SI Assignment

Q1. To find the best probability distribution over the given data, we iterate through a list of common candidate distributions i.e., normal, exponential, gamma and beta distributions, and compare the Akaike Information Criterion scores for each of these distributions. The best fitting distribution will have the lowest AIC score. AIC is defined as:

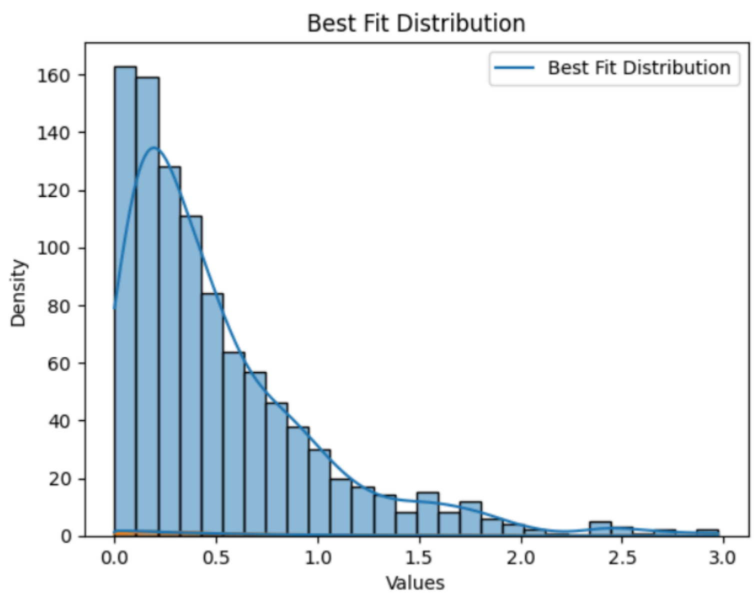
$$\mathrm{AIC} = 2k - 2\ln(\hat{L})$$

Where k is the estimated number of parameters and L is the Maximum Likelihood estimation for a given distribution.

RESULT:

Gamma distribution was adjudged as the best fitting distribution with parameters, shape (α) as 1.067 and scale ($\frac{1}{\beta}$) as 0.482.

GRAPH:



Q2. We use the Maximum Likelihood Estimation and Method of Moments to estimate the parameters of the underlying distribution we found in Q1.

Q3.

a) The estimates for the shape and scale parameters based on MLE as explained in Q1. , are asymptotically unbiased and consistent estimates.

b) Using the Maximum Likelihood Estimation for parameters as explained in Q1. , we get the following results:

```
Best distribution: gamma
Best parameters: (1.0670699752982742, 0.00031981962970932174, 0.4826855809531399)
AIC: 676.4614438683911
```

Q4.

To estimate the parameter of the gamma distribution using Method of Moments technique, we equate the population mean ($\mu = \alpha \cdot \beta$) with the sample mean ($\bar{x}$), we get,

$$\alpha \cdot \beta = \bar{x}$$

Now we equate the population variance($\sigma^2 = \alpha \cdot \beta^2$) with the sample variance (s^2), we get,

$\beta = \frac{s^2}{\bar{x}}$. After substituting $\bar{x} = \frac{\sum_{i \leq n} x_i}{n}$ and $s^2 = \frac{\sum_{i \leq n} (x_i - \bar{x})^2}{n-1}$

we get the following results :

```
Estimated parameters using method of moments:
Shape (alpha): 1.0895374157643936
Scale (1 / beta): 0.47303685230856923
```

Q5.

The Uniform Minimum Variance Unbiased Estimator for the shape (α) parameter can be approximated as:

$$\text{UMVUE}(\alpha) = 2n - 1 / \sum_{i=1}^{n} \left[\log(x_i) - \log(\hat{x})\right]^2$$

Where x_i is the i-th data point and $\hat{x}$ is the population mean.

The UMVUE for the inverse scale (β) parameter can be approximated as:

$\text{UMVUE}(\beta) \approx \frac{\sum_{1 \leqslant i \leqslant n} x_i}{n \cdot \text{UMVUE}(\alpha)}$. Using this, we obtain the following

result :

```
UMVUE estimate for the shape parameter (alpha): 1.1119116489600962
```

```
UMVUE estimate for the inverse scale parameter (beta): 0.46351825714535505
```

Q6.

We give the confidence interval estimates for the population mean parameter, represented by:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Where $\bar{x}$ is the population mean, σ is the population standard deviation and n is the population size. After substituting $\sigma = \sqrt{\frac{\sum_{i \leq n} (x_i - \bar{x})^2}{n}}$ we get the following result:

```
Confidence interval (α = 0.01): (0.47519229180585604, 0.5555904074453473)
Confidence interval (α = 0.05): (0.48480364514122765, 0.5459790541099756)
Confidence interval (α = 0.1): (0.48972133882025043, 0.5410613604309528)
```

Q7.

Here we perform the one sample t-test for hypothesis testing where the null hypothesis is $\mu = \mu_0$ and the alternative hypothesis is given by $\mu \neq \mu_0$. The t-test is given by $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$, where μ_0 is any value near the population mean, s is the sample standard deviation and $\bar{x}$ is the population mean (equal to the sample mean). For the purpose of this experiment, we take the value of μ_0 to be in the range $(\mu - \varepsilon, \mu + \varepsilon)$, where epsilon is in the range of (0, 0.5).

We get the following results:

```
Null hypothesis value (μ_0): 0.5291756492611704
Sample mean (x_bar): 0.5153913496256016
t-statistic: -0.8828128670923465
p-value: 0.3775497665033496
Fail to reject the null hypothesis
```

Q8.
Here we perform the chi-square test of the population variance for the hypothesis testing where the null hypothesis is given by $\sigma^2 = \sigma_0^2$ and the alternate hypothesis is given by $\sigma^2 \neq \sigma_0^2$.

Chi-square test is given by $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$, where n is the population size, σ_0 squared is any value near the population variance and s squared is the sample variance.
For the purpose of this experiment, we take the value if σ_0 squared to be in the range of $(\sigma^2 - \varepsilon, \sigma^2 + \varepsilon)$ where σ^2 is the

population variance and epsilon is taken to be in the range of (0, 0.5).

```
Null hypothesis value (σ^2_0): 0.47387713463419745
Sample variance (s^2): 0.24379910173395988
Chi-square statistic: 513.9629765429279
Critical value: 1073.6426506574246
Fail to reject the null hypothesis: There is no evidence to suggest that the population variance is different from σ^2_0.
```

Q9.
To perform the goodness of fit test over the distribution we estimated in Q1. , we follow the Kolmogorov-Smirmov test, where we compute the p_value and compare it with $\alpha = 0.05$ to perform the hypothesis test, where the null hypothesis is the assumption that data distribution follows the fitted gamma distribution and alternate hypothesis is otherwise. We get the following results:

```
Kolmogorov-Smirnov test:
KS Statistic: 0.02055539347141072
KS p-value: 0.7839204073824722
Fail to reject the null hypothesis: Data follows the fitted gamma distribution.
```