Lab 08 – MATH 240 – Computational Statistics

Brendan Mariano
Colgate University
Mathematics
`bmariano@colgate.edu`

Abstract

This lab provides a template for understanding how

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 sought to determine whether 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.

1.1 Intro Subsection

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

2 Density Functions and Parameters

We began by producing the beta probability density function with four different parameters for alpha and beta: (2,5), (5,5), (5,2), and (.5,.5). Using the data and ggplot citepggplot2, we produced individual graphs for each set of parameters which also had a normal distribution containing the same mean and variance. Additionally, we included a Gaussian distribution in the plot which has the same mean and variance as the Beta distribution. When alpha and beta were equal at (5,5), the beta distribution resembled a normal distribution. When we decreased alpha to two, however, the data became right skewed, and when we decreased beta to two, the distribution became left skewed. Lastly, when we decreased both the alpha and beta to less than one, the distribution was symmetrical but had a greater density on the tails.
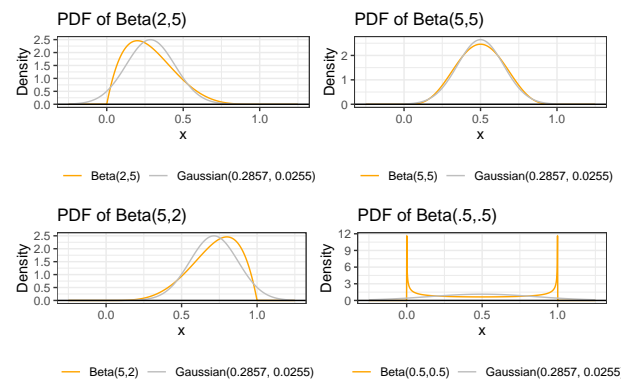

Figure 1

We used data about the death rates worldwide in 2022. We cleaned the data using the tidyverse (Wickham et al., 2019) and made it usable for a beta distribution (having every value between 0 and 1 inclusive). Since our overall goal was to compare the MLE and MOM, we calculated the values for each estimation method using R. We also created a graph using ggplot (Wickham, 2016) comparing the population distribution in a histogram to the MLE and MOM which were different colored lines. After working with the data from the worldwide death rates, we created 1000 samples of data using set seed so that we could calculate that alpha and beta values for each point estimation method. Using the 1,000 values for MLE alpha, MLE, beta, MOM alpha and MOM beta, we were able to create graphs that compared the MLE to the MOM. death rates in from almost every part of the world in 2022. This 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.

2.1 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.

3 Properties

There were four properties or moments, which we analyzed: mean, variance, skewness, and kurtosis. Moments are different indicators for a distribution. They can either be uncentered, indicating that they take they don't account for the mean, or centered which means that they account for the mean. The moments are advantageous because they allow us to calculate the values for a sample. The equations are as follows.

$$\mu_X = \mathbb{E}(X)$$

$$\sigma_X^2 = \text{var}(X) = \mathbb{E}\left[(X - \mu_X)^2\right]$$

$$\text{skew}(X) = \frac{\mathbb{E}\left[(X - \mu_X)^3\right]}{\left(\mathbb{E}\left[(X - \mu_X)^2\right]\right)^{3/2}}$$

$$\text{kurt}(X) = \frac{\mathbb{E}\left[(X - \mu_X)^4\right]}{\left(\mathbb{E}\left[(X - \mu_X)^2\right]\right)^2} - 3$$

When we created our plot comparing the actual death rates in 2022 to the MLE and MOM estimations, we found that they were all distributed very similarly. The MLE and MOM distributions were practically identical and the population histogram followed a very similar shape to the lines.

For our four plots comparing the distribution of Alpha and Beta values for MLE and MOM over 1,000 samples, the MLE had a more defined peak for both the Alpha and Beta values. The alpha value peaked at roughly 8 and the beta value peaked at roughly 950. The distribution of MOM values also peaked at 8 for alpha and 950 for beta, however, the peak is less sharp which suggests more variance within the data.

TALK ABOUT TABLE ADD TABLE AND GRAPHS 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.

3.1 Results Subsection

Subsections can be helpful for the Results section, too. This can be particularly helpful if you have different questions to answer.

4 Example

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.

References

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

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.

5 Appendix

	Element	Bias	Precision	MSE
1	Alpha (MLE)	0.07	2.13	0.48
2	Beta (MLE)	9.11	0.00	7132.70
3	Alpha (MOM)	0.08	1.83	0.55
4	Beta (MLE)	10.29	0.00	8288.46

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.