Question 1

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

$$bias(\hat{\theta}) = \theta - \mathbb{E}\,[\hat{\theta}]$$

Answer 1

	$var(X) = \mathbb{E}\,[X^2] - \mathbb{E}\,[X]^2$	(known)
or,	$var(\hat{\theta}) = \mathbb{E}\,[\hat{\theta}^2] - \mathbb{E}\,[\hat{\theta}]^2$	(1)

	$bias(\hat{\theta}) = \theta - \mathbb{E}[\hat{\theta}]$	(given)
or,	$bias(\hat{\theta})^2 = (\theta - \mathbb{E}[\hat{\theta}])^2$	(squaring both sides)
or,	$bias(\hat{\theta})^2 = \theta^2 - 2\theta\mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}]^2$	(2)

	$MSE(\hat{L}) = \mathbb{E}\,\left[(\hat{L} - L)^2\right]$	(given)
or,	$MSE(\hat{\theta}) = \mathbb{E}\,\left[(\hat{\theta} - \theta)^2\right]$	
or,	$MSE(\hat{\theta}) = \mathbb{E}\,[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2]$	
or,	$MSE(\hat{\theta}) = \mathbb{E}\,[\hat{\theta}^2] - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2$	
or,	$MSE(\hat{\theta}) = \theta^2 - 2\theta\mathbb{E}[\hat{\theta}] + \mathbb{E}\,[\hat{\theta}^2]$	(rearranging)
or,	$MSE(\hat{\theta}) = \left(\theta^2 - 2\theta\mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}]^2\right) + \left(\mathbb{E}\,[\hat{\theta}^2] - \mathbb{E}[\hat{\theta}]^2\right)$	(adding and subtracting $\mathbb{E}[\hat{\theta}]^2$)
or,	$MSE(\hat{\theta}) = bias(\hat{\theta})^2 + var(\hat{\theta})$	(from 2 and 1)

Question 2

Compute the bias of $\hat{L}_{MOM}$ and $\hat{L}_{MLE}$. In general, $\hat{L}_{MLE}$ consistently underestimates L - why? *Hint: What is the pdf for $\hat{L}_{MLE}$?*

Answer 2

	$\hat{L}_{MOM} = 2\bar{X}_n$	(given)
and,	$bias(\hat{\theta}) = \theta - \mathbb{E}[\hat{\theta}]$	(given)
Therefore,	$bias(\hat{L}_{MOM}) = L - \mathbb{E}[2\hat{L}]$	
	$bias(\hat{L}_{MOM}) = L - 2\mathbb{E}[\hat{L}]$	
	$bias(\hat{L}_{MOM}) = L - 2\frac{L}{2}$	(from Estimation notes 13)
	$bias(\hat{L}_{MOM}) = 0$	

Therefore, $\hat{L}_{MOM}$ is unbiased.

$$\hat{L}_{MLE} = \max_{i=1,\dots,n} X_i \qquad \text{(given)}$$

The c.d.f. is:

$$P(\hat{L}_{MLE} \le x) = P\left(\max_{i=1,\dots,n} X_i\right)$$

or, $F(x) = P(X_1 \le x, X_2 \le x, \dots, X_n \le x)$

or, $F(x) = P(X_1 \le x, X_2 \le x, \dots, X_n \le x)$

or, $F(x) = P(X_1 \le x)\, P(X_2 \le x)\, \dots P(X_n \le x)$

or, $F(x) = P(X \le x)^n$

or, $F(x) = \left(\frac{x}{L}\right)^n$ (1)

Calculating the p.d.f.:

$$f(x) = \frac{d}{dx} F(x)$$

or, $f(x) = \frac{d}{dx}\left(\frac{x}{L}\right)^n$ (from 1)

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

or, $f(x) = \frac{n}{L^n} x^{n-1}$ (2)

Calculating the expectation:

$$\mathbb{E}[\hat{L}_{MLE}] = \int_0^L x f(x) dx$$

or, $\mathbb{E}[\hat{L}_{MLE}] = \int_0^L x \frac{n}{L^n} x^{n-1} dx$ (from 2)

or, $\mathbb{E}[\hat{L}_{MLE}] = \int_0^L \frac{n}{L^n} x^n dx$

or, $\mathbb{E}[\hat{L}_{MLE}] = \frac{n}{L^n} \int_0^L x^n dx$

or, $\mathbb{E}[\hat{L}_{MLE}] = \frac{n}{L^n} \left| \frac{x^{n+1}}{n+1} \right|_0^L$

or, $\mathbb{E}[\hat{L}_{MLE}] = \frac{n}{L^n} \frac{L^{n+1}}{n+1}$

or, $\mathbb{E}[\hat{L}_{MLE}] = \frac{nL}{n+1}$ (3)

$$bias(\hat{\theta}) = \theta - \mathbb{E}[\hat{\theta}] \qquad \text{(given)}$$

or, $bias(\hat{L}_{MLE}) = L - \mathbb{E}[\hat{L}_{MLE}]$

or, $bias(\hat{L}_{MLE}) = L - \frac{nL}{n+1}$ (from 3)

or, $bias(\hat{L}_{MLE}) = \frac{nL + L - nL}{n+1}$

or, $bias(\hat{L}_{MLE}) = \frac{L}{n+1}$

Therefore, $\hat{L}_{MLE}$ is biased.

The denominator is always more than 1 (actually more than 2) because n ≥ 1, $\hat{L}_{MLE}$ will always be less than L. Therefore $\hat{L}_{MLE}$ consistently underestimates L.

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

Answer 3

$\hat{L}_{MOM} = 2\bar{X}_n$ (given)
or, $var(\hat{L}_{MOM}) = var(2\bar{X}_n)$
or, $var(\hat{L}_{MOM}) = (2^2)var(\bar{X}_n)$
or, $var(\hat{L}_{MOM}) = 4var(\bar{X}_n)$
or, $var(\hat{L}_{MOM}) = 4var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$
or, $var(\hat{L}_{MOM}) = \frac{4}{n^2}var(\sum_{i=1}^{n} X_i)$
or, $var(\hat{L}_{MOM}) = \left(\frac{4}{n^2}\right)\left(n\frac{(L-0)^2}{12}\right)$ (uniform distribution on the interval *[0, L]*)
or, $var(\hat{L}_{MOM}) = \left(\frac{1}{n}\right)\left(\frac{L^2}{3}\right)$
or, $var(\hat{L}_{MOM}) = \frac{L^2}{3n}$

$\mathbb{E}\left[\hat{L}_{MLE}{}^2\right] = \int_0^L x^2 f(x)dx$
or, $\mathbb{E}\left[\hat{L}_{MLE}{}^2\right] = \int_0^L x^2 \frac{n}{L^n} x^{n-1}dx$ (from Answer 2 … 2)
or, $\mathbb{E}\left[\hat{L}_{MLE}{}^2\right] = \frac{n}{L^n}\int_0^L x^{n+1}dx$
or, $\mathbb{E}\left[\hat{L}_{MLE}{}^2\right] = \frac{n}{L^n} | \frac{x^{n+2}}{n+2} |_0^L$
or, $\mathbb{E}\left[\hat{L}_{MLE}{}^2\right] = \frac{n}{L^n}\frac{L^{n+2}}{n+2}$
or, $\mathbb{E}\left[\hat{L}_{MLE}{}^2\right] = \frac{nL^2}{n+2}$ (1)

$\mathbb{E}[\hat{L}_{MLE}] = \frac{nL}{n+1}$ (from Answer 2 … 3)
or, $\mathbb{E}[\hat{L}_{MLE}]^{\mathbf{2}} = \left(\frac{nL}{n+1}\right)^2$ (2)

$var(X) = \mathbb{E}\,[X^2] - \mathbb{E}\,[X]^2$ (known)
or, $var(\hat{L}_{MLE}) = \mathbb{E}\left[\hat{L}_{MLE}{}^2\right] - \mathbb{E}\left[\hat{L}_{MLE}\right]^2$
or, $var(\hat{L}_{MLE}) = \frac{nL^2}{n+2} - \left(\frac{nL}{n+1}\right)^2$ (from 1 and 2)
or, $var(\hat{L}_{MLE}) = \frac{(nL^2)(n+1)^2-(nL)^2(n+2)}{(n+2)(n+1)^2}$
or, $var(\hat{L}_{MLE}) = \frac{(nL^2)(n^2+2n+1)-(nL)^2(n+2)}{(n+2)(n+1)^2}$
or, $var(\hat{L}_{MLE}) = \frac{(n^3L^2+2n^2L^2+nL^2)-(n^3L^2+(2n^2L^2)}{(n+2)(n+1)^2}$
or, $var(\hat{L}_{MLE}) = \frac{nL^2}{(n+2)(n+1)^2}$ (3)

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

Answer 4

$$MSE(\hat{\theta}) = bias(\hat{\theta})^2 + var(\hat{\theta}) \qquad \text{(proven in Answer 1)}$$

or, $$MSE(\hat{L}_{MOM}) = bias(\hat{L}_{MOM})^2 + var(\hat{L}_{MOM})$$

or, $$MSE(\hat{L}_{MOM}) = 0^2 + \frac{L^2}{3n} \qquad \text{(proven in Answer 2 and 3)}$$

or, $$MSE(\hat{L}_{MOM}) = \frac{L^2}{3n} \qquad (1)$$

$$MSE(\hat{\theta}) = bias(\hat{\theta})^2 + var(\hat{\theta}) \qquad \text{(proven in Answer 1)}$$

or, $$MSE(\hat{L}_{MLE}) = bias(\hat{L}_{MLE})^2 + var(\hat{L}_{MLE})$$

or, $$MSE(\hat{L}_{MLE}) = \left(\frac{L}{n+1}\right)^2 + \frac{nL^2}{(n+2)(n+1)^2} \qquad \text{(proven in Answer 2 and 3)}$$

or, $$MSE(\hat{L}_{MLE}) = \frac{L^2(n+2)+nL^2}{(n+2)(n+1)^2}$$

or, $$MSE(\hat{L}_{MLE}) = \frac{nL^2+2L^2+nL^2}{(n+2)(n+1)^2}$$

or, $$MSE(\hat{L}_{MLE}) = \frac{2nL^2+2L^2}{(n+2)(n+1)^2}$$

or, $$MSE(\hat{L}_{MLE}) = \frac{2L^2(n+1)}{(n+2)(n+1)^2}$$

or, $$MSE(\hat{L}_{MLE}) = \frac{2L^2}{(n+2)(n+1)} \qquad (2)$$

$$MSE_{diff} = MSE(\hat{L}_{MOM}) - MSE(\hat{L}_{MLE})$$

or, $$MSE_{diff} = \frac{L^2}{3n} - \frac{2L^2}{(n+2)(n+1)} \qquad \text{(from 1 and 2)}$$

or, $$MSE_{diff} = \frac{L^2(n+2)(n+1)-(2L^2)(3n)}{(3n)(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{(L^2)(n^2+3n+2)-(2L^2)(3n)}{(3n)(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{(n^2L^2+3nL^2+2L^2)-(6nL^2)}{(3n)(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{n^2L^2-3nL^2+2L^2}{(3n)(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{L^2(n^2-3n+2)}{(3n)(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{L^2(n-1)(n-2)}{(3n)(n+2)(n+1)}$$

MSE_{diff} is positive for n < 1 and n > 2
MSE_{diff} is negative for n between 1 and 2 (exclusive).
MSE_{diff} is 0 for n=1 and n=2

But n ≥ 1, so MSE_{diff} is negative for n < 2 and positive for n > 2.
In general n >>> 1, so MSE_{diff} is positive.

Considering MSE_{diff} as positive, we can say $MSE(\hat{L}_{MLE})$ is less than $MSE(\hat{L}_{MOM})$. Therefore, $\hat{L}_{MLE}$ is a better estimator.

Question 5

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

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

Answer 5

Code:

```
print(chr(27) + "[2J")          #clear terminal

import numpy as np

n = 100          #features
L = 10           #dist. between 0 to L
MAX_ITER = 1000

np.random.seed(0)               #keeping results same

#Generating the (pseudo) random data (uniform distribution)
data = []
for i in range(MAX_ITER):
    data.append(np.random.uniform(0, L, n))

estd_l_mom = []                 #estimated L M.O.M. vector
estd_l_mle = []                 #estimated L M.L.E. vector

# Calculating estimated M.O.M. and M.L.E.
for i in range(len(data)):
    estd_l_mom.append(2*np.mean(data[i]))
    estd_l_mle.append(max(data[i]))

# Calculating estimated MSEs
estd_mse_mom = 0
for val in estd_l_mom:
        estd_mse_mom += pow((val - L),2)
estd_mse_mom = estd_mse_mom/MAX_ITER

estd_mse_mle = 0
for val in estd_l_mle:
        estd_mse_mle += pow((val - L),2)
estd_mse_mle = estd_mse_mle/MAX_ITER

print("Estimated MSE MOM: ",estd_mse_mom)
print("Estimated MSE MLE: ",estd_mse_mle)

# Calculating theoretical MSEs
theoretical_mse_mom = pow(L,2)/(3*n)
theoretical_mse_mle = (2*pow(L,2))/((n+2)*(n+1))

print("Theoretical MSE MOM: ",theoretical_mse_mom)
print("Theoretical MSE MLE: ",theoretical_mse_mle)

# Difference between estimated and theoretical MSEs
print("Absolute difference between theoretical and estimated MSE w.r.t.
MOM: ",abs(theoretical_mse_mom - estd_mse_mom))
```

```
print("Absolute difference between theoretical and estimated MSE w.r.t. MLE: ",abs(theoretical_mse_mle - estd_mse_mle))
```

Output:

```
Estimated MSE MOM:  0.33705516291082566
Estimated MSE MLE:  0.017543456156631155
Theoretical MSE MOM:  0.3333333333333333
Theoretical MSE MLE:  0.01941370607649
Absolute difference between theoretical and estimated MSE w.r.t. MOM:
0.0037218295774923416
Absolute difference between theoretical and estimated MSE w.r.t. MLE:
0.0018702499198588463
```

The difference between the estimated and theoretical MSEs are quite small.
Also, we notice that $MSE(\hat{L}_{MLE}) < MSE(\hat{L}_{MOM})$ for both estimated and theoretical values and hence we can confirm our result from Answer 4 that $MSE(\hat{L}_{MLE})$ is a better estimator.

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

Answer 6
We have proved in Answer 2 that $\hat{L}_{MLE}$ is biased.
We have proved in Answer 4 that $MSE(\hat{L}_{MLE}) < MSE(\hat{L}_{MOM})$.

From the following: $\boldsymbol{MSE(\hat{\theta}) = bias(\hat{\theta})^2 + var(\hat{\theta})}$, we notice that the weightage on bias is more compared to the variance, and yet $MSE(\hat{L}_{MLE}) < MSE(\hat{L}_{MOM})$. This implies that $var(\hat{L}_{MOM})$ is very high compared to $var(\hat{L}_{MLE})$ and the same was proven in Answer 3. As n increases $var(\hat{L}_{MOM})$ increases at a higher rate compared to $var(\hat{L}_{MLE})$, although $bias(\hat{L}_{MOM}) = 0$ and $bias(\hat{L}_{MLE}) > 0$. **$\boldsymbol{\hat{L}_{MLE}}$ has a better bias vs variance trade off and thus we get lower value of $\boldsymbol{MSE(\hat{L}_{MLE})}$ compared to $\boldsymbol{MSE(\hat{L}_{MOM})}$**.

Question 7
Find $P(\hat{L}_{MLE} < L - \varepsilon)$ 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.

Answer 7

$$\hat{L}_{MLE} = \max_{i=1,\dots,n} X_i \qquad \text{(given)}$$

$$F(x) = \left(\frac{x}{L}\right)^n \qquad \text{(from Answer 2 … 1)}$$

Therefore, $$P(\hat{L}_{MLE} \le L - \varepsilon) = \left(\frac{L-\varepsilon}{L}\right)^n$$

or, $$P(\hat{L}_{MLE} \le L - \varepsilon) = \left(1 - \frac{\varepsilon}{L}\right)^n \qquad (1)$$

	$P(L - \hat{L}_{MLE} < \varepsilon) > \delta$	(question requirement)
or,	$1 - P(L - \hat{L}_{MLE} > \varepsilon) > \delta$	
or,	$1 - \delta > P(L - \hat{L}_{MLE} > \varepsilon)$	
or,	$P(L - \hat{L}_{MLE}) < 1 - \delta$	(2)

	$\left(1 - \frac{\varepsilon}{L}\right)^n < 1 - \delta$	(from 1 and 2)
or,	$n\, ln\left(1 - \frac{\varepsilon}{L}\right) < \ln(1 - \delta)$	(taking log on both side)
or,	$n < \frac{\ln(1-\delta)}{ln\left(1-\frac{\varepsilon}{L}\right)}$	

Since δ lies between 0 and 1, this makes $\ln(1 - \delta)$ negative and the inequality is reversed:

or,	$n > -\frac{\ln(1-\delta)}{ln\left(1-\frac{\varepsilon}{L}\right)}$	
or,	$n > \frac{\ln(1-\delta)^{-1}}{ln\left(1-\frac{\varepsilon}{L}\right)}$	
or,	$n > \frac{\ln\left(\frac{1}{1-\delta}\right)}{ln\left(1-\frac{\varepsilon}{L}\right)}$	

Thus, we need at least $\boldsymbol{\frac{\ln\left(\frac{1}{1-\delta}\right)}{ln\left(1-\frac{\varepsilon}{L}\right)} + 1}$ samples to be sure that our estimate was within ε with probability at least δ.

Question 8
Show that

$$\hat{L} = \left(\frac{n+1}{n}\right) \max_{i=1,\dots,n} X_i$$

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

Answer 8

	$\hat{L} = \left(\frac{n+1}{n}\right) \max_{i=1,\dots,n} X_i$	(given…1)

	$bias(\hat{\theta}) = \theta - \mathbb{E}[\hat{\theta}]$	(given)
or,	$bias(\hat{L}) = L - \mathbb{E}[\hat{L}]$	
or,	$bias(\hat{L}) = L - \mathbb{E}\left[\left(\frac{n+1}{n}\right) \max_{i=1,\dots,n} X_i\right]$	(from 1)
or,	$bias(\hat{L}) = L - \left(\frac{n+1}{n}\right) \mathbb{E}\left[\max_{i=1,\dots,n} X_i\right]$	
or,	$bias(\hat{L}) = L - \left(\frac{n+1}{n}\right)\left(\frac{nL}{n+1}\right)$	(from Answer 2 … 3)
or,	$bias(\hat{L}) = 0$	(2)

Thus, $\hat{L} = \left(\frac{n+1}{n}\right) \max_{i=1,\dots,n} X_i$ is an unbiased estimator.

$$var(\hat{L}) = var\left(\left(\frac{n+1}{n}\right) \max_{i=1,\dots,n} X_i\right)$$

or, $$var(\hat{L}) = \left(\frac{n+1}{n}\right)^2 var\left(\max_{i=1,\dots,n} X_i\right)$$

or, $$var(\hat{L}) = \left(\frac{n+1}{n}\right)^2 \left(\frac{nL^2}{(n+2)(n+1)^2}\right)$$ (from Answer 3 … 3)

or, $$var(\hat{L}) = \frac{L^2}{n(n+2)}$$ (3)

$$MSE(\hat{\theta}) = bias(\hat{\theta})^2 + var(\hat{\theta})$$ (proven in Answer 1)

or, $$MSE(\hat{L}) = bias(\hat{L})^2 + var(\hat{L})$$

or, $$MSE(\hat{L}) = 0^2 + \frac{L^2}{n(n+2)}$$ (from 2 and 3)

or, $$MSE(\hat{L}) = \frac{L^2}{n(n+2)}$$

$$MSE(\hat{L}_{MLE}) = \frac{2L^2}{(n+2)(n+1)}$$ (from Answer 2…2)

We need to prove $MSE(\hat{L}) < MSE(\hat{L}_{MLE})$ or $MSE(\hat{L}) - MSE(\hat{L}_{MLE}) < 0$

$$MSE_{diff} = MSE(\hat{L}) - MSE(\hat{L}_{MLE})$$

or, $$MSE_{diff} = \frac{L^2}{n(n+2)} - \frac{2L^2}{(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{L^2(n+1) - 2L^2(n)}{n(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{nL^2 + L^2 - 2nL^2}{n(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{L^2 - nL^2}{n(n+2)(n+1)}$$

or, $$MSE_{diff} = \frac{(1-n)L^2}{n(n+2)(n+1)}$$

Since n ≥ 1, $MSE_{diff} \leq 0$, therefore $MSE(\hat{L}) \leq MSE(\hat{L}_{MLE})$.
In general n has larger values than 1, hence $MSE(\hat{L}) < MSE(\hat{L}_{MLE})$ and the difference is more for larger values of n.

***