Lab 7 and 8 – MATH 240 – Computational Statistics

Jack Schaeffer
Math 240
Professor Cipolli
jschaeffer@colgate.edu

April 1, 2025

Abstract

This lab is an overview of the beta distribution including potential uses and examples. Further examination of the distribution includes analysis of statistic summaries, the effect of sample size, and point estimator methods.

Keywords: Beta distribution; point estimators; moments of distribution

1 Introduction

This assignment is focused on analysis of the beta distribution to gain insight into its uses and properties. The beta distribution is a continuous distribution that models a variable X that has values within [0,1]. The shape and properties of the beta distribution are reliant on the parameters α and β ($\alpha > 0$, $\beta > 0$). The goal during the lab was to gain a better understanding of the beta distribution and its potential use in real world context. This included determining how altering sample sizes and parameters affects the beta distribution and comparing the efficacy of different point estimators in calculating α and β values.

My initial work was focused on examining the effect that α and β have on the distribution before moving into analysis of statistical properties and point estimators. This preparation included testing sample size's importance in producing accurate data, which was helpful for using a beta distribution to model the death rates collected from the World Bank (cite).

2 Density Functions and Parameters

Due to the beta distribution's separate shape parameters α and β, the distribution can have very different appearances and statistical properties depending on parameter values.

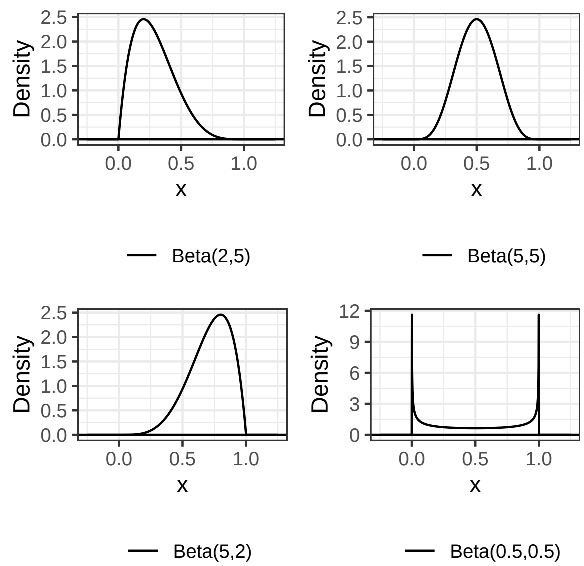

Figure 1: Density plots of the beta distribution with differing parameter values

Values	Mean	Variance	Skew	Kurtosis
Alpha = 2, Beta = 5	0.29	0.03	0.60	-0.12
Alpha = 5, Beta = 5	0.50	0.02	0.00	-0.46
Alpha = 5, Beta = 2	0.71	0.03	-0.60	-0.12
Alpha = 0.5, Beta = 0.5	0.50	0.12	0.00	-1.50

Table 1: Properties in comparison to differing parameter values

Figure 1 showcases the noticeable difference in the shape of the beta distribution depending on parameter values. These differences are reflected in the plots' properties in Table 1.

3 Properties

As is clearly shown from Figure 1 and Table 1, both the beta distribution's shape and characteristics are dependent on α and β. The population-level characteristics can be calculated by

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

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

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

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

The above equations are reflective of the answers seen in Table ˚(efdistrib.tab. For example, when α and β are equal, the skewness is always equal to zero.

4 Estimators

When the exact parameters of a beta distribution are unknown, we have to instead rely on data samples to estimate the distribution. To estimate population-level characteristics, we can use moments of distribution. The kth uncentered moment of distribution is

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

while the kth centered moment of distribution is

$$E[(X - \mu_X)^k] = \int_{\chi} (x - \mu_X)^k f_X(x)dx.$$

Using these moments, we can calculate the population-level characteristics as

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

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

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

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

To test the accuracy of data sampling compared to population-level values, we can overlay the population-level density plot over a histogram of the collected data.

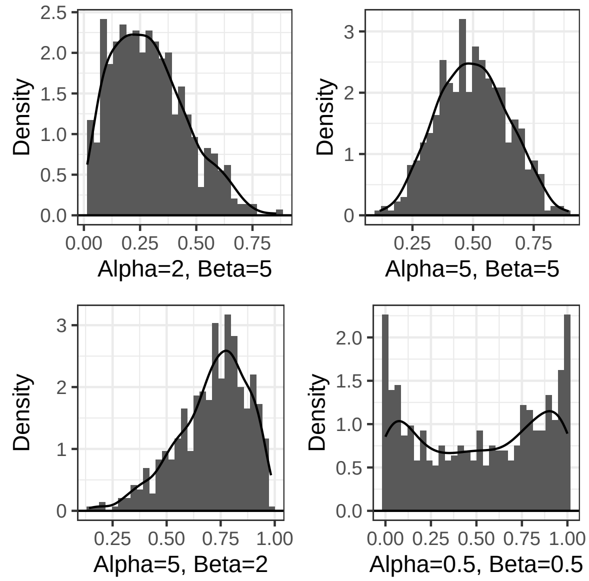

Figure 2: Comparison of estimator versus population-level values

Variable	mean	variance	skewness	kurtosis
Alpha = 2, Beta = 5	0.29	0.03	0.57	2.78
Alpha = 5, Beta = 5	0.50	0.02	0.06	2.54
Alpha = 5, Beta = 2	0.71	0.03	-0.74	3.22
Alpha = 0.5, Beta = 0.5	0.52	0.12	-0.11	1.55

Table 2: Estimated characteristics of the beta distribution for specified parameters

5 Example: Death Rates Data

To demonstrate the beta distribution's use in real life, I used data from the World Bank on death rates in 2022.

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.

6 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.