Evaluation of estimation accuracy

- Suppose that $(X_1, \ldots, X_n)$ is a sample and the distribution of $(X_1, \ldots, X_n)$ depends on a parameter vector θ, where θ is in some space Θ. Suppose that T_n is an estimator of $g(\theta)$, where g is a real-valued function, then it is common to use the mean squared error (or mean square error; MSE)

$$E(T_n - g(\theta))^2$$

to evaluate the estimation accuracy of T_n.

- Example 1. Suppose that $(X_1, \ldots, X_n)$ is a random sample and X_i has a PDF f_θ, where $\theta > 0$ and for $x \in (-\infty, \infty)$,

$$f_\theta(x) = I_{[0,\theta]}(x) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta; \\ 0 & \text{otherwise.} \end{cases}$$

Here for a set $A \subset R^d$, the function I_A is defined on R^d so that

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A; \\ 0 & \text{otherwise.} \end{cases}$$

(a) Let $X_{(n)} = \max(X_1, \ldots, X_n)$. Show that $X_{(n)}$ is the MLE of θ.

(b) For $x \in (-\infty, \infty)$, define

$$F_n(x) = \begin{cases} 0 & \text{if } x \le 0; \\ (x/\theta)^n & \text{if } 0 < x < \theta; \\ 1 & \text{if } x \ge \theta. \end{cases}$$

Show that F_n is the CDF of $X_{(n)}$.

(c) For $x \in (-\infty, \infty)$, define

$$f_n(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} I_{[0,\theta]}(x) = \begin{cases} \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} & \text{if } 0 \le x \le \theta; \\ 0 & \text{otherwise.} \end{cases}$$

Show that f_n is a PDF of $X_{(n)}$.

(d) Show that the MSE of $X_{(n)}$ is $\dfrac{2\theta^2}{(n+1)(n+2)}$.

(e) Show that $2\bar{X}$ is a method of moment estimator of θ, where $\bar{X} = \sum_{i=1}^n X_i/n$ and the MSE of $2\bar{X}$ is $\dfrac{\theta^2}{3n}$.

A sketch of solution.

(a) The likelihood function L is

$$L(\eta) = \begin{cases} \left(\frac{1}{\eta}\right)^n & \text{if } X_{(n)} \le \eta; \\ 0 & \text{otherwise.} \end{cases} = \left(\frac{1}{\eta}\right)^n I_{[X_{(n)},\infty)}(\eta).$$

Note that $L(\eta)$ is maximized when $\eta \in [X_{(n)}, \infty)$ and η is minimized. Since the smallest η in $[X_{(n)}, \infty)$ is $X_{(n)}$, the MLE of θ is $X_{(n)}$.

(b) Let F be the CDF of X_1, then

$$F(x) = \int_{-\infty}^{x} f_\theta(t)dt = \begin{cases} 0 & \text{if } x \le 0; \\ \int_0^x \frac{1}{\theta} dt = (x/\theta) & \text{if } 0 < x < \theta; \\ \int_0^\theta \frac{1}{\theta} dt = 1 & \text{if } x \ge \theta. \end{cases}$$

and for $x \in (-\infty, \infty)$,

$$P(X_{(n)} \le x) = P(\cap_{i=1}^n \{X_i \le x\}) = (F(x))^n.$$

It is clear that $(F(x))^n = F_n(x)$ for $x \in (-\infty, \infty)$ for the F_n given in Part (b), so F_n is the CDF of $X_{(n)}$.

(c) Direct calculation gives

$$\int_{-\infty}^{x} f_n(t)dt = \begin{cases} 0 & \text{if } x \le 0; \\ \int_0^x \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} dt = (x/\theta)^n & \text{if } 0 < x < \theta; \\ \int_0^\theta \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} dt = 1 & \text{if } x \ge \theta, \end{cases}$$

which means $\int_{-\infty}^{x} f_n(t)dt = F_n(x)$ for $x \in (-\infty, \infty)$. Since F_n is the CDF of $X_{(n)}$, f_n is a PDF of $X_{(n)}$.

(d) The MSE of $X_{(n)}$ is

$$\begin{aligned} E(X_{(n)} - \theta)^2 &= \int_{-\infty}^{\infty} (t-\theta)^2 f_n(t)dt \\ &= \int_0^\theta (t-\theta)^2 \frac{n}{\theta} \left(\frac{t}{\theta}\right)^{n-1} dt \\ &= \frac{2\theta^2}{(n+1)(n+2)}. \end{aligned}$$

(e) $E(X_1) = \int_0^\theta x f_\theta(x)dx = \frac{\theta}{2}$. Solving $\bar{X} = \frac{\theta}{2}$ gives $\theta = 2\bar{X}$, so a method of moment eatimator of θ is $2\bar{X}$. The MSE of $2\bar{X}$ is

$$\begin{aligned} E(2\bar{X} - \theta)^2 &= Var(2\bar{X}) + (E(2\bar{X} - \theta))^2 \\ &= \frac{4Var(X_1)}{n} = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}. \end{aligned}$$

- In Example 1, we can also approximate each MSE at given a θ value by generating random samples to compute the squared errors and then computing the average of the squared errors as an approximate MSE. The R commands are given below.

```
set.seed(1)
theta <- 1
n <- 1000
m <- 10^6
```

```
 se <- rep(0, m)
 for (i in 1:m){
   x <- runif(n, 0, theta)
   se[i] <- (max(x) - theta)^2
 }
 mean(se)
#approximate MSE of MLE: 1.985207e-06
2*theta^2/((n+1)*(n+2))
#MSE of MLE: 1.994014e-06

 set.seed(1)
 theta <- 1
 n <- 1000
 m <- 10^6
 se <- rep(0, m)
 for (i in 1:m){
   x <- runif(n, 0, theta)
   se[i] <- (2*mean(x) - theta)^2
 }
 mean(se)
#approximate MSE of MME: 0.0003333407
theta^2/(3*n)
#MSE of MME: 0.0003333333
```

- In Example 1, the MSE of MLE is smaller than the MSE of MME for all $\theta > 0$ when $n \geq 3$. In general, it is impossible to find an estimator with smallest MSE for all $\theta \in \Theta$, but it is possible to find an estimator with smallest MSE among unbiased estimators based on Lehmann-Scheffé Theorem or Cramér-Rao lower bound (Rao-Cramér lower bound).

- Bias of an estimator. Suppose that T_n is an estimator of $g(\theta)$ based on a sample of size n. $E(T_n) - g(\theta)$ is called the bias of T_n. If the bias is 0 for all $\theta \in \Theta$, then T_n is called an unbiased estimator of $g(\theta)$.

- Example 2. Suppose that $(X_1, \ldots, X_n)$ is a random sample and $X_1 \sim N(\mu, \sigma^2)$. Let $\bar{X} = \sum_{i=1}^n X_i/n$ and

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}.$$

  Find the bias of $\sum_{i=1}^n (X_i - \bar{X})^2/n$ and S^2 (as estimators of σ^2).

  Sol. Since

$$(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2,$$

and

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

we have

$$\begin{aligned} E((n-1)S^2) &= E\sum_{i=1}^{n}(X_i-\mu)^2 - En(\bar{X}-\mu)^2 \\ &= nVar(X_1) - nVar(\bar{X}) \\ &= n\sigma^2 - n\left(\frac{\sigma^2}{n}\right) = (n-1)\sigma^2. \end{aligned}$$

so $E(S^2) = \sigma^2$ and the bias of S^2 is $E(S^2)-\sigma^2 = 0$. The bias of $\sum_{i=1}^{n}(X_i - \bar{X})^2/n$ is

$$\begin{aligned} E(n^{-1}(X_i-\bar{X})^2) - \sigma^2 &= E((n-1)S^2/n) - \sigma^2 \\ &= \frac{(n-1)\sigma^2}{n} - \sigma^2 \\ &= -\frac{\sigma^2}{n}. \end{aligned}$$

- In Example 2, the assumption $X_1 \sim N(\mu, \sigma^2)$ can be replaced by that $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$.

- Suppose that $(X_1, \ldots, X_n)$ is a sample and the distribution of $(X_1, \ldots, X_n)$ is determined by a parameter vector θ, where θ is in some space Θ. Consider the estimation of $g(\theta)$, where g is a known real-valued function. Suppose that T_n is an unbiased estimator of $g(\theta)$ and for every W_n that is an unbiased estimator of $g(\theta)$, we have

  $$Var(T_n) \leq Var(W_n)$$

  for all $\theta \in \Theta$. Then T_n is called a (uniformly) minimum variance unbiased estimator (UMVUE or MVUE) of $g(\theta)$.

- Lehmann-Scheffé Theorem. Suppose that $(X_1, \ldots, X_n)$ is a sample whose distribution is determined by a parameter vector θ, where the parameter vector θ is in some space Θ. Suppose that T_n is a statistic that is sufficient and complete, and U_n is an unbiased estimator of $g(\theta)$. Then $E(U_n|T_n)$ is the unique MVUE of $g(\theta)$ if $Var(E(U_n|T_n))$ is finite.