HW1 - Estimation Problems CS 536, Machine Learning

Nitin Reddy Karolla

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 $bias(\hat{\theta}) = \theta - E(\hat{\theta})$.

   We know that,

$$var(\hat{\theta}) = E[(\hat{\theta} - E(\hat{\theta}))^2] \tag{1}$$
$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta \tag{2}$$
$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] \tag{3}$$

   On expanding the equation (1), we have

$$\begin{aligned} var(\hat{\theta}) &= E[(\hat{\theta})^2 - 2\hat{\theta}E(\hat{\theta}) + E^2(\hat{\theta})] \\ &= E(\hat{\theta}^2) + E[-2\hat{\theta}E(\hat{\theta})] + E[E^2(\hat{\theta})] \\ &= E(\hat{\theta}^2) - E[2\hat{\theta}E(\hat{\theta})] + E^2(\hat{\theta}) \\ &= E(\hat{\theta}^2) - 2E(\hat{\theta})E(\hat{\theta}) + E^2(\hat{\theta}) \\ &= E(\hat{\theta}^2) - E^2(\hat{\theta}) \end{aligned}$$
$$\therefore var(\hat{\theta}) = E^2(\hat{\theta}) - E(\hat{\theta}^2) \tag{4}$$

   Squaring the equation (2), we have

$$\begin{aligned} bias^2(\hat{\theta}) &= [E(\hat{\theta}) - \theta]^2 \\ &= E^2(\hat{\theta}) - 2E(\hat{\theta})\theta + \theta^2 \end{aligned}$$
$$\therefore bias^2(\hat{\theta}) = E^2(\hat{\theta}) - 2E(\hat{\theta})\theta + \theta^2 \tag{5}$$

   Adding the equations, (4) and (5), we get

$$bias^2(\hat{\theta}) + var(\hat{\theta}) = E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 \tag{6}$$

   Now on expanding equation (3),

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2) \\ &= E(\hat{\theta}^2) - E(2\hat{\theta}\theta) + E(\theta^2) \\ &= E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 \end{aligned}$$

$$\therefore MSE(\hat{\theta}) = E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2 \tag{7}$$

From equations (6) and (7), we can conclude that,

$$MSE(\hat{\theta}) = bias^2(\hat{\theta}) + var(\hat{\theta}) \tag{8}$$

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

   Let us the calculate the bias for the Method of Moments estimator.

$$bias_{MOM} = E(\hat{L}_{MOM}) - L$$

   But we know that $\hat{L}_{MOM} = 2\bar{X}_N$,

$$\begin{aligned} bias_{MOM} &= E(2\bar{X}_N) - L \\ &= 2E(\bar{X}_N) - L \\ &= 2E(\frac{1}{n}\sum_{n=1}^{N} X_i) - L \\ &= \frac{2}{n}\sum_{n=1}^{N} E(X_i) - L \\ &= \frac{2}{n} nE(X) - L \\ &= 2E(X) - L \\ &= 2\frac{L}{2} - L \\ &= L - L = 0 \end{aligned}$$

   Therefore, we can say that Method of Moment estimate for Uniform distribution is unbiased.

   Now, let us calculate the bias for Maximum Likelihood Estimator.
   To calculate expected value of $\hat{L}_{MLE}$, we first assign it a function $Y = max(x_i)$ and calculate the p.d.f.

$$
\begin{aligned}
Y &= max(x_i) \\
F_y(Y) &= P(Y =< y) \\
&= P(max(x_i) =< y) \\
&= P(x_1 =< y ... x_n =< y) \\
&= P(x_1 =< y)P(x_2 =< y)...P(x_n =< y) \\
&= P^n(x <= y) \\
&= y^n/L^n
\end{aligned}
$$

Now,

$$
f_y = F_y\prime(y) = n\frac{1}{L^n}y^{n-1}
$$

Now let us find the expected value of $\hat{L}_{MLE}$, i.e.

$$
\begin{aligned}
E(\hat{L}_{MLE}) &= E(Y) \\
&= \int_{-\infty}^{\infty} yf(y)dy \\
&= \int_0^L yn\frac{1}{L^n}y^{n-1}dy \\
&= \int_0^L (\frac{y}{L})^n dy \\
&= \frac{n}{n+1}(\frac{1}{L})^n.L^{n+1} \\
&= \frac{n}{n+1}L < L
\end{aligned}
$$

$\hat{L}_{MLE}$ is biased since, E[$\hat{L}_{MLE}$] is not equal to L.

In general, we can say the MLE underestimates the parameter L by an factor of n+1/n as seen from the above deduced equations.

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

   Variance of Method of Moment estimator is as follows -

$$
\begin{aligned}
var_{MOM} &= var\hat{L}_{MOM} \\
&= var(2X_N) \\
&= 4var(X_N) \\
&= 4var(\frac{1}{n}\sum_{i=0}^{n} X_i) \\
&= \frac{4}{n^2}\sum_{i=0}^{n} var(X_i) \\
&= \frac{4}{n^2} var(X)
\end{aligned}
$$

We know that variance of Uniform Distribution(0,L) is $L^2/12$.

$$
\begin{aligned}
var_{MOM} &= \frac{4}{n^2} L^2/12 \\
&= \frac{L^2}{3n}
\end{aligned}
$$

Variance for Maximum Likelihood Estimator is as follows -

$$
\begin{aligned}
var_{MLE} &= var(\hat{L}_{MLE}) \\
&= var(Y) \\
&= E(Y^2) - E(Y)^2
\end{aligned}
$$

Now in order to calculate variance we need to calculate $E(Y^2)$, and we already have value of $E(Y) = \frac{n}{n+1}L$ from previous problem.

$$
\begin{aligned}
E(Y^2) &= \int_{-\infty}^{+\infty} y^2 f_Y(y)dy \\
&= \int_{-\infty}^{+\infty} y^2 ny^{n-1}/L^n dy \\
&= \frac{n}{L^n}\int_0 Ly^{n+1}dy \\
&= \frac{n}{L^n}\frac{L^{n+2}}{n+2} \\
&= \frac{nL^2}{n+2}
\end{aligned}
$$

By putting the values of $E(Y^2)$ and $E(Y)$ into the variance formula, we have -

$$\begin{aligned} var_{MLE} &= \frac{nL^2}{n+2} - (\frac{nL}{n+1})^2 \\ &= \frac{nL^2}{(n+1)^2(n+2)} \end{aligned}$$

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

   As the question states, the better estimator is one which gives the smaller mean squared error. Whereas MSE is the summation of square of bias and variance of the parameter as seen before.

   Let us calculate the MSE for $\hat{L}_{MOM}$ -

$$\begin{aligned} MSE_{MOM} &= bias^2_{MOM} + var_{MOM} \\ &= 0 + L^2/3n \\ &= L^2/3n \end{aligned}$$

   Similary MSE for $\hat{L}_{MLE}$ -

$$\begin{aligned} MSE_{MLE} &= bias^2_{MLE} + var_{MLE} \\ &= (\frac{-L}{n+1})^2 + \frac{nL^2}{(n+1)^2(n+2)} \\ &= \frac{(n+1)L^2}{(n+1)^2(n+2)} \\ &= \frac{2L^2}{(n+1)(n+2)} \end{aligned}$$

   Let us now estimate which is error is smaller.

$$
\begin{aligned}
MSE_{MLE} - MSE_{MOM} &= \frac{2L^2}{(n+1)(n+2)} - L^2/3n \\
&= \frac{6nL^2 - L^2(n+1)(n+2)}{3n(n+1)(n+2)} \\
&= \frac{-L^2(n^2-3n+3)}{3n(n+1)(n+2)} \\
&= \frac{-L^2(n-1)(n-2)}{3n(n+1)(n+2)} < 0
\end{aligned}
$$

Clearly as we can see $MSE_{MLE} - MSE_{MOM}$ is negative for any n ¿2, we can say that Maximum Likelihood produces less error and is better estimate than Method of Moment for Uniform Distribution.

5. **Solution to this question is present in the Jupyter notebook attached with the submission.**

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

   We know that mean square error is the combination of bias and variance present in the model. While increasing the bias in the model decreases the variance and similarly increasing variance will lead to decrease in bias. There is bias-variance trade off that needs to taken care to achieve low error. And this is what happens in the case of MSE of MOM estimator, although it is unbiased it has high variance contributing to the error. On the contrary, by having little bias in MSE of MLE, there is a significant reduction in the variance, leading to an error lower than that of MOM.

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

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

$$P(\hat{L}_{MLE} < L - \epsilon) = P(max_{i=1..n} X_i < L - \epsilon) \tag{9}$$

   Also,

$$P(Y < x) = (\frac{x}{L})^n \tag{10}$$

From (9) and (10), we have

$$\begin{aligned}
P(\hat{L}_{MLE} < L - \epsilon) &= P(max_{i=1..n} X_i < L - \epsilon) \\
&= (\frac{L-\epsilon}{L})^n \\
&= (1 - \frac{\epsilon}{L})^n
\end{aligned}$$

$$P(\hat{L}_{MLE} < L - \epsilon) = (1 - \frac{\epsilon}{L})^n \tag{11}$$

On the other side puting the question into equation form we have,

$$\begin{aligned}
P(L - \hat{L}_{MLE} < \epsilon) &< \delta \\
1 - P(L - \hat{L}_{MLE} > \epsilon) &< \delta
\end{aligned}$$

$$P(L - \hat{L}_{MLE} > \epsilon) < 1 - \delta \tag{12}$$

From equations (11) and (12), we now have

$$\begin{aligned}
P(L - \hat{L}_{MLE} > \epsilon) &< 1 - \delta \\
(\frac{L-\epsilon}{L})^n &< 1 - \delta \\
nln(1 - \frac{\epsilon}{L}) &< ln(1 - \delta)
\end{aligned}$$

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

$$\begin{aligned}
-nln(\frac{L}{L-\epsilon}) &< ln(1-\delta) \\
-n &< \frac{ln(1-\delta)}{ln(\frac{L}{L-\epsilon})} \\
n &> \frac{ln(\frac{1}{1-\delta})}{ln(\frac{L}{L-\epsilon})}
\end{aligned}$$

8. Show that $\hat{L} = (\frac{n+1}{n})max_{i=1..n}X_i$ is an unbiased estimator, and has a smaller MSE still.

From the previous questions we know that $E(max(x_i)) = E(Y) = nL/n+1$.
Applying it to the bias formula, we have

$$\begin{aligned}
bias(\hat{L}) &= E(\hat{L}) - L \\
&= E((\frac{n+1}{n})max_{i=1..n}X_i) - L \\
&= (\frac{n+1}{n})E(Y) - L \\
&= (\frac{n+1}{n})\frac{nL}{n+1} - L \\
&= L - L = 0
\end{aligned}$$

Also, we know that $var(max(X_i)) = var(Y) = \frac{nL^2}{(n+1)^2(n+2)}$. Applying to the variance formula, we have -

$$\begin{aligned}
var(\hat{L}) &= var(\frac{n+1}{n}max(X_i) \\
&= (\frac{n+1}{n})^2 var(max(X_i)) \\
&= (\frac{n+1}{n})^2 var(Y) \\
&= \frac{L^2}{n(n+2)}
\end{aligned}$$

Calculating the MSE -

$$\begin{aligned}
MSE(\hat{L}) &= bias(\hat{L})^2 + var(\hat{L}) \\
&= 0 + \frac{L^2}{n(n+2)} \\
&= \frac{L^2}{n(n+2)}
\end{aligned}$$

On comparison with MSE of MLE -

$$\begin{aligned}
MSE(\hat{L}) - MSE(L_{MLE}) &= \frac{L^2}{n(n+2)} - \frac{2L^2}{(n+1)(n+2)} \\
&= \frac{L^2}{n+2}\frac{((n+1)-2n)}{n(n+1)} \\
&= \frac{-L^2(n-1)}{n(n+1)(n+2)} =< 0
\end{aligned}$$

For all n ¿ 1, MSE of $\hat{L}$ is lower than the MSE of MLE as we can see the difference is less than 0. Therefore $\hat{L}$ is unbiased estimator and yet still has a lower MSE.