

Lab XX – MATH 240 – Computational Statistics

Sankalp Ojha
Colgate University
Mathematics
sojha@colgate.edu

04/03/2025

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

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.

2 Density Functions and Parameters

The Beta distribution is a continuous distribution. A continuous distribution is one which has probability density function (PDF) and a cumulative density function (CDF) to represent its probability. The Beta distribution has two parameters, α and β , which dictate on the distribution looks. Both α and β are positive numbers and correspond to **shape1** and **shape2**, respectively, when write code for the Beta distribution in R.

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

In the beta distribution, $I(x \in [0, 1]) = 1$ when $x \in [0, 1]$ and 0 whenever x does not hold a number in support on the Beta distribution. The support of a random variable X in the Beta distribution is $[0, 1]$.

3 Properties and Key Statistics

The Beta distribution is a very versatile and flexible function when it comes to the shapes it can take on based on the various values of the α and β parameters. The various forms of the Beta distribution can be shown through statistics such as mean, variance, skewness, and excess kurtosis. Something to note about the `kurt()` function in R is that it computes the excess kurtosis. So, if we are interested in finding kurtosis of a distribution, we can simply add three to the excess kurtosis found through the `kurt()` function. The equations which can be used to calculate the statistics for a Beta distribution are written in terms of α and β , as seen below:

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

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

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

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

Another way to compute these statistics is through their centered and uncentered k th moments. We can compute various combinations of centered and uncentered k th moments of the Beta distribution to calculate a certain statistic, as seen below:

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

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

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

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

To generalize these equations to calculate any centered or uncentered k th moment of the Beta distribution, we can use the equations

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

and

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

where the first equation corresponds to a centered k th moment and the second equation corresponds to a uncentered k th moment.

For the purposes of this lab, a function `beta.moment(alpha, beta, k, centered)` was made to calculate the various statistics of the Beta distribution for various values of α and β .

3.1 Power of Larger Sample Sizes

By generating a large random sample from a fixed Beta distribution with parameters $\alpha = 2$ and $\beta = 5$, we were able to prove that larger random sample sizes converge closer to the exact distribution. Using the `cumstats` package Erdely and Castillo (2017), a figure illustrates how the Weak Law Of Large Numbers, which states that as the sample size increases so will the precision increase towards the total population distribution.

4 Estimators And Checking Precision

When dealing with data in real world applications, we don't necessarily know what the values of the parameters α and β will be for the given data set. This is where we use point estimators to calculate the precise values of α and β which will best fit the data set. The two methods to estimate include:

Method of Moments (MOM) and Maximum Likelihood Estimation (MLE). The MOM approach equates the first sample k th moments to the first population k th moments. The goal is to find which values of α and β will match the data set moments best with the population moments. The MLE approach utilizes a likelihood function which aims to find the values of α and β which will maximize the likelihood of matching the shape given data set. As both of these methods are estimates, we would like to know how precise the estimates are, which we can do through calculating Bias, Precision, and Mean Squared Error (MSE). Bias calculates the average difference from the true value, precision calculates the consistency of the estimates, and MSE combines both prior metrics to calculate the error. The equations to calculate each measure are:

$$\text{Bias} = \mathbb{E}(\hat{\theta}) - \theta$$

$$\text{Precision} = \frac{1}{\text{Var}(\hat{\theta})}$$

$$\text{MSE} = \text{Var}(\hat{\theta}) + \left(\mathbb{E}(\hat{\theta}) - \theta \right)^2$$

where $\hat{\theta}$ is a given vector of estimates and θ is the true value.

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.

5 Appendix

Table 1: Population Summary Of 4 Beta Distribution Parameters

	Alpha	Beta	Mean	Variance	Skewness	Excess Kurtosis
1	2.00	5.00	0.29	0.03	0.60	-0.12
2	5.00	5.00	0.50	0.02	0.00	-0.46
3	5.00	2.00	0.71	0.03	-0.60	-0.12
4	0.50	0.50	0.50	0.12	0.00	-1.50

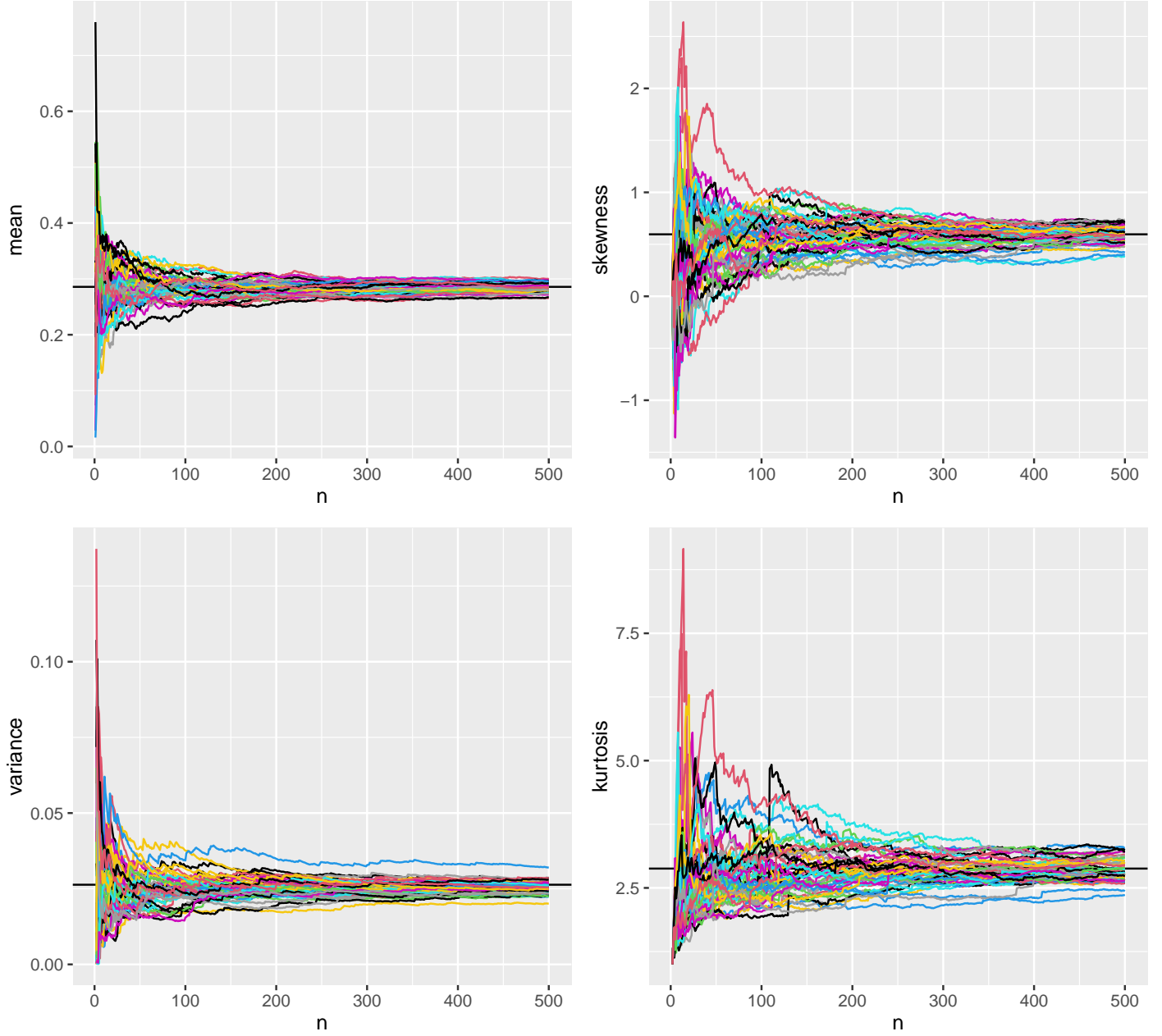


Figure 1: $\alpha=2$ and $\beta=5$ Convergence Simulation