Lab XX – MATH 240 – Computational Statistics

Andrew Li

Colgate University

Mathematics Department

ali@colgate.edu

Abstract

In this lab, we explored the beta distribution has been explored thoroughly by working with its various properties, probability distributions, and parameters. By changing the parameters, we can evaluate the effects on its statistical values such as mean, variance, skewness, and excess kurtosis. To analyze real world data on global deaths from the World Bank, we made 2 point estimators (Method of Moments and Maximum Likelihood Estimations), which both work well but the *MLE* works slightly better.

Keywords: point estimations; parameters; probability distributions;

1 Introduction

The beta distribution is a continuous distribution that can be used to model the variability of a random variable X that ranges from 0 to 1. It is useful for modeling proportions, probabilities, or rates as its statistical characteristics are versatile enough to assume many different shapes based on its input parameters (Given that $\alpha > 0$, $\beta > 0$).

By exploring its properties, the effects of various inputs can be seen to answer our questions about what the beta distribution is, what it can be used for, what are some of its properties, and what useful inferences can be drawn from simulation and real data analysis.

2 Density Functions and Parameters

The beta distribution has a probability density function defined as:

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

Knowing that x stays within 0 to 1, we looked at the different cases of $Beta(\alpha, \beta)$ where $Beta(2,5), Beta(5,5), Beta(5,2),$ and $Beta(0.5, 0.5)$ and explored their properties.

The beta distributions for each plot is graphed together in Figure 1.

3 Properties

We calculated the population moments using numerical integration shown in Table 1.

Since the distribution's shape is affected by its parameters, the population characteristics are also controlled by them. To prove that its possible to approximate what the population distribution might be, we can connect our numerical summaries and graphs to the actual distribution by generating random data and comparing the calculated results against those from the known distribution, shown in Figure 2. The properties of mean, variance, skewness, and excess kurtosis are then compared against those from the population characteristics in Table 1.

Then we explored how the law of large numbers is proven to be true as the increasing sample size decreases the variability in the different properties of the data across different samples. This is shown in Figure 3 and remains true when we ran random samples and the graphical representations of the properties end up converging towards the population values as sample size increased.

	variable	mean	variance	skewness	kurtosis
1	Beta(0.5,0.5)	0.50	0.12	0.00	-1.50
2	Beta(2,5)	0.29	0.03	0.60	-0.12
3	Beta(5,2)	0.71	0.03	-0.60	-0.12
4	Beta(5,5)	0.50	0.02	0.00	-0.46
5	Sample Beta(0.5,0.5)	0.52	0.12	-0.11	1.55
6	Sample Beta(2,5)	0.29	0.03	0.57	2.78
7	Sample Beta(5,2)	0.71	0.03	-0.74	3.22
8	Sample Beta(5,5)	0.50	0.02	0.06	2.54

Table 1: population moments

4 Estimators

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 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.
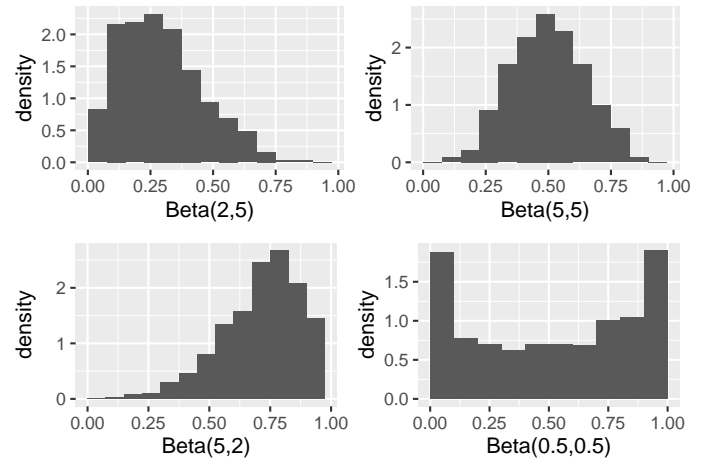
7 Appendix

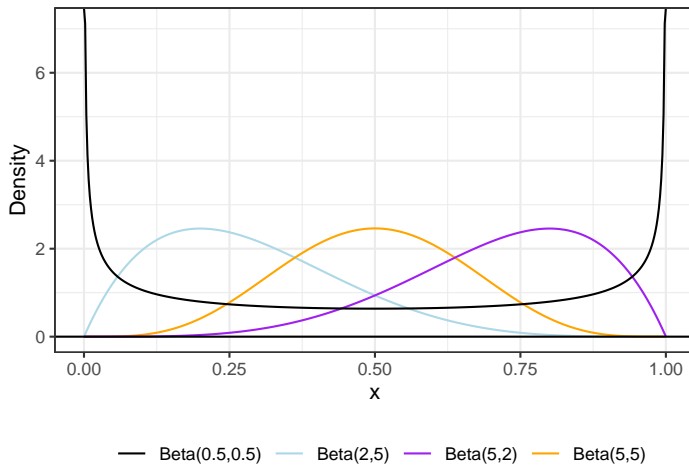Figure 2: Histograms of densities of beta samples
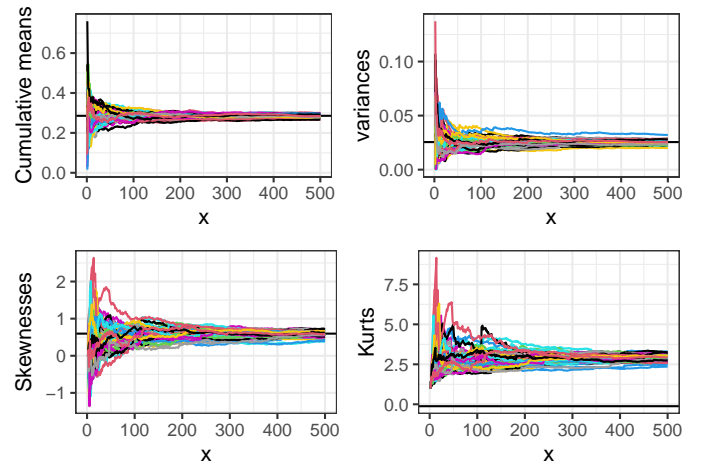

Figure 1: Distributions of different beta plots

Figure 3: Graphical comparison of random samples to population data