Recitation Note: Mean Squared Error

Tyler Maunu
maunut@mit.edu

September 3, 2019

This note will cover a few problems on mean squared error (MSE). We will first understand MSE by deriving the bias-variance decomposition, and then apply this to a few examples. We will then show how some unbiased estimators can be improved (in terms of smaller MSE) by using the technique of shrinkage.

Mean Squared Error of and Estimator

The MSE of an estimator is given be

$$\mathrm{mse}(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta_0)^2, \tag{1}$$

where θ_0 is the true parameter in our model.

The following decomposition of the MSE is useful:

$$\begin{aligned}
\mathrm{mse}(\hat{\theta}) &= \mathbb{E}(\hat{\theta} - \theta_0)^2 \\
&= \mathbb{E}\hat{\theta}^2 + \mathbb{E}\theta_0^2 - 2\mathbb{E}\hat{\theta}\theta_0 \\
&= \mathrm{var}(\hat{\theta}) + (\mathbb{E}\hat{\theta})^2 + \theta_0^2 - 2\theta_0\mathbb{E}\hat{\theta} \\
&= \mathrm{var}(\hat{\theta}) + (\mathbb{E}\hat{\theta} - \theta_0)^2 \\
&= \mathrm{var}(\hat{\theta}) + (\mathrm{bias}(\hat{\theta}))^2.
\end{aligned} \tag{2}$$

This is the bias-variance decomposition of MSE.

For two unbiased estimators of θ, $\hat{\theta}_1$ and $\hat{\theta}_2$, the relative efficiency of $\hat{\theta}_1$ versus $\hat{\theta}_2$ is

$$\mathrm{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\mathrm{var}(\hat{\theta}_1)}{\mathrm{var}(\hat{\theta}_2)}. \tag{3}$$

Questions

1. Calculate the bias, variance and MSE for the following distributions and estimators:

   (a) $X_1, \dots, X_n \overset{i.i.d.}{\sim} \mathsf{Poisson}(\lambda)$, $\hat{\lambda} = \overline{X_n}$.

(b) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Poisson}(\lambda)$, $\hat{\lambda} = X_1$.

(c) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Unif}([0, \theta])$, $\hat{\theta} = \max_i X_i$.

(d) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Unif}([0, \theta])$, $\hat{\theta} = 2\overline{X_n}$.

(e) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} f(x; \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right)$, $\hat{\sigma} = \frac{\sum_i |X_i|}{n}$.

2. What is the relative efficiency of (a) vs (b)? What about the unbiased estimator based on (c) vs (d)?

3. Find the minimum MSE shrinkage estimators of the mean and variance for the $N(\mu, \sigma^2)$ distribution.

Solutions

1. (a)

$$\mathbb{E}\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}X_i = \lambda. \tag{4}$$

$$\text{var}(\overline{X_n}) = \frac{1}{n^2} \sum_{i=1}^{n} \text{var}(X_i) = \frac{\lambda}{n}. \tag{5}$$

$$\text{mse}(\overline{X_n}) = \frac{\lambda}{n} + 0 = \frac{\lambda}{n}. \tag{6}$$

(b)

$$\mathbb{E}X_1 = \lambda. \tag{7}$$

$$\text{var}(X_1) = \lambda. \tag{8}$$

$$\text{mse}(\overline{X_n}) = \lambda + 0 = \lambda. \tag{9}$$

(c) First, the cdf of $\hat{\theta}$ is given by

$$\begin{aligned} F(x) = \mathbb{P}(\hat{\theta} \leq x) &= \mathbb{P}(\max_i X_i \leq x) \\ &= (\mathbb{P}(X_i \leq x))^n \\ &= \left(\frac{x}{\theta}\right)^n. \end{aligned} \tag{10}$$

where $x \in [0, \theta]$. The pdf is

$$f(x) = \frac{nx^{n-1}}{\theta^n}, \ x \in [0, \theta]. \tag{11}$$

We find

$$\mathbb{E}\hat{\theta} = \int_0^\theta \frac{nx^n}{\theta^n} dx = \frac{n}{n+1} \frac{x^{n+1}}{\theta^n}\Big|_0^\theta = \frac{n}{n+1}\theta. \tag{12}$$

and

$$\mathbb{E}\hat{\theta}^2 = \int_0^\theta \frac{nx^{n+1}}{\theta^n} dx = \frac{n}{n+2} \frac{x^{n+2}}{\theta^n}\Big|_0^\theta = \frac{n}{n+2}\theta^2. \tag{13}$$

$$
\begin{aligned}
\operatorname{mse}(\hat{\theta}) &= \operatorname{var}(\hat{\theta}) + (\operatorname{bias}(\hat{\theta}))^2 \\
&= \left(\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\right) + \left(\frac{1}{n+1}\theta\right)^2 \\
&= \theta^2 \frac{n(n+1)^2 - n^2(n+2) + (n+2)}{(n+2)^2(n+1)} \\
&= \theta^2 \frac{n+3}{(n+2)^2(n+1)}.
\end{aligned} \tag{14}
$$

(d)

$$\mathbb{E}\hat{\theta} = \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}X_i = 2\frac{\theta}{2} = \theta. \tag{15}$$

$$\operatorname{var}(\hat{\theta}) = \frac{4}{n^2}\sum_{i=1}^{n}\operatorname{var}(X_i) = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}. \tag{16}$$

$$\operatorname{mse}(\hat{\theta}) = \operatorname{var}(\hat{\theta}) + (\operatorname{bias}(\hat{\theta}))^2 = \frac{\theta^2}{3n} + 0 = \frac{\theta^2}{3n}. \tag{17}$$

(e)

$$\mathbb{E}\hat{\sigma} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|X_i|. \tag{18}$$

We have

$$\mathbb{E}|X_i| = \int_{-\infty}^{\infty}\frac{|x|}{2\sigma}\exp(-|x|/\sigma)dx = \int_{0}^{\infty}\frac{x}{\sigma}\exp(-x/\sigma)dx = \sigma. \tag{19}$$

Thus $\mathbb{E}\hat{\sigma} = \sigma$.

$$\operatorname{var}(\hat{\sigma}) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(|X_i|). \tag{20}$$

Notice

$$\mathbb{E}|X_i|^2 = \int_{0}^{\infty}\frac{x^2}{\sigma}\exp(-x/\sigma) = \sigma^2 + \sigma^2 = 2\sigma^2. \tag{21}$$

Thus, $\operatorname{var}(\hat{\sigma}) = \frac{\sigma^2}{n}$.

$$\operatorname{mse}(\hat{\sigma})) = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n}. \tag{22}$$

2. (a) vs (b) is

$$\frac{\lambda/n}{\lambda} = \frac{1}{n}. \tag{23}$$

The unbiased estimator based on (c) is

$$\hat{\theta} = \frac{n+1}{n}\max_i X_i. \tag{24}$$

This estimator has variance

$$\operatorname{var}(\hat{\theta}) = \left(\frac{n+1}{n}\right)^2 \operatorname{var}(\hat{\theta}) = \left(\frac{n+1}{n}\right)^2 \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\theta^2 = \frac{1}{n(n+2)}\theta^2. \tag{25}$$

Thus, we have a relative efficiency of

$$\frac{\theta^2/(n(n+2))}{\theta^2/(3n)} = \frac{3}{n+2} \tag{26}$$

3. (a) If $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$, then we know that $\overline{X_n}$ is an unbiased estimator of μ. We have

$$\operatorname{mse}(\overline{X_n}) = \operatorname{var}(\overline{X_n}) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}(X_i) = \frac{\sigma^2}{n}. \tag{27}$$

Now consider the estimator $a\overline{X_n}$, for $a \in (0, \infty)$. Notice that this estimator is biased if $a \neq 1$:

$$\mathbb{E}a\overline{X_n} = a\mu. \tag{28}$$

We have thus

$$\begin{aligned} \operatorname{mse}(a\overline{X_n}) &= \operatorname{var}(a\overline{X_n}) + (\operatorname{bias}(a\overline{X_n}))^2 \\ &= a^2\frac{\sigma^2}{n} + ((1-a)\mu)^2 \\ &= \left(\frac{\sigma^2}{n} + \mu^2\right) a^2 - 2\mu^2 a + \mu^2. \end{aligned} \tag{29}$$

Notice that this is a quadratic equation in a that is minimized when

$$a = \frac{\mu^2}{\sigma^2/n + \mu^2}. \tag{30}$$

Notice that this is not 1! In fact, this means that we can achieve smaller MSE using a biased estimator. The MSE of the estimator using this choice of a is

$$\begin{aligned} \operatorname{mse}(a\overline{X_n}) &= \left(\sigma^2/n + \mu^2\right)\left(\frac{\mu^2}{\sigma^2/n + \mu^2}\right)^2 - 2\mu^2\left(\frac{\mu^2}{\sigma^2/n + \mu^2}\right) + \mu^2 \\ &= \frac{\mu^4}{\sigma^2/n + \mu^2} - 2\frac{\mu^4}{\sigma^2/n + \mu^2} + \mu^2 \\ &= \mu^2\frac{\sigma^2/n + \mu^2}{\sigma^2/n + \mu^2} - \frac{\mu^4}{\sigma^2/n + \mu^2} \\ &= \frac{\sigma^2}{n}\frac{\mu^2}{\sigma^2/n + \mu^2} < \frac{\sigma^2}{n}. \end{aligned} \tag{31}$$

(b) For this, we will consider estimation of the variance for a $N(0, \sigma^2)$ distribution. In this case, the unbiased estimator of the variance is given by

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2. \tag{32}$$

Notice the choice of $1/n$ here since the distribution has a fixed mean parameter of 0. We have

$$\operatorname{mse}(\hat{\sigma}^2) = \operatorname{var}(\hat{\sigma}^2) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}(X_i^2) = \frac{2\sigma^4}{n}. \tag{33}$$

Repeating the shrinkage procedure as before, we find

$$\begin{aligned} \operatorname{mse}(\hat{\sigma}^2) &= a^2\frac{2\sigma^4}{n} + ((1-a)\sigma^2)^2 \\ &= \left(\frac{2\sigma^4}{n} + \sigma^4\right)a^2 - 2\sigma^4 a + \sigma^4. \end{aligned} \tag{34}$$

This is minimized when

$$a = \frac{\sigma^4}{2\sigma^4/n + \sigma^4} = \frac{1}{2/n+1} = \frac{n}{2+n}, \tag{35}$$

and it results in an MSE of

$$\begin{aligned} \operatorname{mse}(\hat{\sigma}^2) &= \left(\frac{2\sigma^4}{n} + \sigma^4\right)\left(\frac{n}{2+n}\right)^2 - 2\sigma^4\frac{n}{2+n} + \sigma^4 \\ &= \sigma^4\left(\left(\frac{n}{2+n}\right) - \frac{2n}{2+n} + 1\right) \\ &= \frac{2\sigma^4}{2+n}. \end{aligned} \tag{36}$$