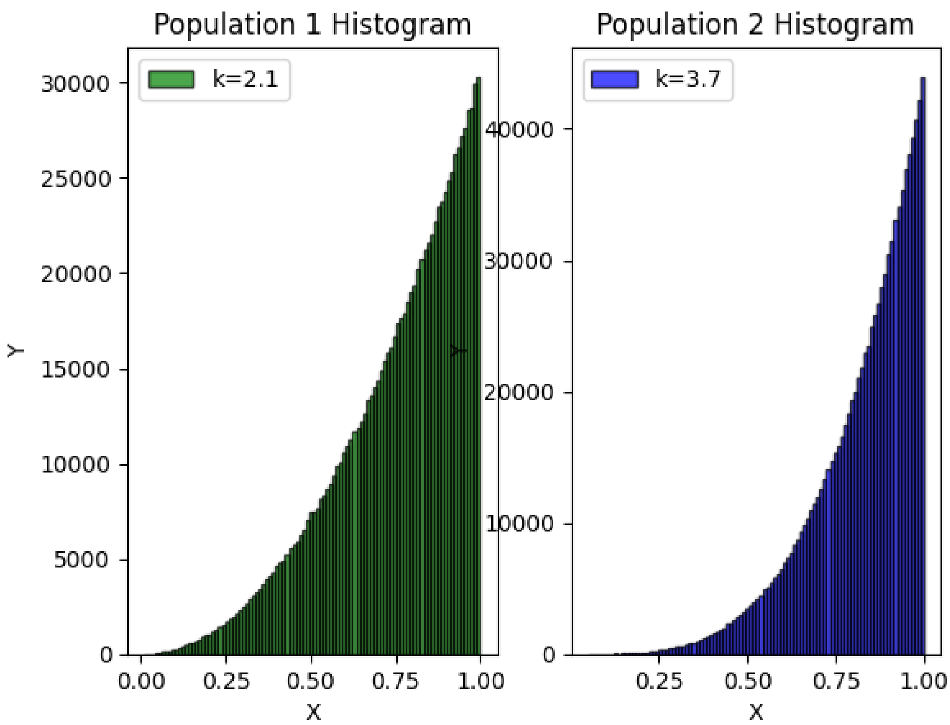
CENG222 HOMEWORK 2 EXPERIMENT REPORT

Ahmet Kurt

290201034

1. You need to synthetically generate two populations with different k values given in the skeleton code. These populations will only be used to sample from. You also need to calculate and print their means and variances and plot the **population histograms.**
   What does these values and plots tell you about the two investment strategies?

   Answer:



Population 1
Mean = 0.75609
Variance = 0.03618

Population 2
Mean = 0.82456
Variance = 0.02158

Population 1 has a lower mean compared to Population 2. Population 2 has a lower variance compared to Population 1. Let's say Population 1 is Strategy 1 and Population 2 is Strategy 2.

Strategy 1 has high potential gains, lower expected return but comes with higher risk, as indicated by the higher variance.

Strategy 2 has more consistent results, higher expected return, and lower risk, as indicated by the lower variance.

We must consider the risk factor:

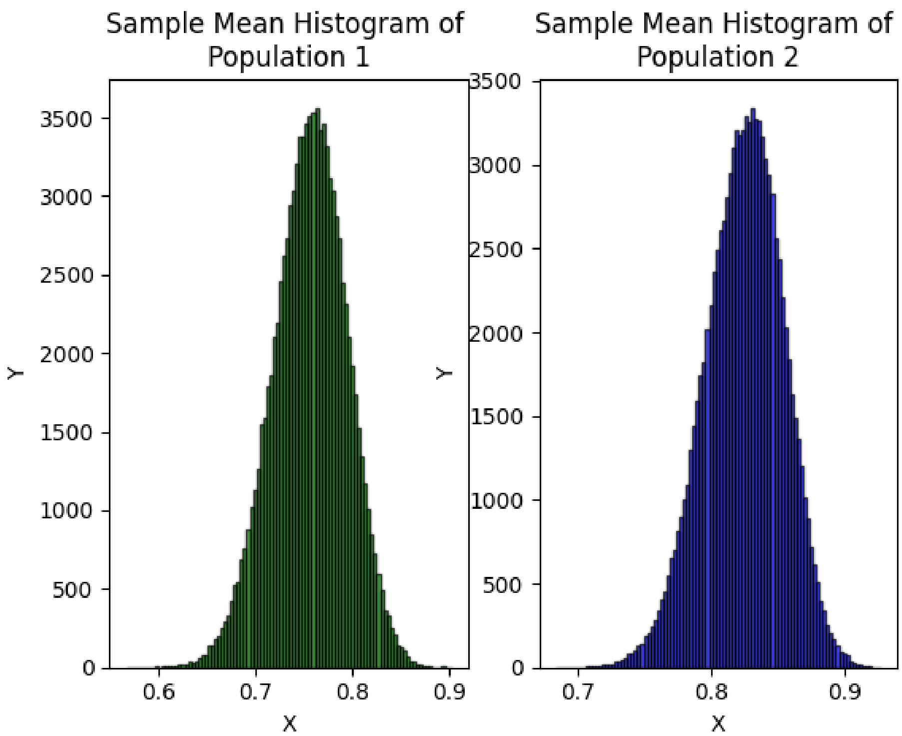
If we are preferring higher potential returns and accepting the higher risk, we may choose Strategy 1.

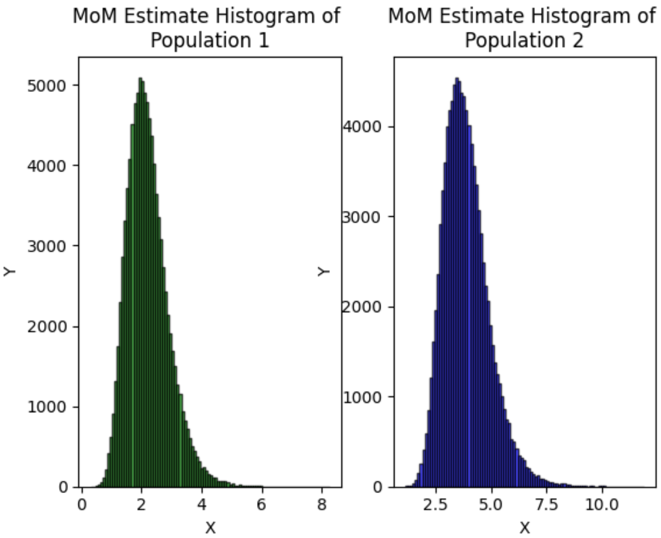
If we are preferring more stable returns and accepting the lower risk, we may choose Strategy 2.

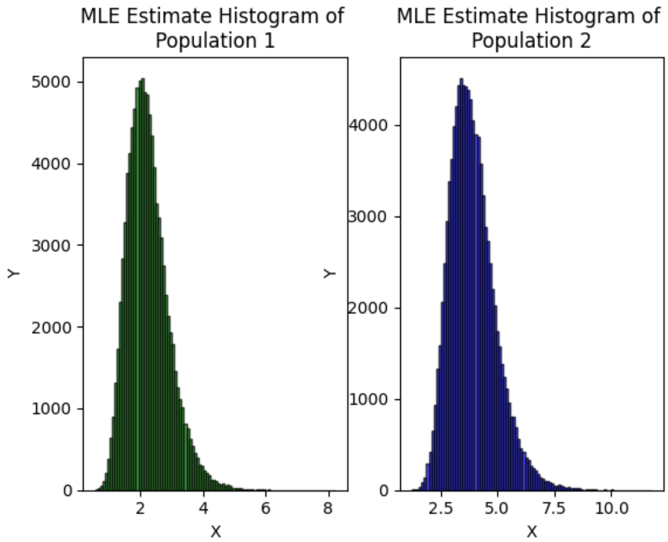
2. You need to calculate and print means of sample means, biased and unbiased sample variances, and MoM and MLE estimates of the parameter k for both populations. You also need to plot the **histograms of sample means, histograms of k estimates using MoM** and **MLE** for both populations.

   What do you observe for sample mean distributions and the mean sample means?
   What do you observe for biased and unbiased sample variances?
   What do you observe for the distributions of MoM and MLE estimates of k? Are these estimators unbiased?

   Answer:







Population 1

Mean of Sample Means: 0.7562197649513888

Mean of Biased Sample Variance: 0.034649607370738994

Mean of Unbiased Sample Variance: 0.03609334101118645

Mean of MoM Estimate: 2.206359011204086

Mean of MLE Estimate: 2.231808807678416

Population 2

Mean of Sample Means: 0.8245950670973012

Mean of Biased Sample Variance: 0.020765203031572924

Mean of Unbiased Sample Variance: 0.021630419824555124

Mean of MoM Estimate: 3.867522214612229

Mean of MLE Estimate: 3.89458744479006

Based on the variances, we can see that the fluctuations in population 1 are (left side) 0.6, while the largest fluctuations in population 2 did not fall below the 0.7.

The unbiased sample variance is slightly higher than the biased sample variance for both populations.

For population 1, the k values was 2.1, and for population 2, it was 3.7. In both populations, the histograms of the Method of Moments (MoM) and Maximum Likelihood Estimation (MLE) estimates are concentrated around 2.1 for population 1 and 3.7 for population 2.

In conclusion, the analysis confirmed that the sample means are centred around the theoretical means and that both MoM and MLE are reliable prediction methods, and the predictions are centred around the expected values. This consistency between different measurements and populations demonstrated the accuracy of statistical methods.

3. You need to calculate and print the ratio of confidence intervals computed with and without using the population standard deviation that contains the population mean for both populations.

   What do you observe for **<u>the confidence intervals computed with and without using the population</u>** standard deviation?

Answer:

<u>Population 1</u>

With population std: 0.97138

Without population std: 0.95263

<u>Population 2</u>

With population std: 0.97079

Without population std: 0.94903

There can be difference between confidence intervals which calculated with the population standard deviation and calculated without it.

This is because, especially for small samples, the sample standard deviation can deviate from the population standard deviation.

4. You need to collect a sample of length 100000*25 from both populations, calculate and print their sample means, biased and unbiased sample variances, MoM and MLE estimates of parameter k and confidence intervals with and without using the population standard deviation.

   What do you observe for **<u>all the sample statistics computed?</u>**
   What do you observe for the **<u>confidence intervals computed with and without</u>** using the population standard deviation?

Answer:

Given the large sample size, the sample means are expected to be close to the theoretical means.

Similarly, both biased and unbiased variances are observed to be very close to the theoretical variances. The same applies to the Method of Moments (MoM) and Maximum Likelihood Estimation (MLE). Consequently, with the large sample size, the confidence intervals are observed to be very similar to each other.

As a result, the confidence intervals derived from these estimations are very similar, reflecting the stability and reliability of the estimates due the large sample size.

MY CALCULATIONS

Ahmet KURT
290201034

① INVERSE TRANSFORMATION METHOD

Uniform random variable : U

The given pdf $f(x) = (k+1)x^k$ for $0 \leq x \leq 1$

The cdf is (cumulative distribution function) = $F(x) = \int_0^x (k+1)x^k dx$

$$= (k+1)\int_0^x x^k dx$$

$$= (k+1)\left[\frac{x^{k+1}}{k+1}\right]\Big|_0^x = x^{k+1}$$

random variable X;

$F(X) = U$; $U = [0,1]$ interval, we solve for X:

$X = F^{-1}(U) = U^{\frac{1}{k+1}}$

② REJECTION METHOD

$\Rightarrow$ $a=0$, $b=1$ and $c=k+1$ (the maximum value of $f(x)$)

$\Rightarrow$ Generated two random variables (U and V)

Transform U and V;

$X = a + (b-a)U$

$Y = cV$

$\Rightarrow$ If $Y \leq f(X) = Y \leq (k+1).X^k$ accept. Otherwise, reject.

③ EXPECTED VALUE OF THE GIVEN DISTRIBUTION

The expected value $E[X]$;

$$E(X) = \int x f(x)\, dx = \int_0^1 x(k+1)x^k dx$$

$$= (k+1)\int_0^1 x^{k+1} dx$$

$$= E(X) = (k+1)\left(\frac{x^{k+2}}{k+2}\right)\Big|_0^1 = \frac{k+1}{k+2}$$

Ahmet KURT
290201034

④ THE VARIANCE OF THE GIVEN DISTRIBUTION

$Var(X) = E(X^2) - (E(X))^2$

Compute the $E(X^2)$;

$$E(X^2) = \int_0^1 x^2 f(x)\,dx = \int_0^1 x^2 (k+1) x^k\,dx$$

$$= (k+1)\int_0^1 x^{k+2}\,dx$$

$$= (k+1)\left(\frac{x^{k+3}}{k+3}\right)\Bigg|_0^1 = \frac{k+1}{k+3}$$ and we know that $E(X) = \frac{k+1}{k+2}$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{k+1}{k+3} - \left(\frac{k+1}{k+2}\right)^2$$

⑤ THE METHOD OF MOMENTS ESTIMATES OF THE K FOR THE DIST.

$\hat{k}$ is obtained by the equating the sample mean to the expected value.

$\mu = \bar{X}$. Set X equal to $E(X)$:

$\bar{X} = \frac{k+1}{k+2}$ Solving for k: $\bar{X}(k+2) = k+1 \Rightarrow \bar{X}k + 2\bar{X} = k+1$

$$\Rightarrow k(\bar{X}-1) = 1 - 2\bar{X}$$

$$\Rightarrow k = \frac{1-2\bar{X}}{\bar{X}-1}$$

⑥ THE MAXIMUM LIKELIHOOD ESTIMATE OF THE K FOR THE DIST.

Likelihood function for a sample $X_1, X_2, \dots, X_n$ is:

$L(k; X_1, X_2, \dots, X_n)$

$= n\ln(k+1) + k\sum_{i=1}^{n} \ln X_i$, taking the derivative with respect to k

Solving for k

$$= \frac{d\,l(k)}{dk} = \frac{n}{k+1} + \sum_{i=1}^{n} \ln X_i = 0$$

$$\Rightarrow \frac{n}{k+1} = -\sum_{i=1}^{n} \ln X_i$$

$$k+1 = -\frac{n}{\sum_{i=1}^{n} \ln X_i} \Longrightarrow k = -\frac{n}{\sum_{i=1}^{n} \ln X_i} - 1$$

$$= \hat{k}\ \frac{-n}{\sum_{i=1}^{n} \ln X_i} - 1$$