Lab 08 – MATH 240 – Computational Statistics

Cristian Palmer
Student
Mathematics
cpalmer@colgate.edu

1 Introduction

Lab 8 is a continuation of the work which we began during lab 7. In lab 7, we were tasked with computing the population moments for four distinct cases of the beta distribution and graphically comparing the different cases. In lab 8 we continued to build on our understanding of the beta distribution by modeling country death rates worldwide with the beta distribution. Our end goal with this lab was to be able to describe the beta distribution. Particularly, this write up aims to provide answers to questions such as: What is the beta distribution? What does it look like? What is it used for? What are its properties? And, what additional information do we gain from the simulations and real data analysis?

2 Density Functions and Parameters

To begin, the beta distribution's probability density function (PDF) is given by:

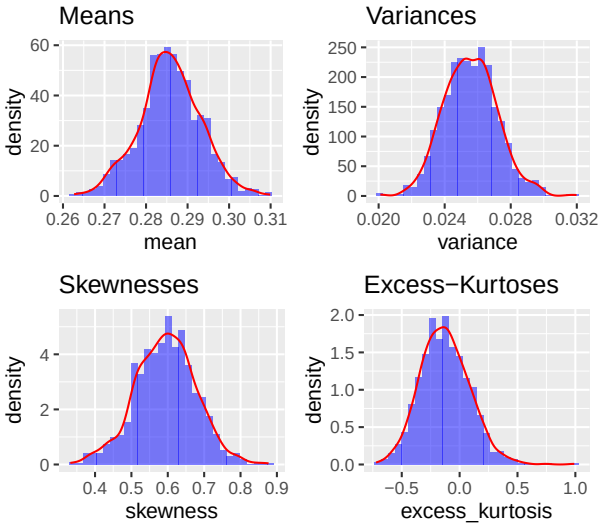
$$f_X(x \mid \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} I(x \in [0,1])$$

This PDF is expressed using the gamma function and involves the random variable x along with parameters alpha and beta. The domain of the beta distribution is restricted to the interval [0,1], meaning $0 \leq x \leq 1$. Additionally, both shape parameters alpha and beta must be strictly positive for the distribution to be properly defined.

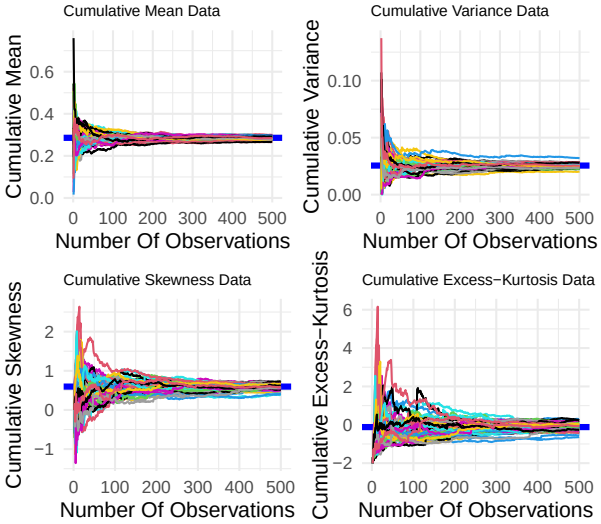
3 Properties

During lab 7 multiple of our tasks saw us calculating certain properties of the beta distribution for our alternate alpha and beta values. Since the distribution's shape is defined by its parameters, the distributions properties, notably its mean, variance, skewness, and excess kurtosis are also described by the alpha and beta parameters.

Pictured at the top of the second column is a 2x2 grid of histograms depicting the respective distributions and density curves for the beta distribution with an alpha of 2, and a beta of 5. We iterated over this specific beta distribution 1000 times, each time generating 500 values for each property which we saved to a tibble and graphed below. We utilized the `patchwork` (Pedersen, 2024) library in order to orient the four graphs in this 2x2 grid fashion.



Furthermore, the `cumstats` (Erdely and Castillo, 2017) package allowed us to compute the cumulative numerical summaries for each property. For 50 iterations of the alpha = 2 and beta = 5 beta distribution, we ran the `cumstats` functions on a sample size of 500 to generate a set of 50 line plots for each property. For each property, the plot below shows how with a large enough sample size, all the properties all average out to their true value. Plotted below is another 2x2 grid made using `patchwork`, however this grid depicts four line graphs, each showing 50 different iterations in which each property averages out to their respective true value as the sample size increases.



4 Estimators

In lab 8 we were tasked with computing the Method of Moments (M.O.M), and Maximum Likelihood (M.L.E) estimators.

5 Example

References

Erdely, A. and Castillo, I. (2017). *cumstats: Cumulative Descriptive Statistics*. R package version 1.0.

Pedersen, T. L. (2024). *patchwork: The Composer of Plots*. R package version 1.3.0.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, 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.

6 Appendix