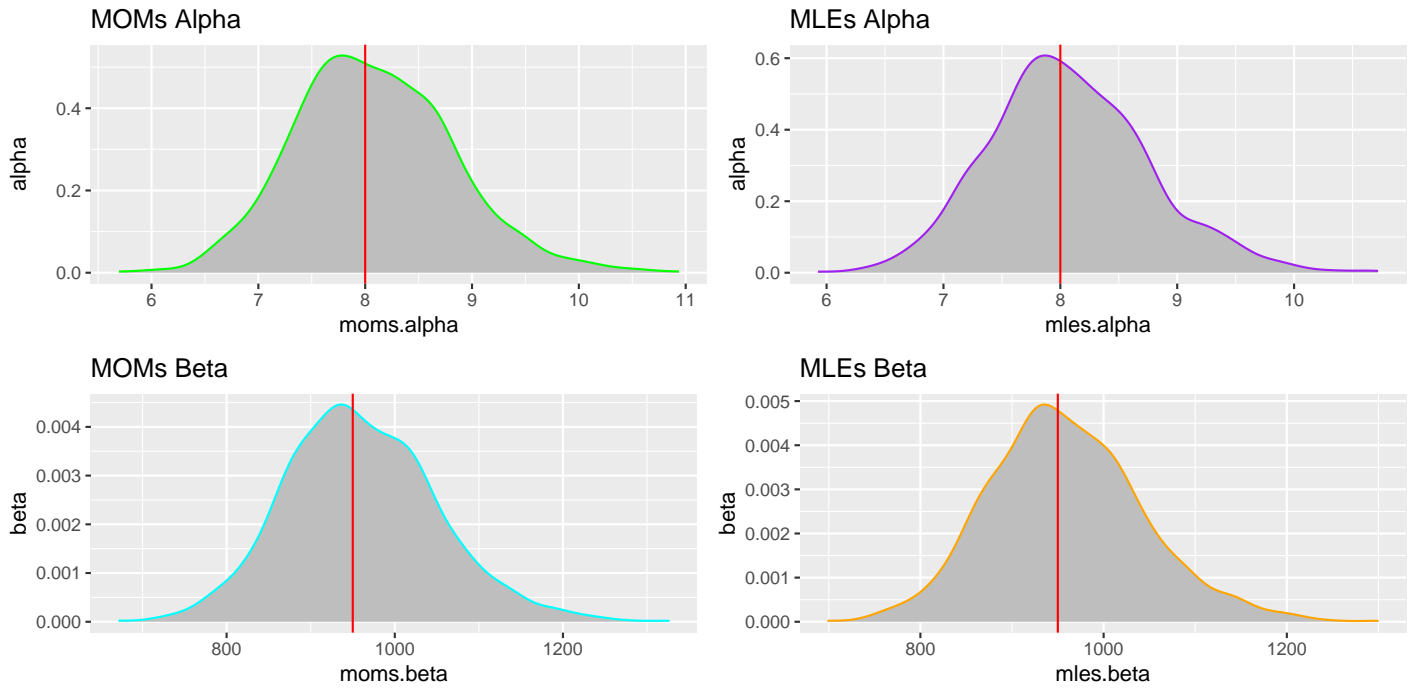


Figure 5: moms mles