Labs 7 and 8 – MATH 240 – Computational Statistics

Pierce Leclerc
Colgate University
Department of Mathematics
pleclerc@colgate.edu

Abstract

In labs 7 and 8, we explored the beta distribution and its properties. We compared summary values of beta distributions under different parameter sets, also analyzing cumulative statistics. We then took death rate data and utilized both the method of moments and maximum likelihood estimates to determine approximate values of the parameters that would provide a matching distribution.

Keywords: Beta Distributions, Random Sampling, Estimators

1 Introduction

The beta distribution is a continuous probability distribution with two free parameters, α and β. It models a random variable X with a support of $[0, 1]$. The parameters can be adjusted to allow for flexibility in the distribution. For example, under specific values of α and β, a beta distribution can be skewed in different directions, have different excess kurtosis values, and other properties.

2 Density Functions and Parameters

Adjusting the parameters of beta distributions, α and β, can drastically alter their visual properties. Figure 1 provides four examples of different beta distributions.
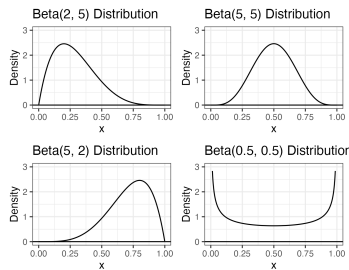

Figure 1: Beta distributions with different values of α and β.

From these same four distributions, we can analyze their properties numerically through summaries. Table 1 highlights the mean, variance, skewness, and excess kurtosis of their respective distributions.

Alpha	Beta	Mean	Variance	Skewness	Excess 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 values of beta distributions from different choices of α and β.

From generating samples of each distribution, we can observe how randomly generated data can approximate specific beta distributions in real-world scenarios. Figure 2 illustrates random sampling of the previous beta distributions with both estimated densities and true population densities superimposed.
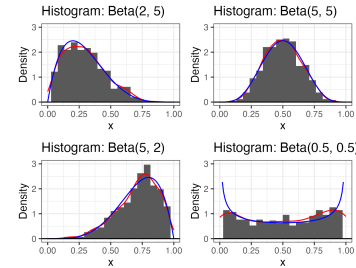

Figure 2: Histograms of random samples from different beta distributions.

3 Properties

From calculating cumulative statistics, we can attempt to observe convergence to true population quantities as the sample size increases through gradual decreases in variability. Figure 3 demonstrates this convergence through cumulative statistics of sampled data from the Beta(2,5) distribution. As sample size increases, for all four measures, the cumulative statistics visually converge to the values that describe the true population distribution (red horizontal lines).
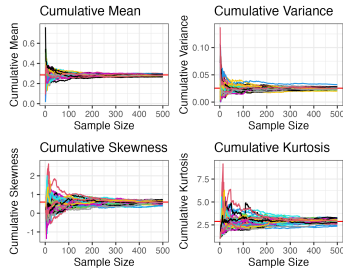
Figure 3: Cumulative statistics from sampled data of the Beta(2,5) distribution compared with sample size.

By storing statistics of interest from samples of the Beta(2,5) distribution, we can also observe their likely true values. Figure 4 displays the results, with each statistic appearing normally distributed after 1000 samples were taken.
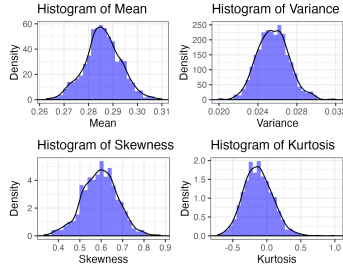

Figure 4: Distributions of summary values under random sampling of the Beta(2,5) distribution. All four appear to be normally distributed.

4   Estimators

2022 death rate data was collected from the World Bank and analyzed through point estimates in an attempt to determine a beta distribution which the data closely approximates as suggested by Fatih (2022). Method of moments and maximum likelihood estimators were obtained, and their distributions were superimposed with the death rate data in Figure 5.
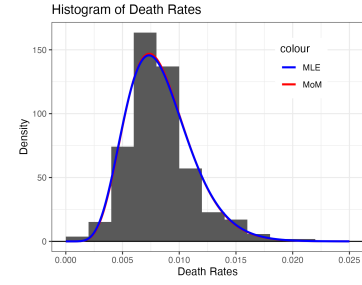

Figure 5: Histogram of 2022 death rates from the world bank, with MLE and MoM estimated distributions superimposed.

From 1000 samples, method of moments and maximum likelihood estimators were calculated, and Figure 6 visualizes the densities of each method's respective estimators in an attempt to determine which method provided the best estimates.
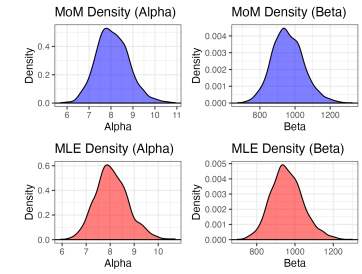

Figure 6: Estimated densities of alpha and beta for the method of moments and maximum likelihood estimates.

To aid these visualizations, Table 2 provides a numerical summary, highlighting bias, precision, and mean squared error (MSE) for the estimators.

Parameter	Method	Bias	Precision	MSE
Alpha	MoM	0.08	1.83	0.55
Alpha	MLE	0.07	2.13	0.48
Beta	MoM	10.29	0.00	8288.46
Beta	MLE	9.11	0.00	7132.70

Table 2: Bias, precision, and mean squared error for the method of moments and maximum likelihood estimates, with $\alpha = 8$ and $\beta = 950$.

From Figure 6 and Table 2, it is evident that the MLE estimators more closely approximate the data, as both bias and MSE are lower for each estimator than their method of moments counterparts.

References

Fatih, C. (2022). Determinants of mortality rates from covid-19: A macro level analysis by extended-beta regression model. *Revista de Salud Pùblica*, 24(2):1.