Uniform Estimators Solutions:

Prerequisite:

$$\hat{L}_{MOM} = 2\bar{X}_n$$
$$\hat{L}_{MLE} = \max_{i=1,\ldots,n} X_i \tag{1}$$

$$\mathrm{MSE}(\hat{L}) = \mathrm{E}\left[\left(\hat{L} - L\right)^2\right] \tag{2}$$

1)

Proof:

$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left[\left(\hat{\theta} - \theta\right)^2\right] = \mathrm{E}\left[\left(\hat{\theta} - \mu + \mu - \theta\right)^2\right]$, where $\mu = \mathrm{E}[\hat{\theta}]$.

Expanding we can easily get:

$$\mathrm{E}\left[\left(\hat{\theta} - \mu + \mu - \theta\right)^2\right] = \int_0^L \left(\hat{\theta} - \mu + \mu - \theta\right)^2 f(x)dx$$
$$= \int_0^L \left(\hat{\theta} - \mu\right)^2 f(x)dx + \int_0^L (\mu - \theta)^2 f(x)dx + \int_0^L 2\left(\hat{\theta} - \mu\right)(\mu - \theta) f(x)dx$$
$$= \mathrm{E}\left[\left(\hat{\theta} - \mu\right)^2\right] + \mathrm{E}[(\mu - \theta)^2] + 2\mathrm{E}\left[\left(\hat{\theta} - \mu\right)(\mu - \theta)\right]$$

It is obviously that:

$$\mathrm{E}\left[\left(\hat{\theta} - \mu\right)^2\right] = Var\left(\hat{\theta}\right)$$

Because from the lecture note ProbabilityNotes.pdf, we can easily tell that:

$$\mathrm{E}[(\mu - \theta)^2] = (\mu - \theta)^2 = (\theta - \mu)^2 = \left(\theta - \mathrm{E}[\hat{\theta}]\right)^2$$

And,

$$\left(\theta - \mathrm{E}[\hat{\theta}]\right)^2 = bias\left(\hat{\theta}\right)^2$$

So,

$$\mathrm{E}[(\mu - \theta)^2] = bias\left(\hat{\theta}\right)^2$$

And since $\mathrm{E}\left[\left(\hat{\theta} - \mu\right)(\mu - \theta)\right] = (\mu - \theta)\mathrm{E}\left[\left(\hat{\theta} - \mu\right)\right] = 0$, we know that:

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{E}\left[\left(\hat{\theta} - \theta\right)^2\right] = bias\left(\hat{\theta}\right)^2 + Var\left(\hat{\theta}\right)$$

2)

In MOM, it is easy to show that, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{2}L$, so we can know:

$$\hat{L}_{MOM} = 2\bar{X}_n = L$$

So,

$$bias\left(\hat{L}\right) = L - E\left[\hat{L}\right] = L - 2E[\bar{X}_n] = L - \frac{2}{n}\sum_{i=1}^{n} E[X_i] = L - \frac{2}{n} \cdot n \cdot \frac{L}{2} = 0$$

Hence, we know $\hat{L}_{MOM}$ is unbiased.

In MLE, we can compute the density of X_i by:

$$f_{\hat{L}}(x) = \frac{d}{dx}P(X_i \le x) = \frac{d}{dx}\left[\left(\frac{x}{L}\right)^n\right] = \frac{nx^{n-1}}{L^n}$$

So we have:

$$E[\hat{L}] = \int_0^L x\, f_{\hat{L}}(x)\, dx = \frac{n}{L^n}\cdot\left[\frac{x^{n+1}}{n+1}\right]_0^L = \frac{nL}{n+1}$$

So,

$$bias(\hat{L}) = L - E[\hat{L}] = \frac{L}{n+1} > 0$$

Hence, we know $\hat{L}_{MLE}$ has bias.

In general, we know the bias of $\hat{L}_{MLE}$ is always greater than 0, which means we always have:

$$L > E[\hat{L}], n \to \infty$$

That is $\hat{L}_{MLE}$ consistently underestimates L.

3)

In MOM:

$$Var(\hat{L}_{MOM}) = Var(2\bar{X}_n) = 4Var(\bar{X}_n) = 4\cdot\frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = 4\cdot\frac{1}{n}Var(X) = 4\cdot\frac{L^2}{12n} = \frac{L^2}{3n}$$

In MLE, from 2). we have:

$$f_{\hat{L}}(x) = \frac{d}{dx}P(X_i \le x) = \frac{d}{dx}\left[\left(\frac{x}{L}\right)^n\right] = \frac{nx^{n-1}}{L^n}$$

$$E[\hat{L}] = \int_0^L x\, f_{\hat{L}}(x)\, dx = \frac{n}{L^n}\cdot\left[\frac{x^{n+1}}{n+1}\right]_0^L = \frac{nL}{n+1}$$

So we can calculate variance:

$$Var(\hat{L}_{MLE}) = \int_0^L x^2\, f_{\hat{L}}(x)\, dx - \left(E[\hat{L}]\right)^2 = \frac{n}{L^n}\cdot\left[\frac{x^{n+2}}{n+2}\right]_0^L - \left(E[\hat{L}]\right)^2 = \frac{nL^2}{n+2} - \left(\frac{nL}{n+1}\right)^2$$

$$= \frac{nL^2}{(n+1)^2(n+2)}$$

4)

In MOM:

$$\mathrm{MSE}(\hat{L}_{MOM}) = Var(\hat{L}_{MOM}) = \frac{L^2}{3n}$$

In MLE:

$$\mathrm{MSE}(\hat{L}_{MLE}) = bias(\hat{L}_{MOM})^2 + Var(\hat{L}_{MOM}) = \left(\frac{L}{n+1}\right)^2 + \frac{nL^2}{(n+1)^2(n+2)} = \frac{2L^2}{(n+1)(n+2)}$$

We can easy to tell that because:

$$\mathrm{MSE}(\hat{L}_{MOM}) - \mathrm{MSE}(\hat{L}_{MLE}) = \left(\frac{1}{3n} - \frac{2}{(n+1)(n+2)}\right)\cdot L^2 = \frac{n^2+3n+2-6n}{3n(n+1)(n+2)}\cdot L^2$$

$$= \frac{n^2-3n+2}{3n(n+1)(n+2)}\cdot L^2 = \frac{(n-1)(n-2)}{3n(n+1)(n+2)}\cdot L^2$$

And when $n \to \infty$, n > 0, $(n-1)(n-2) > 0$, $(n+1)(n+2) > 0$, so we know

$\text{MSE}(\hat{L}_{MOM}) - \text{MSE}(\hat{L}_{MLE}) > 0$. That is:

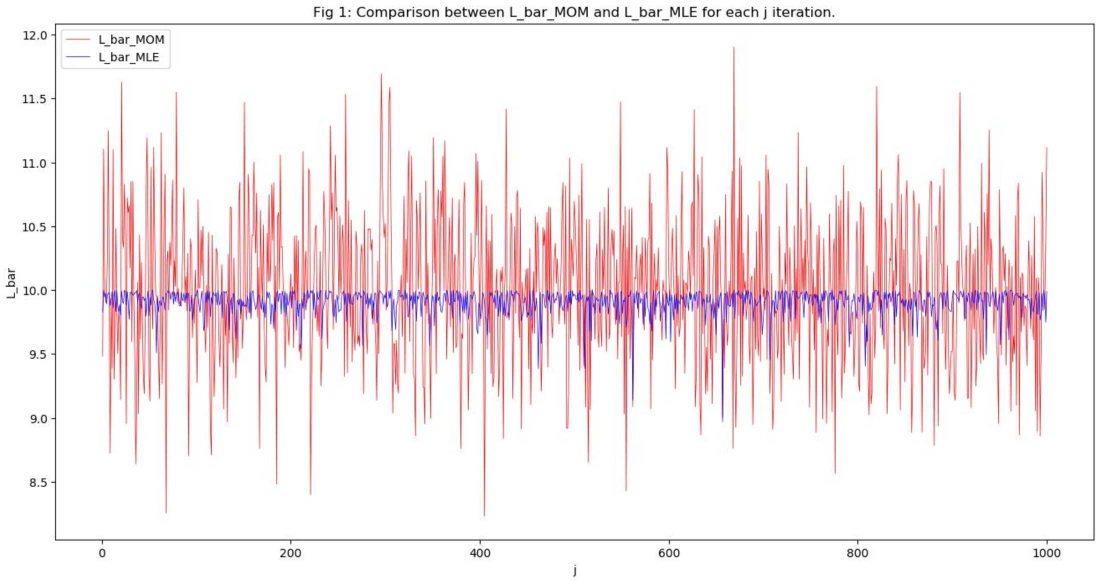
$$\text{MSE}(\hat{L}_{MOM}) > \text{MSE}(\hat{L}_{MLE})$$

$$\frac{\text{MSE}(\hat{L}_{MOM})}{\text{MSE}(\hat{L}_{MLE})} > 1, \text{if } \text{MSE}(\hat{L}_{MLE}) \neq 0$$

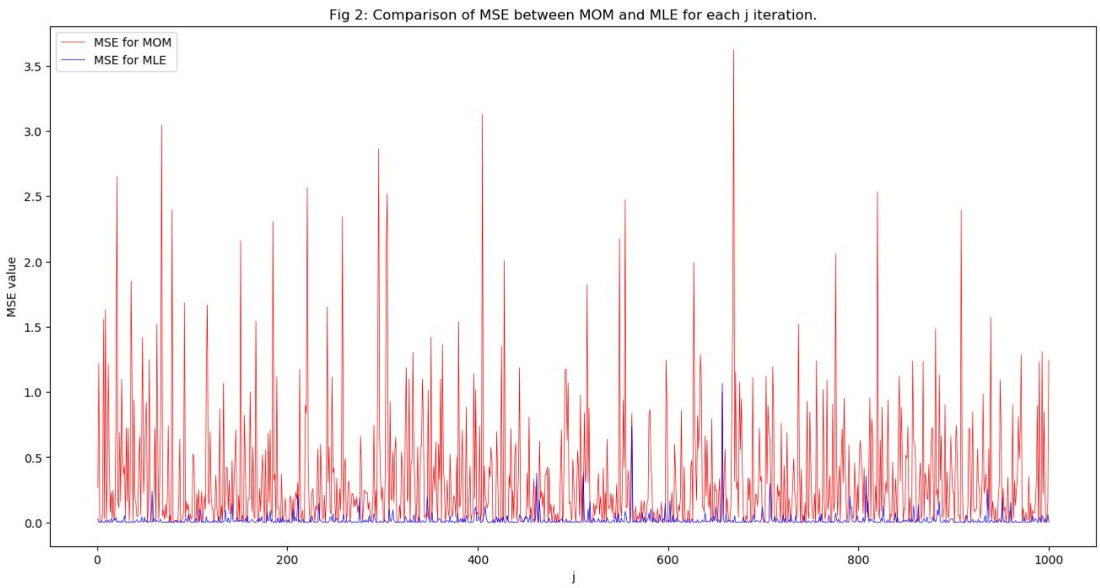
So, MLE is better. MLE is more efficient.

5)

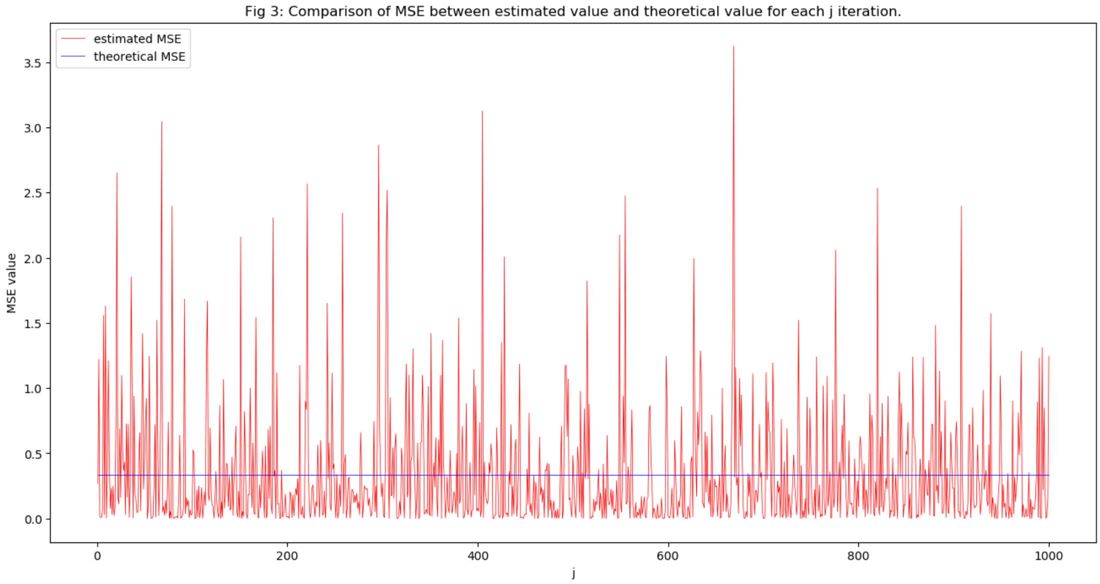
I use Python to program some scripts, so we can tell the different between data from pictures. The code and result data is in the attach.



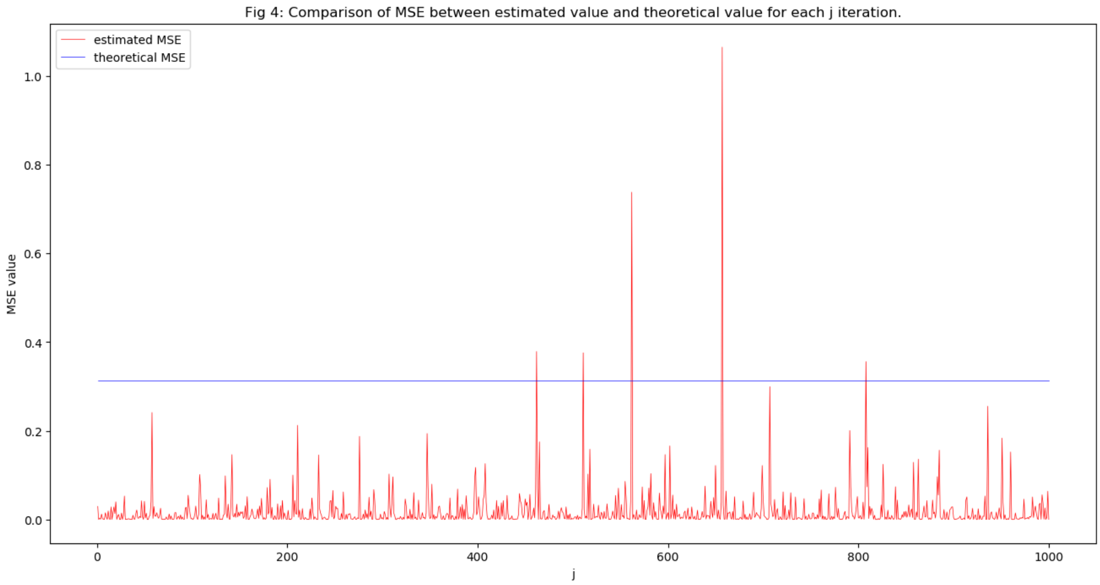
From Figure 1 above, we can know that: the estimated value for $\hat{L}_{MLE}$ is more precise than $\hat{L}_{MOM}$, they are closer to the true value than $\hat{L}_{MOM}$.



From Figure 2 above, we can know that: the estimated value for $\text{MSE}(\hat{L}_{MLE})$ is smaller than $\text{MSE}(\hat{L}_{MOM})$, which means the MLE is better estimated the true value of L than MOM because it has smaller error.



From Figure 3 above, we can know that: the estimated value for $\text{MSE}(\hat{L}_{MOM})$ is larger than the theoretical value of $\text{MSE}(\hat{L}_{MOM})$ we have calculated in previous questions.



From Figure 3 above, we can know that: the estimated value for $\text{MSE}(\hat{L}_{MLE})$ is smaller than the theoretical value of $\text{MSE}(\hat{L}_{MLE})$ we have calculated in previous questions, which is contrary to the result of Figure 3.

6)

"Unbiased" is often misunderstood to mean "superior". That is only true if an unbiased estimator has superior precision too. But biased estimators often have smaller overall error than unbiased ones. Clearly both criteria must be considered for an estimator to be judged superior to another.

Bias is the average difference between the estimator and the true value. Precision is the standard deviation of the estimator. One measure of the overall variability is the Mean Squared Error. The MSE is also the sum of the square of the precision and the square of the bias. Often the overall variability of a biased estimator is smaller than that for an unbiased estimator. Being unbiased isn't always a good thing if it also results in greater overall variability, which we talk about here is all known as **bias–variance tradeoff**.

The bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa.

In this question, we know that the MOM just considers the features of the data but ignore the distribution of data, which means it takes mean and variance of the data into consideration but ignores the overall distribution of the data. So, the MOM will have smaller precision, that is it will have larger variance for estimator if there are some extreme values, it will not "superior". Only if the data set is larger enough, we can have "superior" estimated.

While the MLE is to find the parameter, which has the maximum likelihood of occurrence, so it will make the samples in a tight distributed area. So, the MLE will have smaller variance of estimator than the MOM. Even the MLE has bias, the overall MSE of MLE will smaller than MSE of the MOM.

7)

From the above questions, we have already known the expect and variance of $\hat{L}_{MLE}$, and $\hat{L}_{MLE}$ is strictly positive since $\hat{L}_{MLE} \in [0, L]$. So, we can use the **Markov Inequality** that:

$$P(\hat{L}_{MLE} > L - \varepsilon) \leq \frac{E[\hat{L}_{MLE}]}{L - \varepsilon}$$

$$E[\hat{L}_{MLE}] = \int_0^L x\, f_{\hat{L}}(x)\, dx = \frac{n}{L^n} \cdot \left[\frac{x^{n+1}}{n+1}\right]_0^L = \frac{nL}{n+1}$$

$$Var(\hat{L}_{MLE}) = \int_0^L x^2\, f_{\hat{L}}(x)\, dx - \left(E[\hat{L}]\right)^2 = \frac{n}{L^n} \cdot \left[\frac{x^{n+2}}{n+2}\right]_0^L - \left(E[\hat{L}]\right)^2 = \frac{nL^2}{n+2} - \left(\frac{nL}{n+1}\right)^2$$

$$= \frac{nL^2}{(n+1)^2(n+2)}$$

$$P(\hat{L}_{MLE} < L - \varepsilon) \geq \delta$$

So,

$$P(\hat{L}_{MLE} \geq L - \varepsilon) \leq 1 - \delta$$

$$P(\hat{L}_{MLE} > L - \varepsilon) \leq \frac{E[\hat{L}_{MLE}]}{L - \varepsilon} = 1 - \delta$$

That is:

$$\frac{nL}{n+1} \cdot \frac{1}{L - \varepsilon} = 1 - \delta, \qquad 0 < \varepsilon < L, 0 < \delta < 1$$

Hence,

$$\frac{nL}{L - \varepsilon} = (1 - \delta)n + (1 - \delta)$$

$$\left[\frac{L}{L - \varepsilon} - (1 - \delta)\right] n = (1 - \delta)$$

$$\left(1+\frac{\varepsilon}{L-\varepsilon}-1+\delta\right)n=(1-\delta)$$

$$\frac{\delta L-\delta\varepsilon+\varepsilon}{L-\varepsilon}n=(1-\delta)$$

Therefore,

$$n=\frac{(1-\delta)(L-\varepsilon)}{\delta L+\varepsilon-\delta\varepsilon}$$

So, we need at least $O\left(\frac{(1-\delta)(L-\varepsilon)}{\delta L+\varepsilon-\delta\varepsilon}\right)$ many samples.

8)

In MLE, from 2). we have:

$$E\left[\hat{L}_{MLE}\right]=\int_0^L x\,f_{\hat{L}}(x)\,dx=\frac{n}{L^n}\cdot\left[\frac{x^{n+1}}{n+1}\right]_0^L=\frac{nL}{n+1}$$

Since now we have $\hat{L}=\left(\frac{n+1}{n}\right)\max_{i=1,\dots,n}X_i=\frac{n+1}{n}\hat{L}_{MLE}$, therefore we know that:

$$E\left[\hat{L}\right]=\frac{n+1}{n}\int_0^L x\,f_{\hat{L}}(x)\,dx=\frac{n+1}{n}\cdot\frac{n}{L^n}\cdot\left[\frac{x^{n+1}}{n+1}\right]_0^L=L$$

So,

$$bias\left(\hat{L}\right)=L-E\left[\hat{L}\right]=0$$

So, it is an unbiased estimator.

And variance of $\hat{L}_{MLE}$ is:

$$Var\left(\hat{L}_{MLE}\right)=\int_0^L x^2\,f_{\hat{L}}(x)\,dx-\left(E\left[\hat{L}\right]\right)^2=\frac{n}{L^n}\cdot\left[\frac{x^{n+2}}{n+2}\right]_0^L-\left(E\left[\hat{L}\right]\right)^2=\frac{nL^2}{n+2}-\left(\frac{nL}{n+1}\right)^2$$

$$=\frac{nL^2}{(n+1)^2(n+2)}$$

So we can calculate variance:

$$Var\left(\hat{L}\right)=Var\left(\left(\frac{n+1}{n}\right)\hat{L}_{MLE}\right)=\left(\frac{n+1}{n}\right)^2Var\left(\hat{L}_{MLE}\right)=\left(\frac{n+1}{n}\right)^2\cdot\frac{nL^2}{(n+1)^2(n+2)}$$

$$=\frac{L^2}{n(n+2)}$$

So,

$$\mathrm{MSE}\left(\hat{L}\right)=Var\left(\hat{L}\right)=\frac{L^2}{n(n+2)}$$

Compare to the MOM, we can have that:

$$\lim_{n\to\infty}\left(\frac{\mathrm{MSE}\left(\hat{L}\right)}{\mathrm{MSE}\left(\hat{L}_{MOM}\right)}\right)=\lim_{n\to\infty}\left(\frac{\frac{L^2}{n(n+2)}}{\frac{L^2}{3n}}\right)=\lim_{n\to\infty}\left(\frac{3}{n+2}\right)\to 0$$

So,

$$\lim_{n\to\infty}\left(\frac{\mathrm{MSE}\left(\hat{L}\right)}{\mathrm{MSE}\left(\hat{L}_{MOM}\right)}\right)<1$$

Which means $\mathrm{MSE}\left(\hat{L}\right)<\mathrm{MSE}\left(\hat{L}_{MOM}\right)$. We can still have a smaller MSE.