3

Evaluating the Goodness of an Estimator: Bias, Mean-Square Error, Relative Efficiency

Consider a population parameter θ for which estimation is desired. For example, θ could be the population mean (traditionally called μ) or the population variance (traditionally called σ^2). Or it might be some other parameter of interest such as the population median, population mode, population standard deviation, population minimum, population maximum, population range, population kurtosis, or population skewness.

As previously mentioned, we will regard parameters as numerical characteristics of the population of interest; as such, a parameter will be a fixed number, albeit unknown. In Stat 252, we will assume that our population has a distribution whose density function depends on the parameter of interest. Most of the examples that we will consider in Stat 252 will involve continuous distributions.

Definition 3.1. An *estimator* $\hat{\theta}$ is a statistic (that is, it is a random variable) which after the experiment has been conducted and the data collected will be used to estimate θ.

Since it is true that any statistic can be an estimator, you might ask why we introduce yet another word into our statistical vocabulary. Well, the answer is quite simple, really. When we use the word estimator to describe a particular statistic, we already have a statistical estimation problem in mind.

For example, if θ is the population mean, then a natural estimator of θ is the sample mean. If θ is the population variance, then a natural estimator of θ is the sample variance. More specifically, suppose that $Y_1, \ldots, Y_n$ are a random sample from a population whose distribution depends on the parameter θ. The following estimators occur frequently enough in practice that they have special notations.

- sample mean: $\overline{Y} := \frac{1}{n}\sum_{i=1}^{n} Y_i$
- sample variance: $S^2 := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$

- sample standard deviation: $S = \sqrt{S^2} \geq 0$
- sample minimum: $Y_{(1)} := \min\{Y_1, \ldots, Y_n\}$
- sample maximum: $Y_{(n)} := \max\{Y_1, \ldots, Y_n\}$
- sample range: $R = Y_{(n)} - Y_{(1)}$

For example, suppose that we are interested in the estimation of the population minimum. Our first choice of estimator for this parameter should probably be the sample minimum. Only once we've analyzed the sample minimum can we say for certain if it is a good estimator or not, but it is certainly a natural first choice. But the sample mean $\overline{Y}$ is also an estimator of the population minimum. Indeed, any statistic is an estimator. However, even without any analysis, it seems pretty clear that the sample mean is not going to be a very good choice of estimator of the population minimum. And so this is why we introduce the word estimator into our statistical vocabulary.

Notation. In Stat 251, if we assumed that the random variable Y had an $\text{Exp}(\theta)$ distribution, then we would write the density function of Y as

$$f_Y(y) = \begin{cases} \theta e^{-\theta y}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

In Stat 252, to emphasize the dependence of the distribution on the parameter, we will write the density function of Y as

$$f(y|\theta) = \begin{cases} \theta e^{-\theta y}, & y > 0, \\ 0, & y \leq 0. \end{cases}$$

In Stat 252, the estimation problem will most commonly take the following form. Suppose that θ is the parameter of interest. We will take $Y_1, \ldots, Y_n$ to be a random sample with common density $f(y|\theta)$, and we will find a suitable estimator $\hat{\theta} = g(Y_1, \ldots, Y_n)$ for some real-valued function of the random sample g.

Three of the measures that we will use to assess the goodness of an estimator are its bias, its mean-square error, and its standard error.

Definition 3.2. If $\hat{\theta}$ is an estimator of θ, then the *bias* of θ is given by

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

and the *mean-square error* of $\hat{\theta}$ is given by

$$\text{MSE}(\hat{\theta}) := \mathbb{E}(\hat{\theta} - \theta)^2.$$

We say that $\hat{\theta}$ is *unbiased* if $B(\hat{\theta}) = 0$.

Exercise 3.3. Show that $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + [B(\hat{\theta})]^2$. In particular, conclude that if $\hat{\theta}$ is unbiased, then $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta})$.

Definition 3.4. If $\hat{\theta}$ is an estimator of θ, then the *standard error* of $\hat{\theta}$ is simply its standard deviation. We write $\sigma_{\hat{\theta}} := \sqrt{\mathrm{Var}(\hat{\theta})}$.

Example 3.5. Let $Y_1, \ldots, Y_n$ be a random sample from a population whose density is

$$f(y|\theta) = \begin{cases} 3\theta^3 y^{-4}, & \theta \le y, \\ 0, & \text{otherwise}, \end{cases}$$

where $\theta > 0$ is a parameter. Suppose that we wish to estimate θ using the estimator $\hat{\theta} = \min\{Y_1, \ldots, Y_n\}$.

(a) Compute $B(\hat{\theta})$, the bias of $\hat{\theta}$.
(b) Compute $\mathrm{MSE}(\hat{\theta})$, the mean-square error of $\hat{\theta}$.
(c) Compute $\sigma_{\hat{\theta}}$, the standard error of $\hat{\theta}$.

Solution. (a) Since $B(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$, we must first compute $\mathbb{E}(\hat{\theta})$. To determine $\mathbb{E}(\hat{\theta})$, we need to find the density function of $\hat{\theta}$, which requires us first to find the distribution function of $\hat{\theta}$. As we know from Stat 251, there is a "trick" for computing the distribution function of a minimum of random variables. That is, since $Y_1, \ldots, Y_n$ are i.i.d., we find

$$\begin{aligned} P(\hat{\theta} > x) = P(\min\{Y_1, \ldots, Y_n\} > x) &= P(Y_1 > x, \ldots, Y_n > x) \\ &= [P(Y_1 > x)]^n. \end{aligned}$$

We know the density of Y_1, and so if $x \ge \theta$, we compute

$$P(Y_1 > x) = \int_x^\infty f(y|\theta)\,\mathrm{d}y = \int_x^\infty 3\theta^3 y^{-4}\,\mathrm{d}y = \theta^3 x^{-3}.$$

Therefore, we find

$$P(\hat{\theta} > x) = [P(Y_1 > x)]^n = \theta^{3n} x^{-3n} \quad \text{for } x \ge \theta$$

and so the distribution function for $\hat{\theta}$ is

$$F(x) = P(\hat{\theta} \le x) = 1 - P(\hat{\theta} > x) = 1 - \theta^{3n} x^{-3n}$$

for $x \ge \theta$, and $F(x) = 0$ for $x < \theta$. Finally, we differentiate to conclude that the density function for $\hat{\theta}$ is

$$f(x) = F'(x) = \begin{cases} 3n\theta^{3n} x^{-3n-1}, & x \ge \theta \\ 0, & x < \theta. \end{cases}$$

Now we can determine $\mathbb{E}(\hat{\theta})$ via

$$\begin{aligned}\mathbb{E}(\hat{\theta}) = \int_{-\infty}^{\infty} x \cdot f(x)\,\mathrm{d}x = \int_{\theta}^{\infty} x \cdot 3n\theta^{3n} x^{-3n-1}\,\mathrm{d}x &= 3n\theta^{3n}\int_{\theta}^{\infty} x^{-3n}\,\mathrm{d}x \\ &= 3n\theta^{3n} \cdot \frac{\theta^{-3n+1}}{3n-1} \\ &= \frac{3n}{3n-1}\theta.\end{aligned}$$

Hence, the bias of $\hat{\theta}$ is given by

$$B(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta = \frac{3n}{3n-1}\theta - \theta = \frac{\theta}{3n-1}.$$

(b) As for the mean-square error, we have by definition $\mathrm{MSE}(\hat{\theta}) = \mathbb{E}(\hat{\theta} - \theta)^2$ and so

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= \int_{-\infty}^{\infty} (x-\theta)^2 f(x)\,\mathrm{d}x = \int_{\theta}^{\infty} (x-\theta)^2 \cdot 3n\theta^{3n}x^{-3n-1}\,\mathrm{d}x \\ &= 3n\theta^{3n}\left[\int_{\theta}^{\infty} x^{1-3n}\,\mathrm{d}x - 2\theta\int_{\theta}^{\infty} x^{-3n}\,\mathrm{d}x + \theta^2\int_{\theta}^{\infty} x^{-3n-1}\,\mathrm{d}x\right] \\ &= 3n\theta^{3n}\left[\frac{\theta^{2-3n}}{3n-2} - 2\theta \cdot \frac{\theta^{1-3n}}{3n-1} + \theta^2 \cdot \frac{\theta^{-3n}}{3n}\right] \\ &= 3n\theta^2\left[\frac{1}{3n-2} - \frac{2}{3n-1} + \frac{1}{3n}\right] \\ &= \theta^2\left[\frac{(3n-1)(3n-2) - 9n(n-1)}{(3n-1)(3n-2)}\right] \\ &= \frac{2\theta^2}{(3n-1)(3n-2)}.\end{aligned}$$

(c) From the previous exercise, we know that $\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + [B(\hat{\theta})]^2$ and so

$$\begin{aligned}\mathrm{Var}(\hat{\theta}) = \mathrm{MSE}(\hat{\theta}) - [B(\hat{\theta})]^2 &= \frac{2\theta^2}{(3n-1)(3n-2)} - \left[\frac{\theta}{3n-1}\right]^2 \\ &= \frac{\theta^2}{3n-1}\left[\frac{2}{3n-2} - \frac{1}{3n-1}\right] \\ &= \frac{3n\theta^2}{(3n-1)^2(3n-2)}.\end{aligned}$$

Therefore, the standard error of $\hat{\theta}$ is

$$\sigma_{\hat{\theta}} = \sqrt{\mathrm{Var}(\hat{\theta})} = \frac{\theta}{(3n-1)}\sqrt{\frac{3n}{3n-2}}.$$

Example 3.5 (continued). Observe that $\hat{\theta}$ is not unbiased; that is, $B(\hat{\theta}) \neq 0$. According to our Stat 252 criterion for evaluating estimators, this particular

$\hat{\theta}$ is not preferred. However, as we will learn later on, it might not be possible to find *any* unbiased estimators of θ. Thus, we will be forced to settle on one which is biased. Since

$$\lim_{n\to\infty} B(\hat{\theta}) = \lim_{n\to\infty} \frac{\theta}{3n-1} = 0$$

we say that $\hat{\theta}$ is *asymptotically unbiased*. If no unbiased estimators can be found, the next best thing is to find asymptotically unbiased estimators.

Definition 3.6. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ, then the *efficiency of* $\hat{\theta}_1$ *relative to* $\hat{\theta}_2$ is

$$\text{Eff}(\hat{\theta}_1, \hat{\theta}_2) := \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}.$$

Remark. We can use the relative efficiency to decide which of the two unbiased estimators is preferred.

- If
$$\text{Eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} > 1,$$
then $\text{Var}(\hat{\theta}_2) > \text{Var}(\hat{\theta}_1)$. Thus, $\hat{\theta}_1$ has smaller variance that $\hat{\theta}_2$, and so $\hat{\theta}_1$ is preferred.
- On the other hand, if
$$\text{Eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)} < 1,$$
then $\text{Var}(\hat{\theta}_2) < \text{Var}(\hat{\theta}_1)$. Thus, $\hat{\theta}_2$ has smaller variance that $\hat{\theta}_1$, and so $\hat{\theta}_2$ is preferred.

Example 3.7. Suppose that $Y_1, \ldots, Y_n$ are a random sample from a Uniform$(0, \theta)$ population where $\theta > 0$ is a parameter. As you will show on Assignment #2, both

$$\hat{\theta}_1 := 2\overline{Y} \quad \text{and} \quad \hat{\theta}_2 := (n+1)\min\{Y_1, \ldots, Y_n\}$$

are unbiased estimators of θ. Compute $\text{Eff}(\hat{\theta}_1, \hat{\theta}_2)$, and decide which estimator is preferred.

Solution. Since the random variables $Y_1, \ldots, Y_n$ have common density

$$f(y|\theta) = \begin{cases} 1/\theta, & 0 \le y \le \theta, \\ 0, & \text{otherwise}, \end{cases}$$

we deduce $\mathbb{E}(Y_i) = \theta/2$ and $\text{Var}(Y_i) = \theta^2/12$ for all i. Since $Y_1, \ldots, Y_n$ are i.i.d., we find

$$\operatorname{Var}(\hat{\theta}_1) = \operatorname{Var}(2\overline{Y}) = 4\operatorname{Var}(\overline{Y}) = \frac{4}{n}\operatorname{Var}(Y_1) = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}.$$

As for $\hat{\theta}_2$, recall that the density of $Y_{(1)} := \min\{Y_1, \dots, Y_n\}$ is

$$f_{Y_{(1)}}(x|\theta) = \begin{cases} n\theta^{-n}(\theta - x)^{n-1}, & 0 \le x \le \theta, \\ 0, & \text{otherwise,} \end{cases}$$

so that

$$\mathbb{E}(Y_{(1)}) = \frac{\theta}{n+1}$$

and

$$\operatorname{Var}(Y_{(1)}) = \int_0^\theta \left(x - \frac{\theta}{n+1}\right)^2 n\theta^{-n}(\theta - x)^{n-1}\, dx = \frac{2\theta^2}{(n+1)(n+2)} - \frac{\theta^2}{(n+1)^2}.$$

Therefore,

$$\operatorname{Var}(\hat{\theta}_2) = (n+1)^2 \operatorname{Var}(Y_{(1)}) = \frac{n\theta^2}{n+2}.$$

We now find

$$\operatorname{Eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\operatorname{Var}(\hat{\theta}_2)}{\operatorname{Var}(\hat{\theta}_1)} = \frac{\frac{n\theta^2}{n+2}}{\frac{\theta^2}{3n}} = \frac{3n^2}{n+2} > 1$$

provided $n > 1$. Since $\operatorname{Eff}(\hat{\theta}_1, \hat{\theta}_2) > 1$, we conclude that $\operatorname{Var}(\hat{\theta}_2) > \operatorname{Var}(\hat{\theta}_1)$ so that $\hat{\theta}_1 = 2\overline{Y}$ is preferred to $\hat{\theta}_2 = (n+1)\min\{Y_1, \dots, Y_n\}$.

Example 3.8. Suppose that $Y_1, \dots, Y_n$ are a random sample with a common density depending on a parameter θ. Suppose further that $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased estimators of θ based on $Y_1, \dots, Y_n$, and that

$$\operatorname{Eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\operatorname{Var}(\hat{\theta}_2)}{\operatorname{Var}(\hat{\theta}_1)} = \frac{2n^3 + 5n + 1}{2n^3 + n^2 - 1}.$$

Which estimator is preferred?

Solution. We notice that $2n^3 + 5n + 1 < 2n^3 + n^2 - 1$ if and only if $n > 5$. Therefore, if the sample size is $n = 1$, 2, 3, 4, or 5, then $\hat{\theta}_1$ is preferred, but if $n > 5$, then $\hat{\theta}_2$ is preferred