Lab 7 and 8 – MATH 240 – Computational Statistics

Anya Suko

27 March 2025

1 Introduction of the Beta Distribution

The Beta distribution is a continuous distribution, often used to model a random variable X that ranges from 0 to 1, making it useful for modeling proportions, probabilities, and rates. It is known for being remarkably flexible with regards to its shape- it can be symmetric or left/right skewed, depending on its parameters that define its shape: $\alpha > 0$, and $\beta > 0$.

2 Density Function and Parameters

The beta distribution's probablilty density function, which gives the likeliness that a continuous random variable takes on a specific value, is given by

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

where the density is zero anywhere outside of the range [0,1]. Since the distribution is directly impacted by its parameters (α and β), the population level characteristics are also described by those parameters.

A series of example distributions were created with different α and β values in order to demonstrate the effect that both α and β have on the distribution, and therefore the population level characteristics.

When α and β were set to 2 and 5, respectively,the population level statistics were the following:

	Statistic	Value
1	Mean	0.29
2	Variance	0.03
3	Skewness	0.60
4	Kurtosis	-0.12

When α and β were set to 5 and 5, respectively,the population level statistics were the following:

	Statistic	Value
1	Mean	0.50
2	Variance	0.02
3	Skewness	0.00
4	Kurtosis	-0.46

When α and β were set to 5 and 2, respectively,the population level statistics were the following:

	Statistic	Value
1	Mean	0.71
2	Variance	0.03
3	Skewness	-0.60
4	Kurtosis	-0.12

When α and β were set to .5 and .5, respectively,the population level statistics were the following:

	Statistic	Value
1	Mean	0.50
2	Variance	0.12
3	Skewness	0.00
4	Kurtosis	-1.50

Graphical representations of these distributions with α and β values can be seen in the appendix.

Upon comparing the results of the population level characteristics for the different parameter values, it can be noted that with equal α and β values, the mean is exactly one half, and the skewness is zero. However, with increases to both α and β value, the variance is smaller, and the kurtosis is also lower. When α is greater than β, when compared to equal parameters,the mean increases, and the skweness is negative (left skewed). When α is less than β, when compared to equal parameters, the mean is smaller, and the skewness is positive(right skewed). Here it is evident that the α and β parameters significantly impact the population level characterics.

3 Moments

The mean, variance, skewness, and kurtosis are all calculated using a moment of the distribution. In order to demonstrate the effect that using moments has on these values, the characteristics from the previous example were re-calculated:

$\alpha = 2$ and $\beta = 5$

	Statistic	Value
1	Mean	0.29
2	Variance	0.03
3	Skewness	0.03
4	Kurtosis	-0.12

$\alpha = 5$ and $\beta = 5$

	Statistic	Value
1	Mean	0.50
2	Variance	0.02
3	Skewness	-0.00
4	Kurtosis	-0.46

$\alpha = 5$ and $\beta = 2$

	Statistic	Value
1	Mean	0.71
2	Variance	0.03
3	Skewness	-0.03
4	Kurtosis	-0.12

$\alpha = .5$ and $\beta = .5$

	Statistic	Value
1	Mean	0.50
2	Variance	0.12
3	Skewness	-0.00
4	Kurtosis	-1.50

Here, it is evident that both avenues of calculating these statistics output nearly the same values. Since the formula for the moment is derived from the PDF funciton, it makes sense why these calculations would result in such similar values.

4 Population Distribution Approximation

The goal of summarizing data is to approximate the underlying population distribution. This section examines how graphical and numerical summaries reflect the actual population by generating random samples from known distributions and comparing sample-based estimates to theoretical values.

$\alpha = 2$ and $\beta = 5$

	Mean	Variance	Skewness	Excess_Kurtosis
1	0.29	0.03	0.03	-0.12

$\alpha = 5$ and $\beta = 5$

	Mean	Variance	Skewness	Excess_Kurtosis
1	0.72	0.02	-0.03	-0.12

$\alpha = 5$ and $\beta = 2$

	Mean	Variance	Skewness	Excess_Kurtosis
1	0.50	0.02	-0.00	-0.46

$\alpha = .5$ and $\beta = .5$

	Mean	Variance	Skewness	Excess_Kurtosis
1	0.52	0.12	-0.00	-1.50

Here, the difference in statistic calculations from the previous section examples comes from the fact that the data are samples, and therefore there is some variation. In the graphical representation of the repeated sampling, it can be seen that the outcome follows similar patterns to the actual population distribution, but that there are some places which are not exact, which is a result of the randomness of the sampling. While some variation occurs due to sampling, these summaries provide useful approximations for key distribution characteristics.

5 Sample Size and the Beta Distribution

This section demonstrates how cumulative numerical summaries—mean, variance, skewness, and excess kurtosis—change as more data points are added (sample size increases. The goal is to see how these statistics approach their true population values.

(Figure(s) is in the Appendix)

After careful examination of all four summary statistic graphs for the $\alpha = 2$ and $\beta = 5$ beta distribution, it can be seen that as sample size increases, the summary statistics each convergs towards the true value for the population for that summary statistic.This demonstrates that variablilty decreases as sample size increases

6 Modeling Variation with Sampling Distributions

For this section, in order to evaluate how statistics vary, each statistic was calculated 1000 times from a different sample of the population to create sampling distributions.

(Figure(s) in Appendix)

These distributions show that for every statistic, the sampling distribution shows something that reflects a bell curve, meaning that with lots of repeated samples, the sample values will get closer to the true population statistic values.

7 2022 Death Rates Example

World-wide Death rates(which have been collected through he World Bank, 2022) can be modeled with a beta distribution. First, in order to determine the beta distribution parameters, it is vital to complete both a Method of Moments (M.O.M) and Maximum Likelihood Estimante (M.L.E) to calculate α and β.

Using the M.O.M procedure results in $\alpha = 8.426105$ and $\beta = 1003.461$.

Using the M.L.E procedure results in $\alpha = 8.271409$ and $\beta = 985.0477$.

While these results are slightly different, they do give insight into where the true parameter value lie. Both of the M.L.E. results are slightly less than the M.O.M. results.

If $\alpha = 8$ and $\beta = 950$, and new data is used to generate new samples repeatedly, it creates a set of estimates for α and β for both the M.O.M. and M.L.E , which can then be summarized to see each distribution.

A graphical representation of this step can be found in the appendix.

In order to determine which kind of point estimation to use, it is best practice to look at the bias, precision, and mean squared error calculations. Bias should be as close to zero as possible, the precision should be as close to infinity as possible, and the mean squared error should be as small as it possibly can be. Below is a table demonstrating these statistics for each of the four distributions.

	Method	Parameter	Bias	Precision	MSE
1	MOM	Alpha	0.08	1.83	83.54
2	MOM	Beta	10.29	0.00	8265.64
3	MLE	Alpha	0.07	2.13	83.46
4	MLE	Beta	9.11	0.00	7132.70

For both kinds of point estimation, it is clear that the values are nearly identical, with few small changes. The best estimator for α is the M.L.E., and the best estimator for β is also the M.L.E. It would make the most sense to use the M.L.E. method to estimate parameters for the Beta Distribution of world wide death rates.

8 References

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... & Yutani, H. (2019). *Welcome to the tidyverse.* Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686

Dahl, D. B. (2023). *xtable: Export Tables to LaTeX or HTML.* R package version 1.8-4. https://CRAN.R-project.org/package=xtable

Becker, R. A., & Wilks, A. R. (2023). *cumstats: Cumulative Statistical Functions.* R package version 1.0. https://CRAN.R-project.org/package=cumstats

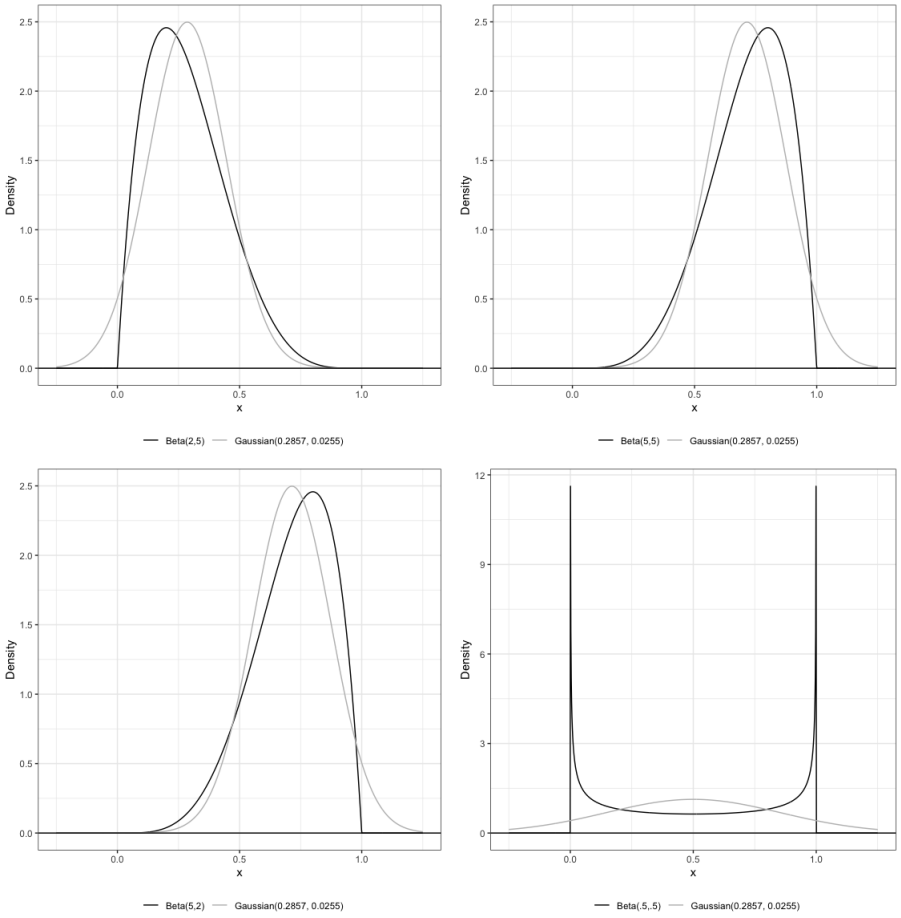
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer. https://ggplot2.tidyverse.org

World Bank. (2022). *Determinants of Mortality Rates from COVID-19: A Macro Level Analysis by Extended-Beta Regression Model.* Policy Research Working Paper No. 10149. The World Bank. https://doi.org/10.1596/1813-9450-10149

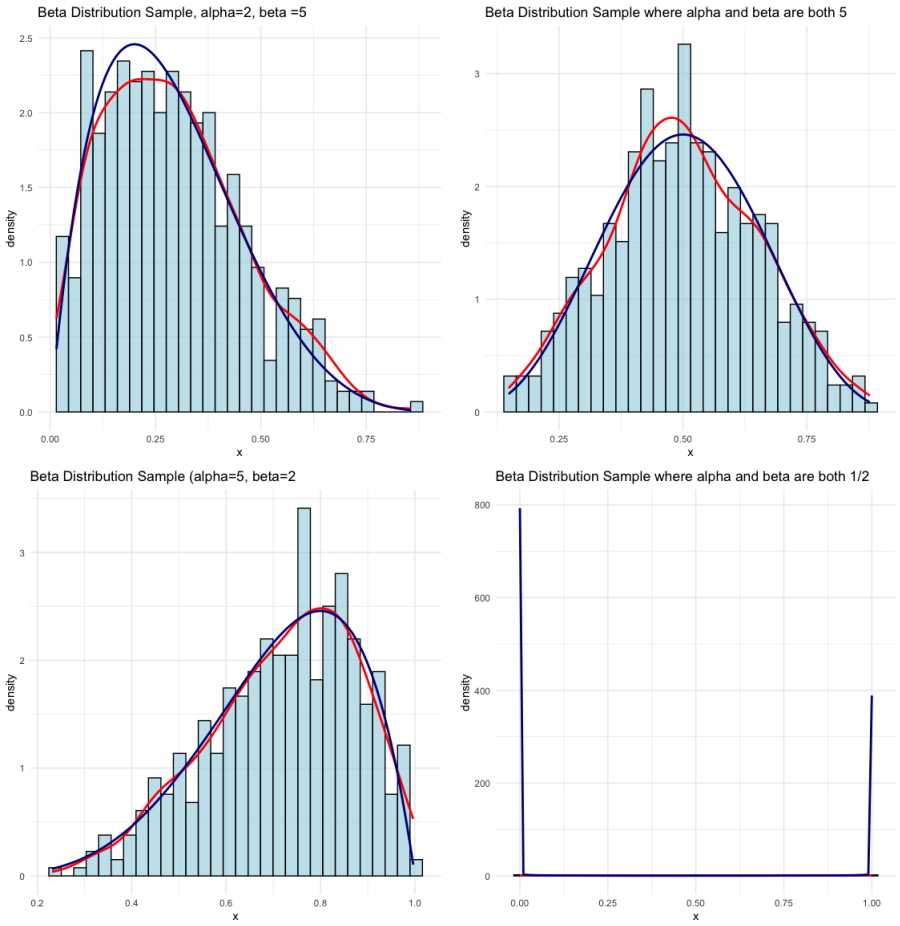
9 Appendix

Below are all of the graphs for the sections above, in the same order that they are mentioned

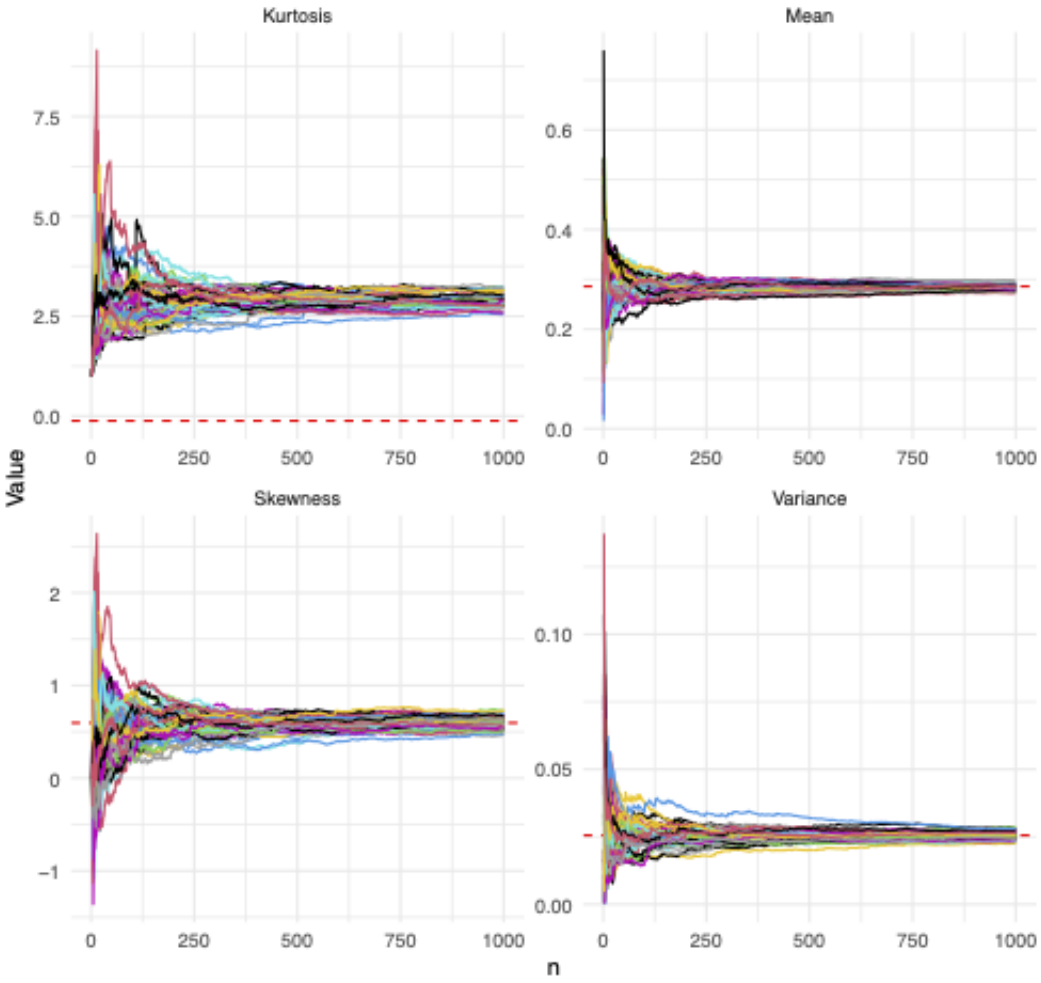
9.1 Density Function and parameters



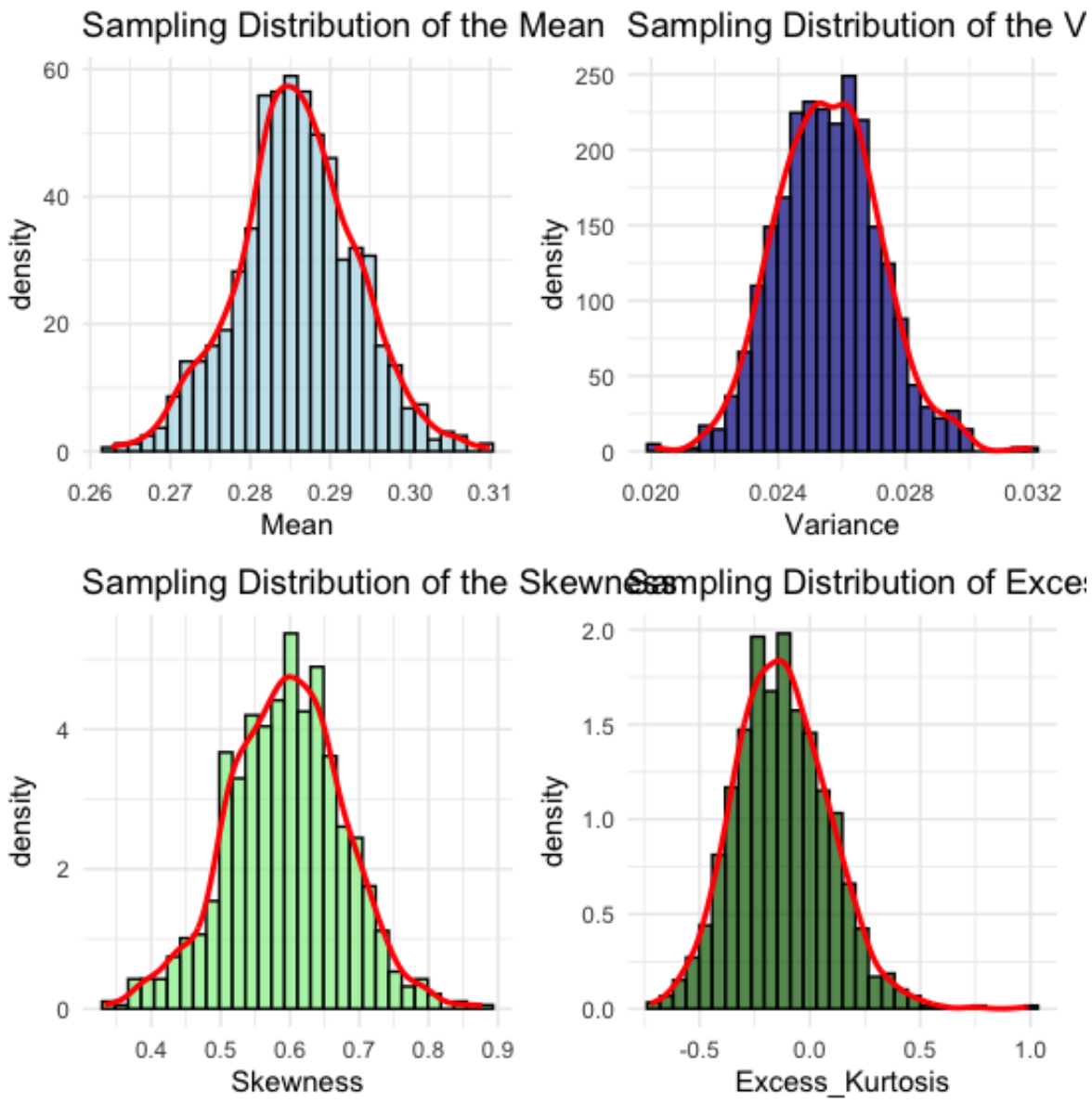
9.2 Population Distribution and Approximation



9.3 Sample Size and the Beta Distribution



9.4 Modeling Variance with a Sampling Distribution



9.5 2022 Death Rates Example