NetID: sj735 Machine-Learning Assig^n 1 [Uniform Estimators]

1.)

Assuming here $\hat{\theta} = \hat{L}$, $\theta = L$

We have,
$$MSE = E[(\hat{\theta}-\theta)^2]$$
$$= E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2]$$
$$= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2$$
$$= E[\hat{\theta}^2] + \theta^2 - 2\theta E[\hat{\theta}] + E[\hat{\theta}]^2 - E[\hat{\theta}]^2$$
$$= \left(E[\hat{\theta}^2] - E[\hat{\theta}]^2\right) + \left(\theta - E[\hat{\theta}]\right)^2$$

Given, $Bias(\hat{\theta}) = \theta - E[\hat{\theta}]$, We know that $Var(\hat{\theta}) = E[\hat{\theta}^2] - E[\hat{\theta}]^2$

$$\boxed{MSE = Bias(\hat{\theta})^2 + Var(\hat{\theta})}$$

3.) <u>Calculate Variance</u> :

* $\hat{L}_{mle}$: $\max_{i=1\ldots n} X_i$

First we need to calculate ~~density~~ CDF, Y:

$$P(Y \le x) = P(\max_i X_i \le x) = P(X_1 \le x, X_2 \le x, \ldots X_n \le x)$$

Since these are iid

$$P(Y \le x) = P(X_1 \le x) . P(X_2 \le x) \ldots\ldots P(X_n \le x)$$

$$= P(X \le x)^n \quad — (1)$$

X is uniform distribution;

$\therefore$ density$_X$ $= \frac{1}{L}$ over the interval $[0, L]$

$\therefore$ above probability is simply $\left(\frac{x}{L}\right)^n$

$$\therefore P(Y \le x) = \left(\frac{x}{L}\right)^n \quad — (2)$$

Taking derivative of (2), we will get density of CDF

$$f(x) = n . \frac{1}{L}\left(\frac{x}{L}\right)^{n-1}$$

Hence $$\boxed{f(x) = \begin{cases} 0 & \text{if } x \le 0 \text{ or } x \ge L \\ \frac{n}{L}\left(\frac{x}{L}\right)^{n-1} & \text{if } 0 < x < L \end{cases}}$$

Scanned with CamScanner

We can now compute mean & variance of MLE

for convenience

$\hat{L} = \hat{\theta}$

$L = \theta$

$$E[\hat{\theta}_{MLE}] = \int_0^{\theta} x f(x)\,dx = \frac{n}{\theta^n}\int_0^{\theta} x^n\,dx = \frac{n}{n+1}\cdot\theta \quad \text{i.e } \underline{\underline{\frac{nL}{n+1}}}$$

$$E[\hat{\theta}^2_{MLE}] = \int_0^{\theta} x^2 f(x)\,dx = \frac{n}{\theta^n}\int_0^{\theta} x^{n+1}\,dx = \frac{n}{n+2}\theta^2 \quad \text{i.e } \frac{nL^2}{n+2}$$

$$\therefore Var[\hat{\theta}_{MLE}] = E[\hat{\theta}^2_{MLE}] - E[\hat{\theta}_{MLE}]^2 = \underline{\underline{\frac{nL^2}{(n+2)(n+1)^2}}}$$

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

PDF for uniformly distributed $[0, L]$ will be given as $\frac{1}{L}$

$$\therefore \text{Mean, } E[\hat{L}_{mom}] = \int_0^{L} x f(x) = \frac{1}{L}\int_0^{L} x\,dx = \underline{\underline{\frac{L}{2}}}$$

$$E[\hat{L}^2_{mom}] = \int_0^{L} x^2 f(x) = \frac{1}{L}\cdot\frac{x^3}{3}\Big|_0^L = \underline{\underline{\frac{L^2}{3}}}$$

$$\therefore Var[\hat{L}_{mom}] = E[\hat{L}^2_{mom}] - E[\hat{L}_{mom}]^2$$

$$= \frac{L^2}{3} - \left(\frac{L}{2}\right)^2 = \underline{\underline{\frac{L^2}{12}}}$$

$\therefore$ ~~Var~~ Here $\hat{L}_{mom} = 2\bar{X}_n$

$$\therefore Var[\hat{L}_{mom}] = \frac{4}{n}\left[\frac{L^2}{12}\right]$$

$$= \frac{L^2}{3n}$$

Scanned with CamScanner

2.)

Continuing our observations which were calculated in solution ③,

We are going to calculate bias for $\hat{L}_{MOM}$ & $\hat{L}_{MLE}$

→ $$\text{bias}(\hat{L}_{mom}) = L - E[\hat{L}_{mom}]$$ [Given $\hat{L}_{mom} = 2\bar{X}_n$]

$$= L - E[2\bar{X}_n]$$

$$= L - 2E[\bar{X}_n]$$

$$= L - 2\left(\frac{L}{2}\right) = \underline{\underline{0}}$$

∴ $\hat{L}_{mom}$ is unbiased

→ $$\text{bias}(\hat{L}_{MLE}) = L - E[\hat{L}_{MLE}]$$

$$= L - \frac{nL}{n+1}$$

$$= \frac{nL + L - nL}{n+1} = \frac{L}{n+1}$$

∴ $\hat{L}_{MLE}$ is unbiased

Since, the factor $\left(\frac{1}{n+1}\right)$ is coming while calculating the bias for $\hat{L}_{MLE}$,

It consistently underestimates L

H.P.

4.)

Mean Square Error, MSE is given by

$$MSE = Bias(\hat{L})^2 + var(\hat{L})$$

$$\therefore MSE_{MLE} = \left[Bias(\hat{L}_{MLE})\right]^2 + var(\hat{L}_{MLE})$$

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

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

And. $$MSE_{MOM} = \left[Bias(\hat{L}_{MOM})\right]^2 + var(\hat{L}_{MOM})$$

$$= 0 + \frac{L^2}{3n}$$

$$= \frac{L^2}{3n}$$

MLE has a higher bias, however its variance is significantly lower than the variance of MME (of order $O\left(\frac{1}{n^2}\right)$ against $O\left(\frac{1}{n}\right)$).

$\Rightarrow$ MSE is significantly improved in the case of MLE (also going from $O\left(\frac{1}{n}\right)$ to $O\left(\frac{1}{n^2}\right)$: by doing is little bit of trade-off

Scanned with CamScanner

with bias, we greatly decreased the MSE.

$\therefore$ ~~$\hat{L}_{MLE}$~~ $\therefore$ MLE is a better estimator than MOM

$O\left(\frac{1}{n^2}\right) \longrightarrow O\left(\frac{1}{n}\right)$

(5.)

Please check the attached R-code file for simulation.

- We have already calculated the Theoritical MSE's of both MLE & MOM

i.e

$$MSE_{MLE} = \frac{2L^2}{(n+2)(n+1)} \quad \& \quad MSE_{mom} = \frac{L^2}{3n}$$

for $n=100$, $L=10$

$$MSE_{MLE} = \frac{2(10)^2}{102 * 101} \quad \& \quad MSE_{mom} = \frac{(10)^2}{3*100}$$

$$= \frac{200}{102 * 101} \qquad = \frac{100}{300}$$

$$= 0.019413 \qquad = 0.333$$

$\therefore$ We can make a conclusion that, MSE_{MLE} is less than MSE_{mom} theoritically, which is what we observed while calculating it programaticalty.

6.

While calculating the value of MSE for $\hat{l}_{MLE}$, we have seen that

MLE has a higher bias, however its variance is significantly lower than the variance of MOM.

i.e $Var_{MLE} = O\left(\frac{1}{n^2}\right)$

$Var_{MOM} = O\left(\frac{1}{n}\right)$

Although, the bias for MOM is zero while the bias $(\hat{l}_{MLE})$ is $O\left(\frac{1}{n}\right)$

For higher values of n, the MSE for $\hat{l}_{MLE}$ is consistently lower than MSE for $\hat{l}_{mom}$.

i.e

By doing a little bit of trade-off with "bias", we greatly decreased the MSE for MLE

(7.) We are here trying to find $P(\hat{L}_{MLE} < L - \epsilon)$ as a function of L, ϵ, n and . we Know that,

$$P(\hat{L}_{MLE} < L - \epsilon) = P(\max_{i=1..n} X_i < L - \epsilon)$$

Also. $$P(Y < x) = \left(\frac{x}{L}\right)^n$$

$$\therefore P(\hat{L}_{MLE} < L - \epsilon) = P(\max_{i=1..n} X_i < L - \epsilon)$$
$$= \left(\frac{L-\epsilon}{L}\right)^n$$

$$\therefore P(\hat{L}_{MLE} < L - \epsilon) = \left(1 - \frac{\epsilon}{L}\right)^n \quad \text{--- (1)}$$

$$\Rightarrow P(L - \hat{L}_{MLE} < \epsilon) < \delta$$
$$\Rightarrow 1 - P(L - \hat{L}_{MLE} > \epsilon) < \delta$$
$$\Rightarrow P(L - \hat{L}_{MLE} > \epsilon) < 1 - \delta \quad \text{--- (2)}$$

from (1) & (2)

$$P(L - \hat{L}_{MLE} > \epsilon) < 1 - \delta$$
$$\left(\frac{L-\epsilon}{L}\right)^n < 1 - \delta$$

$$n \ln\left(1-\frac{\epsilon}{L}\right) < \ln(1-\delta)$$

$$\Rightarrow \cancel{\boxed{n > \frac{\ln\left(\frac{1}{1-\delta}\right)}{L}}}$$

Since $1-\frac{\epsilon}{L}$ is negative value

$$-n \ln\left(\frac{L}{L-\epsilon}\right) < \ln(1-\delta)$$

$$\Rightarrow -n < \frac{\ln(1-\delta)}{\ln\left(\frac{L}{L-\epsilon}\right)}$$

$$\Rightarrow \boxed{n > \frac{\ln\left(\frac{1}{1-\delta}\right)}{\ln\left(\frac{L}{L-\epsilon}\right)}}$$

⑧

From the previous solutions, we have calculated the

$$E[\hat{L}_{MLE}] = \frac{nL}{n+1}$$

i.e $E(\max. X_i) = \frac{n}{n+1}L$ —— ①

Now, Let multiply the eq^n ① by $\left(\frac{n+1}{n}\right)$

we will get

$$E[\hat{L}] = E\left[\frac{n+1}{n} \cdot \max X_i\right]$$

$$= \frac{n+1}{n} \times \frac{n \cdot L}{n+1}$$

$$= L$$

$\therefore$ Bias $(\hat{L}) = L - E[\hat{L}]$
$= L - L = 0$

$\therefore$ $\hat{L}$ is an unbiased estimator.

Since, the bias $(\hat{L})$ is zero.

and $\because$ $MSE(\hat{L}) = (Bias)^2 + Var(\hat{L})$

$\therefore$ $MSE(\hat{L})$ will be of order of $O\left(\frac{1}{n}\right)$ which is a smaller MSE still.