**Answer 1**

The objective function (or loss function since minimizing) is:

SE = Squared Error

This represents the ordinary least squares problem of regression.

We have to find and which minimizes the loss function (minimizes the squared error to regression line).

So, we take the partial derivative of the loss function w.r.t. and and set it equal to 0.

and

or,

or,

or,

or,

The yellow highlighted values are the mean values or the expected values, so we can rewrite as:

or,

(1)

or,

or,

or,

or,

The yellow highlighted values are the mean values or the expected values, so we can rewrite as:

or,

(2)

Now we just need the solve the two simultaneous equations (1) and (2) to find and

Rewriting equation (2)

or,

(3)

Rewriting equation (1)

Substituting in the above,

or,

or,

or,

(4)

or,

or,

or,

Using the above, rewriting equation (4)

Substituting from above in (3)

**Answer 2**

From (4)

Previously, we divided by m to get mean, let us undo that (i.e. multiply numerator and denominator by m), then:

(5)

Given true linear model

Substituting the above in (5),

or,

or,

or,

or,

or,

But the mean of errors/noise is 0(given), therefore

and because

Therefore,

This means that is an unbiased estimate of w i.e. it is correct on average.

Now we know

or,

or,

or,

(6)

From (3),

or,

or,

or,

since and expected value of an expected value is just that.

Substituting from (6)

or,

**Alternatively**, in Answer 1, we proved the following:

or,

or,

or,

or,

or,

or,

or,

But , therefore,

This means that is an unbiased estimate of w i.e. it is correct on average.

From (3),

or,

or,

or (since ),

or,

The variances of these estimators are asked to be proven in the next question, so please refer to Answer 3

**Answer 3**

**Finding Variance of :**

or,

or (given ,

or,

or,

or,

or (m-1 for sample variance; for population it would have been m),

or,

or (for large m, m-1=m),

**Finding Variance of :**

or,

or,

or,

Evaluating the above yellow highlighted term

or,

or,

or,

Therefore,

or,

or,

**Answer 4**

**Question 5**

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

- For j = 1, . . . , 1000:

\*Simulate , …, and compute values for and

- 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 for both estimated and theoretical values and hence we can confirm our result from Answer 4 that is a better estimator.

**Question 6**

You should have shown that , 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 is biased.

We have proved in Answer 4 that .

From the following: , we notice that the weightage on bias is more compared to the variance, and yet . This implies that is very high compared to and the same was proven in Answer 3. As n increases increases at a higher rate compared to , although and .  **has a better bias vs variance trade off and thus we get lower value of compared to** .

**Question 7**

Find 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**

(given)

(from Answer 2 … 1)

Therefore,

or, (1)

(question requirement)

or,

or,

or, (2)

(from 1 and 2)

or, (taking log on both side)

or,

Since lies between 0 and 1, this makes negative and the inequality is reversed:

or,

or,

or,

Thus, we need at least samples to be sure that our estimate was within ε with probability at least *δ.*

**Question 8**

Show that

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

**Answer 8**

(given…1)

(given)

or,

or, (from 1)

or,

or, (from Answer 2 … 3)

or, (2)

**Thus, is an unbiased estimator.**

or,

or, (from Answer 3 … 3)

or, (3)

(proven in Answer 1)

or,

or, (from 2 and 3)

or,

(from Answer 2…2)

We need to prove or

or,

or,

or,

or,

or,

Since n ≥ 1, ≤ 0, therefore .

In general n has larger values than 1, hence and the difference is more for larger values of n.

References

1. Lecture Notes and Videos
2. <https://www.khanacademy.org/math/statistics-probability/describing-relationships-quantitative-data/more-on-regression/v/proof-part-4-minimizing-squared-error-to-regression-line>
3. <https://en.wikipedia.org/wiki/Covariance>
4. <https://math.stackexchange.com/questions/2947788/what-is-varx-y>
5. <http://www.unm.edu/~lspear/geog525/24linreg2.pdf>

\*\*\*