Matthew Wilder
MATH 326 - Spring 2022
Homework 02 - (§8.1 - 8.4)
Due: Wednesday 02/02/22 at 23:59

1. Suppose Y_1, Y_2, Y_3 and Y_4 are an iid random sample with $N(\mu, \sigma^2)$. For each of the following, determine which of the following estimators of μ are unbiased. Then for each unbiased estimator, calculate the mean square error. Which estimator has the lowest MSE?

   A. Y_1 B. $Y_1 + Y_2$ C. $\frac{Y_1+2Y_2+2Y_3+Y_4}{6}$ D. $\bar{Y}$

   Solution: We need to use the following formulas:

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

$$MSE(\hat{\theta}) = \underbrace{V(\hat{\theta})}_{\text{Precision}} + \underbrace{B(\hat{\theta})^2}_{\text{Accuracy}}$$

   And by definition of the normal distributions defined on $Y_1, \ldots, Y_4$

$$\theta = \mu \qquad E(Y_i) = \mu \qquad V(Y_i) = \sigma^2$$

A.) Y_1

$$\begin{aligned} B(Y_1) &= E(Y_1) - \mu \\ &= \mu - \mu \\ &= 0 \end{aligned}$$

Therefore Y_1 is unbiased.

$$\begin{aligned} \text{MSE}(Y_1) &= V(Y_1) + B(Y_1)^2 \\ &= V(Y_1) + 0^2 \\ &= V(Y_1) \\ &= \sigma^2 \end{aligned}$$

B.) $Y_1 + Y_2$

$$\begin{aligned} B(Y_1 + Y_2) &= E(Y_1 + Y_2) - \mu \\ &= E(Y_1) + E(Y_2) - \mu \\ &= (\mu + \mu) - \mu \\ &= \mu \end{aligned}$$

Therefore $Y_1 + Y_2$ is biased.

C.) $\frac{Y_1+2Y_2+2Y_3+Y_4}{6}$

$$\begin{aligned} B\left(\frac{Y_1+2Y_2+2Y_3+Y_4}{6}\right) &= E\left(\frac{Y_1+2Y_2+2Y_3+Y_4}{6}\right) - \mu \\ &= \frac{E(Y_1)+2E(Y_2)+2E(Y_3)+E(Y_4)}{6} - \mu \\ &= \frac{6E(Y_i)}{6} - \mu \\ &= \mu - \mu \\ &= 0 \end{aligned}$$

Therefore $\frac{Y_1+2Y_2+2Y_3+Y_4}{6}$ is unbiased.

$$\begin{aligned} \text{MSE}\left(\frac{Y_1+2Y_2+2Y_3+Y_4}{6}\right) &= V\left(\frac{Y_1+2Y_2+2Y_3+Y_4}{6}\right) + 0^2 \\ &= \frac{1}{36}\big(V(Y_1)+2^2V(Y_2)+2^2V(Y_3)+V(Y_4)\big) \\ &= \frac{10}{36}V(Y_i) \\ &= \frac{5}{18}\sigma^2 \end{aligned}$$

D.) $\bar{Y}$

By definition, $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{Y_1+Y_2+Y_3+Y_4}{4}$

$$\begin{aligned} B\left(\frac{Y_1+Y_2+Y_3+Y_4}{4}\right) &= E\left(\frac{Y_1+Y_2+Y_3+Y_4}{4}\right) - \mu \\ &= \frac{E(Y_1)+E(Y_2)+E(Y_3)+E(Y_4)}{4} - \mu \\ &= \frac{4E(Y_i)}{4} - \mu \\ &= \mu - \mu \\ &= 0 \end{aligned}$$

Therefore $\bar{Y}$ is unbiased.

$$\begin{aligned} \text{MSE}\left(\frac{Y_1+Y_2+Y_3+Y_4}{4}\right) &= V\left(\frac{Y_1+Y_2+Y_3+Y_4}{4}\right) + 0^2 \\ &= \frac{1}{16}\left(V(Y_1)+V(Y_2)+V(Y_3)+V(Y_4)\right) \\ &= \frac{4}{16}V(Y_i) \\ &= \frac{1}{4}\sigma^2 \end{aligned}$$

$\boxed{\text{The estimator } \bar{Y} \text{ has the lowest MSE at } \sigma^2/4.}$

2. In class we explored the estimators $\hat{p}_1 = \bar{Y}$ and $\hat{p}_2 = \frac{\sum Y_i + 1}{n+2}$ for population proportion p. For $n = 15$, what values of p is $\hat{p}_2$ the better estimator with respect to MSE?

Solution: $\hat{p}_1 = \bar{Y} = \frac{\sum_{i=1}^{15} Y_i}{15}$ by definition of $\bar{Y}$. $\hat{p}_2 = \frac{\sum_{i=1}^{15} Y_i + 1}{15+2}$. Using formulas and equations from the notes, it can be shown that

$$MSE(\hat{p}_1) = \frac{p(1-p)}{15}$$

Using the notes we can show that

$$MSE(\hat{p}_2) = \frac{15p(1-p) + (1-2p)^2}{(15+2)^2}$$

Setting $MSE(\hat{p}_1) = MSE(\hat{p}_2)$ we can find their intersection:

$$\frac{p(1-p)}{15} = \frac{15p(1-p) + (1-2p)^2}{(15+2)^2}$$

$$\begin{aligned}
\frac{p-p^2}{15}\,\text{lcd}(15,289) &= \frac{15p(1-p)+(1-2p)^2}{289}\,\text{lcd}(15,289) \\
\frac{p-p^2}{15}\cdot 4335 &= \frac{15p(1-p)+(1-2p)^2}{289}\cdot 4335 \\
289(p-p^2) &= 15\left[15p(1-p) + (1+2p)^2\right] \\
289p - 289p^2 &= 225p - 225p^2 + 15 - 60p + 60p^2 \\
124p^2 - 124p + 15 &= 0 \\
p &= \frac{124 \pm \sqrt{(-124)^2 - 4(124)(15)}}{2(124)} \\
p &= \frac{31 \pm 4\sqrt{31}}{62} = \frac{1}{2} \pm \frac{2\sqrt{31}}{31}
\end{aligned}$$

Therefore the intersection is at $p = \frac{1}{2} \pm \frac{2\sqrt{31}}{31}$. To show which is better, we'll look at their first derivatives.

$$MSE(\hat{p}_1)'(p) = \frac{1}{15}(1-2p) \qquad \text{and} \qquad MSE(\hat{p}_2)'(p) = \frac{11}{289}(1-2p)$$

Because $\frac{11}{289} < \frac{1}{15}$ and the intersections are at $p = \frac{1}{2} \pm \frac{2\sqrt{31}}{31}$,

$$MSE(\hat{p}_2) < MSE(\hat{p}_1) \;\text{ for }\; p \in \left(\frac{1}{2} - \frac{2\sqrt{31}}{31},\quad \frac{1}{2} + \frac{2\sqrt{31}}{31}\right)$$

$$\boxed{\text{Therefore } \hat{p}_2 \text{ is better for } p \in \left(\tfrac{1}{2} - \tfrac{2\sqrt{31}}{31},\quad \tfrac{1}{2} + \tfrac{2\sqrt{31}}{31}\right)}$$

3. Suppose $Y_1, Y_2, Y_3, \ldots, Y_n$ is an iid random sample from a distribution with the following density function

$$f(y) = \frac{3y^2}{\theta^3} \quad \text{on support} \quad y \in (0, \theta)$$

Consider two estimators of θ:

$$\hat{\theta}_1 = \bar{Y} \qquad \text{and} \qquad \hat{\theta}_2 = \max(Y_1, Y_2, Y_3, \ldots, Y_n)$$

(a) Show that $\hat{\theta}_1$ is a biased estimator of θ

Solution: We need to compute the bias for $\hat{\theta}_1$. But note that $\hat{\theta}_1 = \bar{Y} = \frac{Y_1+\cdots+Y_n}{n}$. Thus

$$\begin{aligned}
B(\hat{\theta}_1) &= B(\bar{Y}) \\
&= B\left(\frac{Y_1 + \cdots + Y_n}{n}\right) \\
&= E\left(\frac{Y_1 + \cdots + Y_n}{n}\right) - \theta \\
&= E\left(\frac{n \cdot Y_i}{n}\right) - \theta \\
&= E(Y_i) - \theta \\
&= \int_0^\theta y \cdot f(y) - \theta \\
&= \int_0^\theta y \cdot \frac{3y^2}{\theta^3}\, dy - \theta \\
&= \frac{3}{\theta^3} \int_0^\theta y^3\, dy - \theta \\
&= \frac{3}{4\theta^3} \left[y^4\right]_{y=0}^{y=\theta} - \theta \\
&= \frac{3}{4\theta^3} \cdot \theta^4 - \theta \\
&= \frac{3}{4}\theta - \theta \\
&= -\frac{\theta}{4} \\
&\neq 0
\end{aligned}$$

$$\boxed{\text{Therefore } \hat{\theta}_1 \text{ is biased}}$$

(b) Define a multiple of $\hat{\theta}_1$ that is an unbiased estimator of θ. Call this new estimator $\tilde{\theta}_1$.

Solution: We know from (a) that $E(\hat{\theta}_1) = \frac{3}{4}\theta$. But we want this to equal just θ. Therefore we can find a constant $c > 0$ such that $E(c\hat{\theta}_1) = \theta = \tilde{\theta}_1$

$$\begin{aligned}
\tilde{\theta}_1 &= E(c\hat{\theta}_1) \\
&= E(c\bar{Y}) \\
&= E\left(\frac{cY_1 + \cdots + cY_n}{n}\right) \\
&= cE\left(Y_i\right) \\
&= c\frac{3}{4}\theta \\
&= 1
\end{aligned}$$

$$c\frac{3}{4}\theta \qquad \Longrightarrow \qquad c = \frac{4}{3}$$

$$\boxed{\text{Therefore } \tilde{\theta}_1 = \tfrac{4}{3}\hat{\theta}_1}$$

(c) Compute the MSE for $\tilde{\theta}_1$.

$$\begin{aligned}
\mathrm{MSE}(\tilde{\theta}_1) &= \mathrm{Var}(\tilde{\theta}_1) - \mathrm{B}(\tilde{\theta}_1)^2 \\
&= \mathrm{Var}\left(\frac{4}{3}\hat{\theta}_1\right) - 0^2 \\
&= \mathrm{Var}\left(\frac{4}{3}\bar{Y}\right) \\
&= \mathrm{Var}\left(\frac{4}{3n}\left[Y_1 + \cdots + Y_n\right]\right) \\
&= \left(\frac{4}{3n}\right)^2 \mathrm{Var}\left(Y_1 + \cdots + Y_n\right) \\
&= \frac{16}{9n^2}\left(\underbrace{\mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_n)}_{n \text{ iid}}\right) \\
&= \frac{16}{9n^2} n \, \mathrm{Var}(Y_i) \\
&= \frac{16}{9n} \mathrm{Var}(Y_i) \\
&= \frac{16}{9n}\left(E\left(Y^2\right) - \left[E\left(Y\right)\right]^2\right) \\
&= \frac{16}{9n}\left(\int_0^\theta y^2 f(y) - \left[\int_0^\theta y f(y)\right]^2\right) \\
&= \frac{16}{9n}\left(\int_0^\theta \frac{3y^4}{\theta^3}\, dy - \left[\int_0^\theta \frac{3y^3}{\theta^3}\, dy\right]^2\right) \\
&= \frac{16}{9n}\left(\frac{3}{\theta^3}\int_0^\theta y^4\, dy - \left[\int_0^\theta \frac{3y^3}{\theta^3}\, dy\right]^2\right) \\
&= \frac{16}{9n}\left(\frac{3}{5\theta^3}\left[y^5\right]_{y=0}^{y=\theta} - \left[\frac{3}{4\theta^3}\left[y^4\right]_{y=0}^{y=\theta}\right]^2\right) \\
&= \frac{16}{9n}\left(\frac{3}{5}\theta^2 - \left[\frac{3}{4}\theta\right]^2\right) \\
&= \frac{16}{9n}\left(\frac{3}{5}\theta^2 - \frac{9}{16}\theta^2\right) \\
&= \frac{1}{15n}\theta^2
\end{aligned}$$

$$\boxed{\text{Therefore } \mathrm{MSE}(\tilde{\theta}_1) = \tfrac{1}{15n}\theta^2}$$

(d) Show that $\hat{\theta}_2$ is a biased estimator of θ

Solution: We need to compute the bias for $\hat{\theta}_2$. But note that $\hat{\theta}_2 = \max(Y_1, \ldots, Y_n)$. We will first compute a closed form for the maximum.

$$\begin{aligned}
\max(Y_1, \ldots, Y_n) &= n\left[F(y)\right]^{n-1} f(y) \\
&= n\left(\frac{y^3}{\theta^3}\right)^{n-1}\left(\frac{3y^2}{\theta^3}\right) \\
&= n\left(\frac{y^{3n-3}}{\theta^{3n-3}}\right)\left(\frac{3y^2}{\theta^3}\right) \\
&= n\left(\frac{3y^{3n-3+2}}{\theta^{3n-3+3}}\right) \\
&= 3n\left(\frac{y^{3n-1}}{\theta^{3n}}\right)
\end{aligned}$$

$$\begin{aligned}
B(\hat{\theta}_2) &= B(\max(Y_1, \ldots, Y_n)) \\
&= B\left(3n\frac{y^{3n-1}}{\theta^{3n}}\right) \\
&= E\left(3n\frac{y^{3n-1}}{\theta^{3n}}\right) - \theta \\
&= \int_0^\theta y \cdot \frac{3ny^{3n-1}}{\theta^3} - \theta \\
&= \frac{3n}{\theta^3}\int_0^\theta y^{3n} - \theta \\
&= \frac{3n}{\theta^3}\left[\frac{y^{3n+1}}{3n+1}\right]_{y=0}^{y=\theta} - \theta \\
&= \frac{3n}{3n+1} \cdot \frac{\theta^{3n+1}}{\theta^3} - \theta \\
&= \frac{3n}{3n+1}\theta - \theta \\
&\neq 0
\end{aligned}$$

$\boxed{\text{Therefore } \hat{\theta}_2 \text{ is biased}}$

(e) Define a multiple of $\hat{\theta}_2$ that is an unbiased estimator of θ. Call this new estimator $\tilde{\theta}_2$.

Solution: We need the reciprocal of the coefficient on $E(\hat{\theta}_2)$. From part (d), $E(\hat{\theta}_2) = \frac{3n}{3n+1}\,\theta$, so the inverse of $\frac{3n}{3n+1}$ is $\frac{3n+1}{3n}$. Therefore,

$$\tilde{\theta}_2 = \frac{3n+1}{3n}\,\hat{\theta}_2$$

$\boxed{\text{Therefore } \tilde{\theta}_2 = \frac{3n+1}{3n}\,\hat{\theta}_2}$

(f) Compute the MSE for $\tilde{\theta}_2$.

Solution:

$$\begin{aligned}
\text{MSE}(\tilde{\theta}_2) &= \text{Var}(\tilde{\theta}_2) - \text{B}(\tilde{\theta}_2)^2 \\
&= \text{Var}\left(\frac{3n+1}{3n}\hat{\theta}_2\right) - 0^2 \\
&= \left(\frac{3n+1}{3n}\right)^2 \text{Var}(\hat{\theta}_2) \\
&= \left(\frac{3n+1}{3n}\right)^2 \text{Var}\left(3n\frac{y^{3n-1}}{\theta^{3n}}\right) \\
&= \left(\frac{3n+1}{3n}\right)^2 (3n)^2\,\text{Var}\left(\frac{y^{3n-1}}{\theta^{3n}}\right) \\
&= (3n+1)^2\,\text{Var}\left(\frac{y^{3n-1}}{\theta^{3n}}\right) \\
&= (3n+1)^2 \left(E\left(\left(\frac{y^{3n-1}}{\theta^{3n}}\right)^2\right) - \left[E\left(\frac{y^{3n-1}}{\theta^{3n}}\right)\right]^2\right) \\
&= (3n+1)^2 \left(\int_0^\theta y^2 f(y) - \left[\int_0^\theta y f(y)\right]^2\right) \\
&= (3n+1)^2 \left(\int_0^\theta y^2 \cdot \frac{y^{3n-1}}{\theta^{3n}}dy - \left[\int_0^\theta y \cdot \frac{y^{3n-1}}{\theta^{3n}}dy\right]^2\right) \\
&= (3n+1)^2 \left(\int_0^\theta \frac{y^{3n+1}}{\theta^{3n}}dy - \left[\int_0^\theta \frac{y^{3n}}{\theta^{3n}}dy\right]^2\right) \\
&= (3n+1)^2 \left(\frac{[y^{3n+2}]_{y=0}^{y=\theta}}{(3n+2)\cdot\theta^{3n}} - \left[\frac{[y^{3n+1}]_{y=0}^{y=\theta}}{(3n+1)\cdot\theta^{3n}}\right]^2\right) \\
&= (3n+1)^2 \left(\frac{\theta^{3n+2}}{(3n+2)\cdot\theta^{3n}} - \left[\frac{\theta^{3n+1}}{(3n+1)\cdot\theta^{3n}}\right]^2\right) \\
&= (3n+1)^2 \left(\frac{\theta^2}{(3n+2)} - \frac{\theta^2}{(3n+1)^2}\right) \\
&= \theta^2 \left(\frac{(3n+1)^2}{(3n+2)} - \frac{(3n+1)^2}{(3n+1)^2}\right) \\
&= \theta^2 \left(\frac{(3n+1)^2}{3n+2} - 1\right) \\
&= \frac{9n^2+3n-1}{3n+2}\,\theta^2
\end{aligned}$$

$$\boxed{\text{Therefore MSE}(\tilde{\theta}_2) = \tfrac{9n^2+3n-1}{3n+2}\,\theta^2}$$

4. In a study of the relationship between birth order and college success, an investigator found that 126 in a sample of 180 college graduates were firstborn or only children; in a sample of 100 nongraduates of comparable age and socioeconomic background, the number of firstborn or only children was 54. Estimate the difference in the proportions of firstborn or only children for two populations from which these samples were drawn. Give a bound of error of estimation.

Solution: The information provided can be rearranged into a table,

	Nongraduate	Graduate
First Born or Only Child	54	126
Not First Born or Only Child	46	54

Which can be rewritten into their respective probabilities,

	Nongraduate	Graduate
First Born or Only Child	0.54	0.70
Not First Born or Only Child	0.46	0.30

Denote non-graduates by $\hat{p}_1$ and graduates by $\hat{p}_2$. Then,

$$\hat{p}_2 - \hat{p}_1 = 0.7 - 0.54 = 0.16$$

Computing the error bound, we get

$$2\left(\sqrt{\frac{0.7 \cdot 0.3}{180} + \frac{0.54 \cdot 0.46}{100}}\right) \approx 0.1208415$$

Thus the difference in proportion is 0.16 and the error bound is $E \approx 0.1208415$

5. Suppose $\hat{\theta}_1, \hat{\theta}_2$, and $\hat{\theta}_3$ are all unbiased estimators of θ. Suppose that $\text{Var}(\hat{\theta}_i) = 1 + i$ for $i = 1, 2, 3$. Let $X = a\hat{\theta}_1 + b\hat{\theta}_2 + c\hat{\theta}_3$, where a, b, and c are non-negative constants with $a + b + c = 1$

   (a) Show that X is unbiased for θ

Solution: To compute bias, we need to find $E(X)$,

$$\begin{aligned} E(X) &= E(a\hat{\theta}_1 + b\hat{\theta}_2 + c\hat{\theta}_3) \\ &= E(a\theta + b\theta + c\theta) \\ &= \theta E(a + b + c) \\ &= \theta E(1) \\ &= \theta \end{aligned}$$

$$\begin{aligned} B(X) &= E(X) - \theta \\ &= \theta - \theta \\ &= 0 \end{aligned}$$

Therefore, X is unbiased for θ.

(b) Assuming that the $\hat{\theta}_i$'s are independent, find a, b, and c that minimize $\text{Var}(X)$.

Solution: We can explicitly state the variances of $\hat{\theta}_i$. Thus,

$$\begin{aligned}\text{Var}(\hat{\theta}_1) &= 1 + 1 = 2\\ \text{Var}(\hat{\theta}_2) &= 1 + 2 = 3\\ \text{Var}(\hat{\theta}_3) &= 1 + 3 = 4\end{aligned}$$

Using this information, we can compute $\text{Var}(X)$. Because $\hat{\theta}_1, \hat{\theta}_2$, and $\hat{\theta}_3$ are independent, $\text{Cov}(\hat{\theta}_i, \hat{\theta}_j) = 0$ and thus,

$$\begin{aligned}\text{Var}(X) &= \text{Var}(a\hat{\theta}_1 + b\hat{\theta}_2 + c\hat{\theta}_3)\\ &= \text{Var}(a\hat{\theta}_1) + \text{Var}(b\hat{\theta}_2) + \text{Var}(c\hat{\theta}_3)\\ &= a^2 \text{Var}(\hat{\theta}_1) + b^2 \text{Var}(\hat{\theta}_2) + c^2 \text{Var}(\hat{\theta}_3)\\ &= 2a^2 + 3b^2 + 4c^2\end{aligned}$$

Let $f(a, b, c) = 2a^2 + 3b^2 + 4c^2$ and $g(a, b, c) = a + b + c$ with constraints $a \geq 0, b \geq 0$, and $c \geq 0$ and $a + b + c = 1$. Then,

$$\begin{aligned}\nabla f(a, b, c) &= \lambda \nabla g(a, b, c)\\ \left\langle \frac{\partial f}{\partial a}, \frac{\partial f}{\partial b}, \frac{\partial f}{\partial c} \right\rangle &= \lambda \left\langle \frac{\partial g}{\partial a}, \frac{\partial g}{\partial b}, \frac{\partial g}{\partial c} \right\rangle\\ \langle 4a, 6b, 8c \rangle &= \lambda \langle 1, 1, 1 \rangle\\ \langle 4a, 6b, 8c \rangle &= \langle \lambda, \lambda, \lambda \rangle\end{aligned}$$

This equality implies that,

$$4a = \lambda \qquad 6b = \lambda \qquad 8c = \lambda$$

$$a = \frac{\lambda}{4} \qquad b = \frac{\lambda}{6} \qquad c = \frac{\lambda}{8}$$

From our constraint, $a + b + c = 1$. Therefore,

$$\frac{\lambda}{4} + \frac{\lambda}{6} + \frac{\lambda}{8} = 1 \quad \Longrightarrow \quad \lambda = \frac{24}{13}$$

Substituting back into a, b, and c, we get that $f(\frac{6}{13}, \frac{4}{13}, \frac{3}{13})$ is a critical point. Comparing to a point in the region, say, $f(0, 0, 1)$, then

$$\text{Var}(\hat{\theta}_3) = 4 > \frac{12}{13} = \text{Var}(\frac{6}{13}\hat{\theta}_1 + \frac{4}{13}\hat{\theta}_2 + \frac{3}{13}\hat{\theta}_3)$$

Therefore, $(\frac{6}{13}, \frac{4}{13}, \frac{3}{13})$ must be the minimum, since it is the only turning point and we have show a value greater than it. Thus,

$$\boxed{a = \frac{6}{13} \qquad b = \frac{4}{13} \qquad c = \frac{3}{13}}$$