Lab 7-8 – MATH 240 – Computational Statistics

Jackson Colby
Colgate University
Mathematics
jcolby@colgate.edu

1 Introduction

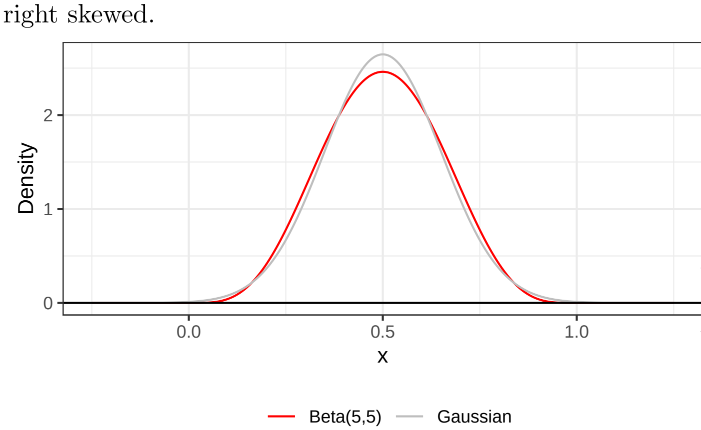
2 Density Functions and Parameters

The Beta distribution with parameters α and β is given by the probability density function:

$$f_X(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad \text{for } x \in [0,1]$$

The probability function for the Beta distribution takes values of 0 everywhere outside of [0,1].

The following plot shows the comparison between a Beta distribution with $\alpha = 5$ and $\beta = 5$ and a Gaussian distribution with the same mean and variance. This figure shows that when alpha and beta are the same or close to the same the beta distribution has a similar density to the normal distribution. If α is greater than β then the distribution will be left skewed and if α is less than β then the distribution will be right skewed.



A table containing the four given cases is below.

alpha	beta	mean	variance	skewness	kurtosis
2.00	5.00	0.29	0.03	0.60	-0.12
5.00	5.00	0.50	0.02	0.00	-0.46
5.00	2.00	0.71	0.03	-0.60	-0.12
0.50	0.50	0.50	0.12	0.00	-1.50

Table 1: Summary of Beta Distribution Statistics

As seen in the table, for larger α and β the variance is lower. All the Beta distributions are platykurtic, the graphs being more platykurtic when α and β are the same or similar.

3 Properties

The population level properties for the Beta Distribution are as follows: The mean is given by:

$$E(X) = \frac{\alpha}{\alpha+\beta}$$

The variance is given by:

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The skewness is given by:

$$\text{Skew}(X) = \frac{2(\beta-\alpha)}{(\alpha+\beta+1)\sqrt{(\alpha+\beta+2)\alpha\beta}}$$

The excess kurtosis is given by:

$$\text{Kurt}(X) = \frac{6(\alpha-\beta)^2}{(\alpha+\beta+1)\,(\alpha+\beta+2)} - \frac{\alpha\beta(\alpha+\beta+2)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}$$

A function called beta.moment was utilized to test both centered and uncentered moments of the population-level characteristics to the approximations calculated.

For the case alpha=2, beta=5: the statistics using the population level were mean=0.285714, var=0.0255102, skew=0.047619, kurt=0.001874. These statistics compared to the table above are very close to the values of the approximation, meaning the characteristics described above are true for the Beta Distribution.

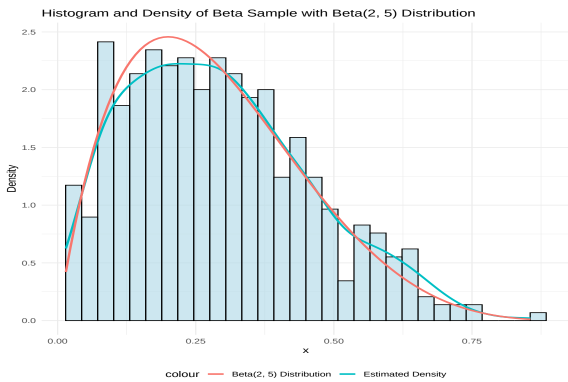
4 Estimators

For the case where alpha=2 and beta=5, the following summary statistics were calculated for a sample of 500.

mean_sample	variance_sample	skewness_sample	kurtosis_sample
0.29	0.03	0.57	-0.23

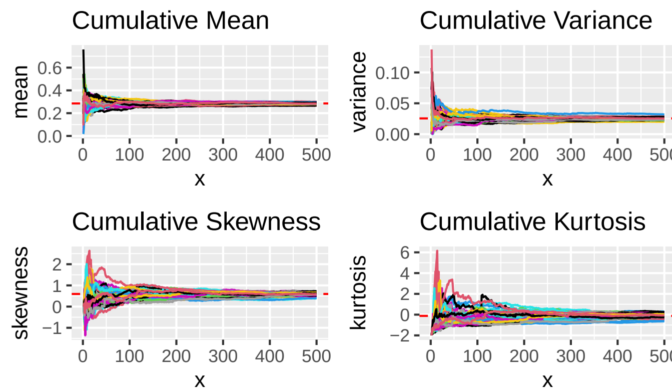
Table 2: Summary Statistics for Beta Sample

Below is a distribution showing the estimated density and the actual Beta Distribtuion over the sample.



As seen in the plot, the estimated density does a good job of fitting the data, and fits similarly compared to the actual Beta(2,5) Distribution.

Utilizing the packages cumstats (Bengtsson, 2022) and patchwork (Pedersen, 2020), we created plots that tracked the path of the summary statistics for different samples as n goes to 500.



As seen in the plot above, as n (the sample size), gets larger all the samples converge to the same values meaning that any random sample is sufficient for calculating the mean, variance, etc. of the Beta Distribution, and the sample size is important because the larger sample size leads to a better approximation.

The variation of the mean, variance, etc. can be shown in the plots below in the Appendix. All four of the plots follow close to normal distributions, showing that the samples of all mean, variance, etc. follow close to a Gaussian distribution, forming a bell shaped curve.

5 Example with Death Rates Data

References

Bengtsson, H. (2022). *cumstats: Cumulative Summary Statistics*. R package version 0.1.1.
Pedersen, P. S. (2020). *patchwork: The Composer of Plots*. R package version 1.1.1.

6 Appendix

6.1 Estimators