Shengjie Li
RUID: 188008047

CS 536 : Estimation Problems

Uniform Estimators

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables, uniformly distributed on $[0, L]$ (i.e., with density $1/L$ on this interval). In the posted notes on estimation, it is shown that the method of moments and maximum likelihood estimators for L are given by

$$\begin{aligned} \hat{L}_{\text{MOM}} &= 2\overline{X}_n \\ \hat{L}_{\text{MLE}} &= \max_{i=1,\ldots,n} X_i. \end{aligned} \tag{1}$$

We want to consider the question of which estimator is better. Recall the definition of the mean squared error of an estimator as

$$\text{MSE}(\hat{L}) = \mathbb{E}[(\hat{L} - L)^2]. \tag{2}$$

Note: the answers to homework zero may also be useful here.

1) Show that in general, $\text{MSE}(\hat{\theta}) = \text{bias}(\hat{\theta})^2 + \text{var}(\hat{\theta})$, where var is the variance, and bias is given by

$$\text{bias}(\hat{\theta}) = \theta - \mathbb{E}[\hat{\theta}]. \tag{3}$$

Solution:

$$\begin{aligned} \text{MSE}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta} - \theta)^2] \\ &= \mathbb{E}[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2] \\ &= \mathbb{E}[\hat{\theta}^2] - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2 \\ &= \mathbb{E}[\hat{\theta}^2] - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2 + \mathbb{E}[\hat{\theta}]^2 - \mathbb{E}[\hat{\theta}]^2 \\ &= (\theta - \mathbb{E}[\hat{\theta}])^2 + (\mathbb{E}[\hat{\theta}^2] - \mathbb{E}[\hat{\theta}]^2) \\ &= \text{bias}(\hat{\theta})^2 + \text{var}(\hat{\theta}) \end{aligned}$$

2) Show that $\hat{L}_{\text{MOM}}$ is *unbiased*, but that $\hat{L}_{\text{MLE}}$ has bias. In general, $\hat{L}_{\text{MLE}}$ consistently underestimates L - why?

Solution:

$$\begin{aligned}
\text{bias}(\hat{L}_{\text{MOM}}) &= L - \mathbb{E}[\hat{L}_{\text{MOM}}] \\
&= L - \mathbb{E}[2\overline{X}_n] \\
&= L - 2\mathbb{E}[\overline{X}_n] \\
&= L - 2\mathbb{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] \\
&= L - \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] \\
&= L - \frac{2}{n} n\mathbb{E}[X] \\
&= L - 2\mathbb{E}[X] \\
&= 0
\end{aligned}$$

Thus $\hat{L}_{\text{MOM}}$ is unbiased.

The cdf of MLE:

$$F(x) = \mathbb{P}(\max_{i=1,\ldots,n} X_i \leq x) = \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x) = \mathbb{P}(X_1 \leq x)\ldots\mathbb{P}(X_n \leq x) = (\frac{x}{L})^n.$$

Thus, the pdf of MLE is:

$$f(x) = F'(x) = \frac{n \cdot x^{n-1}}{L^n}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}[\hat{L}_{\text{MLE}}] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_0^L x \frac{n \cdot x^{n-1}}{L^n} dx \\
&= \int_0^L n(\frac{x}{L})^n dx \\
&= \frac{n}{n+1} \cdot \frac{x^{n+1}}{L^n}\Big|_0^L \\
&= \frac{n}{n+1} \cdot \frac{L^{n+1}}{L^n} \\
&= \frac{n}{n+1} L.
\end{aligned}$$

We have

$$\text{bias}(\hat{L}_{\text{MLE}}) = L - \mathbb{E}[\hat{L}_{\text{MLE}}] = L - \frac{n}{n+1}L = \frac{1}{n+1}L \neq 0.$$

Thus, $\hat{L}_{\text{MLE}}$ has bias.

$$\mathbb{P}(\hat{L}_{\text{MLE}} \geq L) = \mathbb{P}(\max_{i=1,\ldots,n} X_i \geq L) = 0.$$

Thus, $\hat{L}_{\text{MLE}}$ consistently underestimates L.

3) Compute the variance of $\hat{L}_{\text{MOM}}$ and $\hat{L}_{\text{MLE}}$.

Solution:

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}[(X-\mathbb{E}(X))^2] = \mathbb{E}[(X-\frac{X}{2})^2] = \frac{1}{4}\mathbb{E}[X^2] = \frac{1}{4}\int_{-\infty}^{\infty} x^2 f(x)dx = \frac{1}{4}\int_0^L x^2\frac{1}{L}dx \\
&= \frac{1}{4}\frac{1}{3}\frac{1}{L}x^3\Big|_0^L = \frac{1}{12L}L^3 = \frac{L^2}{12} \\
\text{Var}(\hat{L}_{\text{MOM}}) &= \text{Var}(2\overline{X}_n) = 4\text{Var}(\overline{X}_n) = 4\text{Var}(\frac{1}{n}\sum_{i=1}^n X_i) = \frac{4}{n^2}\sum_{i=1}^n \text{Var}(X_i) = \frac{4}{n^2}n\text{Var}(X) = \frac{L^2}{3n} \\
\text{Var}(\hat{L}_{\text{MLE}}) &= \mathbb{E}[\hat{L}_{\text{MLE}}^2] - \mathbb{E}[\hat{L}_{\text{MLE}}]^2 \\
&= \int_{-\infty}^{\infty} x^2 f(x)dx - (\frac{n}{n+1}L)^2 \\
&= \int_0^L x^2\frac{n\cdot x^{n-1}}{L^n}dx - \frac{n^2L^2}{(n+1)^2} \\
&= n\int_0^L \frac{x^{n+1}}{L^n}dx - \frac{n^2L^2}{(n+1)^2} \\
&= \frac{n}{n+2}\frac{x^{n+2}}{L^n}\Big|_0^L - \frac{n^2L^2}{(n+1)^2} \\
&= \frac{nL^2}{n+2} - \frac{n^2L^2}{(n+1)^2} \\
&= \frac{nL^2((n+1)^2 - n(n+2))}{(n+2)(n+1)^2} \\
&= \frac{nL^2}{(n+2)(n+1)^2}
\end{aligned}$$

4) Which one is the better estimator, i.e., which one has the smaller mean squared error?

Solution:

$$\begin{aligned}
\text{MSE}(\hat{L}_{\text{MOM}}) &= \text{bias}(\hat{L}_{\text{MOM}})^2 + \text{var}(\hat{L}_{\text{MOM}}) = \frac{L^2}{3n} \\
\text{MSE}(\hat{L}_{\text{MLE}}) &= \text{bias}(\hat{L}_{\text{MLE}})^2 + \text{var}(\hat{L}_{\text{MLE}}) = (\frac{1}{n+1}L)^2 + \frac{nL^2}{(n+2)(n+1)^2} = \frac{2L^2}{(n+2)(n+1)}
\end{aligned}$$

Let $\text{MSE}(\hat{L}_{\text{MLE}}) \leq \text{MSE}(\hat{L}_{\text{MOM}})$, solve $\frac{2L^2}{(n+2)(n+1)} \leq \frac{L^2}{3n}$, we get $n \leq 1$ or $n \geq 2$.

Thus, MLE always has the smaller MSE.

5) Experimentally verify your computations in the following way: Taking $n = 100$ and $L = 10$,

- For $j = 1, \ldots, 1000$:
- Simulate $X_1^j, \ldots, X_n^j$ and compute values for $\hat{L}^j_{\mathrm{MOM}}$ and $\hat{L}^j_{\mathrm{MLE}}$
- For $n = 100, L = 10$, simulate $X_1, \ldots, X_n$, and compute values for $\hat{L}_{\mathrm{MOM}}$ and $\hat{L}_{\mathrm{MLE}}$.
- Estimate the mean squared error for each population of estimator values.
- How do these estimated MSEs compare to your theoretical MSEs?

Solution:

```
Estimated MSEs          Theoretival MSEs
0.334586                0.333333
0.018118                0.019414
L^MOM:10.346003012988131
L^MLE:9.9819040666619774
```

Source code:

```
import numpy as np

n = 100
L = 10

LJMOM_list = []
LJMLE_list = []
for j in range(0, 1000):
    X = np.random.uniform(0, L, n)
    mu = np.average(X)
    LJMOM = 2 * mu
    LJMLE = np.max(X)
    LJMOM_list.append(LJMOM)
    LJMLE_list.append(LJMLE)
mu_MOM = np.average(LJMOM_list)
mu_MLE = np.average(LJMLE_list)
var_MOM = np.average(np.array(LJMOM_list) ** 2) - mu_MOM ** 2
var_MLE = np.average(np.array(LJMLE_list) ** 2) - mu_MLE ** 2
biasMOM = L - mu_MOM
biasMLE = L - mu_MLE
MSEMOM = biasMOM ** 2 + var_MOM
MSEMLE = biasMLE ** 2 + var_MLE

X = np.random.uniform(0, L, n)
mu = np.average(X)
LMOM = 2 * mu
```

```
LMLE = np.max(X)
print('{:<20s} {:<20s}'.format('Estimated MSEs', 'Theoretival MSEs'))
print('{:<20f} {:<20f}'.format(MSEMOM, 1.0 * L * L / 3 / n))
print('{:<20f} {:<20f}'.format(MSEMLE, 2.0 * L * L / (n + 2) / (n + 1)))
print('L^MOM:' + repr(LMOM))
print('L^MLE:' + repr(LMLE))
```

6) You should have shown that $\hat{L}_{\text{MLE}}$, while biased, has a smaller error over all. Why? The mathematical justification for it is above, but is there an explanation for this?

Solution:

Because we prefer smaller $\hat{L}$ than larger $\hat{L}$ (The same as what we would do if given two estimations 10 and 1000), and $\hat{L}_{MLE}$ constantly underestimates L, so $\hat{L}_{MLE}$ performs well in this circumstance.

7) Find $\mathbb{P}(\hat{L}_{\text{MLE}} < L - \epsilon)$ as a function of L, ϵ, n. Estimate how many samples I would need to be sure that my estimate was within ϵ with probability at least δ.

Solution:

$$\mathbb{P}(\hat{L}_{\text{MLE}} < L - \epsilon) = F(L - \epsilon) = (\frac{L - \epsilon}{L})^n$$

The probability of my estimate was within ϵ is:

$$\mathbb{P}(\hat{L}_{\text{MLE}} \geq L - \epsilon) = 1 - \mathbb{P}(\hat{L}_{\text{MLE}} < L - \epsilon). = 1 - (\frac{L - \epsilon}{L})^n$$

Thus, the number of samples needed to make sure the probability is at least δ is:

$$\begin{aligned}
\mathbb{P}(\hat{L}_{\text{MLE}} \geq L - \epsilon) \geq \delta &\to 1 - (\frac{L - \epsilon}{L})^n \geq \delta \\
&\to 1 - \delta \geq (\frac{L - \epsilon}{L})^n \\
&\to \ln(1 - \delta) \geq n \ln(\frac{L - \epsilon}{L}) \\
&\to n \geq \frac{\ln(1 - \delta)}{\ln(\frac{L - \epsilon}{L})}
\end{aligned}$$

8) Show that

$$\hat{L} = (\frac{n + 1}{n}) \max_{i=1,\ldots,n} X_i, \tag{4}$$

is an unbiased estimator, and has a smaller MSE still.

Solution:

$$\begin{aligned}
\mathbb{E}[\hat{L}] &= \mathbb{E}[(\frac{n+1}{n})\max_{i=1,\dots,n} X_i] \\
&= (\frac{n+1}{n})\mathbb{E}[\max_{i=1,\dots,n} X_i] \\
&= (\frac{n+1}{n})\mathbb{E}[\hat{L}_{\mathrm{MLE}}] \\
&= (\frac{n+1}{n})(\frac{n}{n+1})L \\
&= L
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\hat{L}) &= \mathrm{Var}((\frac{n+1}{n})\max_{i=1,\dots,n} X_i) \\
&= (\frac{n+1}{n})^2\mathrm{Var}(\max_{i=1,\dots,n} X_i) \\
&= (\frac{n+1}{n})^2\mathrm{Var}(\hat{L}_{\mathrm{MLE}}) \\
&= (\frac{n+1}{n})^2\frac{nL^2}{(n+2)(n+1)^2} \\
&= \frac{L^2}{n(n+2)}
\end{aligned}$$

$$\begin{aligned}
\mathrm{bias}(\hat{L}) &= L - \mathbb{E}[\hat{L}] \\
&= 0 \\
\mathrm{MSE}(\hat{L}) &= \mathrm{bias}(\hat{L})^2 + \mathrm{Var}(\hat{L}) \\
&= 0 + \frac{L^2}{n(n+2)} \\
&= \frac{L^2}{n(n+2)}
\end{aligned}$$

Thus, $\hat{L}$ is an unbiased estimator.

Let $\mathrm{MSE}(\hat{L}) \leq \mathrm{MSE}(\hat{L}_{\mathrm{MLE}})$, solve $\frac{L^2}{n(n+2)} \leq \frac{2L^2}{(n+2)(n+1)}$, we get $n \geq 1$.

Thus, $\hat{L}$ has a small MSE.